GI-Edition

Lecture Notes in Informatics

Ralf H. Reussner, Anne Koziolek,
Robert Heinrich (Hrsg.)

INFORMATIK 2020

Back to the Future

28. September – 2. Oktober 2020
Karlsruhe

307

Proceedings

Ralf H. Reussner,Anne Koziolek, Robert Heinrich (Hrsg.) INFORMATIK 2020

GESELLSCHAFT
FÜR INFORMATIK

Ralf H. Reussner, Anne Koziolek, Robert Heinrich (Hrsg.)

50. Jahrestagung der Gesellschaft für Informatik
INFORMATIK 2020

Back to the Future

28. September – 2. Oktober 2020
Karlsruhe, Deutschland

Gesellschaft für Informatik e.V. (GI)

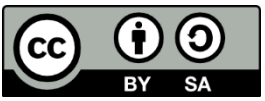
Lecture Notes in Informatics (LNI) - Proceedings
Series of the Gesellschaft für Informatik (GI)

Volume P-307

ISBN 978-3-88579-701-2
ISSN 1617-5468

Volume Editors
Prof. Dr. Ralf H. Reussner
Karlsruher Institut für Technologie (KIT) & FZI Forschungszentrum Informatik
Am Fasanengarten 5, D-76131 Karlsruhe, Germany
reussner@kit.edu | reussner@fzi.de
Prof. Dr. Anne Koziolek
Karlsruher Institut für Technologie (KIT)
Am Fasanengarten 5, D-76131 Karlsruhe, Germany
anne.koziolek@kit.edu
Dr. Robert Heinrich
Karlsruher Institut für Technologie (KIT)
Am Fasanengarten 5, D-76131 Karlsruhe, Germany
robert.heinrich@kit.edu

printed by Köllen Druck+Verlag GmbH, Bonn



Vorwort

Dass diese Jahrestagung eine ganz besondere sein würde, war früh klar: Es handelte sich nämlich um die 50. Jahrestagung der Gesellschaft für Informatik. So war das Motto „Back to the Future“ naheliegend: Wir blickten zurück auf die reiche Geschichte der Informatik im deutschsprachigen Raum. Kaum eine Disziplin beeinflusst den Alltag so stark wie die Informatik, bis hin zu den mittlerweile schon so oft beschworenen „disruptiven“ Veränderungen in unserer Wirtschaft, aber auch in Gesellschaft und Wissenschaft. Daher wollten wir natürlich auch nach vorne blicken und genau diese Veränderungen sowie die, die sich abzeichnen, diskutieren.

Dann kam die Corona-Pandemie. Am Tagungsmotto hatte sich nichts geändert, aber an der Tagungsorganisation alles: Nun war klar, dass diese 50. Jahrestagung auch gleichzeitig die erste virtuelle Jahrestagung der Gesellschaft für Informatik werden würde. Dabei kamen Fragen auf, wie man gerade den so wichtigen informellen Austausch, den wir an den bisherigen Jahrestagungen so sehr schätzten, ins Virtuelle verlagern könne. Sollten wir es überhaupt versuchen? Was macht man mit dem klassischen Empfang oder dem festlichen Dinner? Wie kann man den Bezug zum ursprünglichen Ort, der Stadt Karlsruhe, eigentlich herstellen, wenn alles online übertragen wird? Schnell aber fassten wir den in der Informatik typischen Beschluss: „It’s not a bug, it’s a feature!“ Vielleicht würden wir ja in einem rein virtuellen Format Teilnehmende erreichen können, die wir bei einer klassischen Jahrestagung nicht hätten erreichen können.

Und so kam der nun doppeldeutige Titel der Eröffnungsveranstaltung „Disruption – Dank Informatik“. Dieser spielt einerseits auf den Slogan der GI „... dank Informatik“ an. Andererseits betrachtet er die Informatik nicht nur, wie oben erläutert, als Verursacherin von Disruption, sondern auch als Hilfe zur Bewältigung der Disruptionen, die durch die Corona-Pandemie uns alle betreffen.

Was die Teilnehmenden erleben konnten, war der Versuch einer virtuellen Jahrestagung, bei der wir zum einen ein vielfältiges fachliches Angebot bereitstellten, welches insbesondere auch Schwerpunkte der Karlsruher Informatikforschung widerspiegelte. Zum anderen machten wir aber eben auch den Versuch, ein attraktives Pausenangebot für den Austausch herzustellen ebenso wie die Möglichkeit, Aussteller virtuell zu besuchen. Im Sinne des Programmhefts einer klassisch-physischen Tagung stellte unser „Taschenführer“ der virtuellen Tagung Informationen zusammen, die den Teilnehmenden Orientierung über das Programm, vor allem aber Überblick über die angebotenen Systeme der virtuellen Tagung gab. Das reichte vom Registrierungsprozess über die Teilnahme an den Veranstaltungen und fachlichen „Sessions“, bis hin zum Pausenprogramm.

Ein besonderes Highlight war die erwähnte Eröffnungsveranstaltung mit einer Impuls-Keynote von Martin Hubschneider (CEO der CAS AG) über digitale Souveränität in einer digitalen Ökonomie. In der darauffolgenden Podiumsdiskussion diskutierte er mit Frau Saskia Esken, MdB, Frau Professorin Dorothea Wagner (Vorsitzender des Wissenschaftsrats) Frau Professorin Ina Schieferdecker (Ministerialdirektorin des BMBF) und Ulrich Steinbach (Ministerialdirektor Wissenschaftsministerium Baden-Württemberg).

Das fachliche Programm reflektierte zum einen etablierte Schwerpunkte der Karlsruher Informatik: Data Science, Robotik und KI, Internet und Gesellschaft, sichere und zuverlässige Systeme, Software Engineering sowie autonomes Fahren, Mobilitätssysteme, Energieinformatik und digitale Gesundheit. Diese Themen zeigten aber auch die gegenwärtige Rolle der Informatik als zentraler Problemlöser für verschiedene Herausforderungen moderner Gesellschaften. Insofern wurden bewusst auch Themen herausgestellt, die einen Diskurs jenseits der rein technischen Betrachtung ermöglichen.

Eine Besonderheit dieses Jahres war der „Digital Innovation Day“. Hier versuchten studentische Teams aus dem ganzen Bundesgebiet, von Dienstagnachmittag bis Donnerstagfrüh IT-Innovationen zu Themen, die von Sponsoren vorgegeben werden, in neue Softwareprodukte umzusetzen. Dazu winkte neben dem Preisgeld von 2000 € eine sichtbare Auszeichnung auf der Konferenz.

Die beiden Keynotes stellten natürliche Highlights des Programms dar. Am Mittwochmorgen beleuchtete Herr Professor Dr. Ralf Herrtwich (NVIDIA) kritisch Visionen zum automatisierten Fahren; wo stehen wir und was wird kommen? Am Donnerstagmorgen verdeutlichte Frau Eva Zauke (Chief Knowledge Officer SAP SE), wie digitale Technologien neue Geschäftsmodelle ermöglichen und zeigte anhand von Beispielen, welche Rolle Unternehmenssoftware heute spielen kann.

Zwischen den Keynotes, am Mittwoch, beleuchteten wir ein Thema, das wir ganz gemäß der Informatik in einem regulären Ausdruck so formulierten: „Informatik in (der | die) Schule“. Beide Varianten des Titels sollten ausdrücken, dass wir zum einen über den gegenwärtigen Stand sprechen, zum anderen aber auch über Perspektiven des Ausbaus des Informatikunterrichts in den verschiedenen Schulformen. Mit im Panel vertreten waren Angehörige des Leitungsgremiums der GI-Fachgruppe Informatiklehrerinnen und -lehrer Baden-Württemberg, Professor Dr. Bernhard Standl von der PH Karlsruhe und Herr Thomas Menzel vom Kulturministerium Baden-Württemberg. Eingeleitet wurde es von einem Impulsvortrag von Dirk Fox (Gründer und Geschäftsführer der Secorvo Security Consulting GmbH), der seine bekannte Initiative für extra-curriculare Informatikbildung in Schulen vorstellte.

Am Donnerstag wurde in der Abschlussveranstaltung gemäß des Tagungsmottos „Back to the Future“ der Blick auf die letzten 50 Jahrestagungen der GI gerichtet: Professor Lockemann stellte als Herausgeber ein Buch vor, in dem, soweit möglich, die Organisatorinnen und Organisatoren der Jahrestagungen der GI zu Wort kamen. Es wurde eine Revue der Informatik im deutschsprachigen Raum der letzten Jahre, mit ihren Akteuren und Themen in Forschung, Lehre und Wirtschaft, denn nur, wenn man versteht, wo man herkommt, kann man den Weg in die Zukunft suchen. Und wer hätte den Blick in die Zukunft besser darstellen können, als IT-Start-Up-Unternehmen, die uns ihre neuen Geschäftsmodelle und -ideen in unterhaltsamen „Pitches“ vorstellten. Gemäß der Blicke auf erbrachte Leistungen und zukünftige Chancen wurden in dieser Abschlussveranstaltung auch die neuen Fellows und Junior-Fellows der GI ausgezeichnet. Ein besonderer Höhepunkt zum Abschluss wurde die Vergabe der Klaus-Tschira-Medaille für besondere Verdienste für die Informatik.

Eingerahmt wurde das Tagungsprogramm von einem reichhaltigen Workshop-Programm, welches einerseits zu ersten Mal rein virtuell stattfand, andererseits aber in bewährter Manier die Breite und Aktualität von Themen repräsentierte, die in Fachgruppen und Arbeitskreisen der GI diskutiert werden.

An dieser Stelle bleibt uns herzlichen Dank zu sagen an:

- Professor Dr. Alexander Mädche als Financial Chair
- Dr. Thomas Kühn als Publicity Chair
- Das Organisations-Team aus den KIT-Lehrstühlen und FZI-Gruppen, stellvertretend für alle Elena Kienhöfer für administrative Aufgaben und Jan Keim für die Technik als die führenden Personen des Teams
- Die Kollegen Professor Dr. Andreas Oberweis und Professor Dr. Dr h.c. Lockemann für unermüdliche Unterstützung und Hilfe durch reichhaltige Erfahrungen
- Die Chairs der fachlichen Sessions
- Die Organisatorinnen und Organisatoren der Workshops
- Alle Student Volunteers und unsere studentische Hilfskraft Ilona-Dewi Kusardi
- Die Sponsoren: andrena, Capgemini, CAS, De Gruyter Oldenbourg, Google, IBM, SAP
- Die Partner: AVDATA, FZI – Forschungszentrum Informatik in Karlsruhe und Berlin, Karlsruher Institut für Technologie (KIT), Messe Karlsruhe, Springer Vieweg, Weizenbaum-Institut Berlin
- Den Medienpartner Tagesspiegel
- Die Geschäftsstellen der GI in Bonn und Berlin
- Dr. Christopher Gerking für die Erstellung dieses Tagungsbandes

Hinzu kommt die Hilfe vieler GI-Mitglieder, die durch ihre Netzwerke auf vielfältige Art geholfen haben. Hier sieht man, dass die GI auch das ist: ein Netzwerk von Informatik-Experten, auf das man sich (auch als Organisator einer Jahrestagung) verlassen kann. Mit dieser Erfahrung kann man auch getrost die nächsten 50 Jahre erwarten.

Karlsruhe, im Oktober 2020

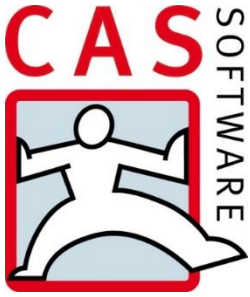
Ralf Reussner	Anne Koziolek	Robert Heinrich
(Gesamtleiter)	(Programmkoordinatorin)	(Workshop Chair)

Sponsoren

Wir danken den folgenden Unternehmen für die Unterstützung der Tagung als Sponsor.

andrena objects ag

CAS Software AG

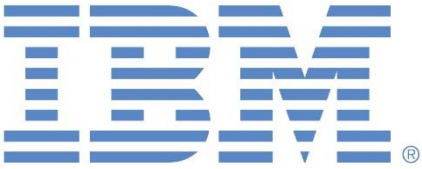


Capgemini Deutschland GmbH

SAP SE

Google Germany GmbH

IBM Deutschland GmbH



De Gruyter Oldenbourg

Partner

Wir danken den folgenden Unternehmen und Institutionen für die partnerschaftliche Unterstützung der Tagung.

Der Tagesspiegel

Weizenbaum-Institut

Springer Vieweg

Technik

Wir danken den folgenden Unternehmen für die technische Unterstützung der Tagung.

Karlsruher Messe- und Kongress GmbH

AVDATA GmbH

Tagungsleitung

Gesamtleitung:	Ralf H. Reussner, KIT & FZI
Programmkoordination:	Anne Koziolek, KIT
Workshops:	Robert Heinrich, KIT
Lokale Organisation:	Jan Keim, KIT
	Elena Kienhöfer, KIT
	Peter C. Lockemann, KIT & FZI
	Andreas Oberweis, KIT & FZI
	Gunther Schiefer, KIT
Finanzen:	Alexander Mädche, KIT
Pressearbeit:	Thomas Kühn, KIT
Tagungsband:	Christopher Gerking, KIT

Lokales Organisationsteam

Demian Frister	Karlsruher Institut für Technologie (KIT)
Sebastian Hahner	Karlsruher Institut für Technologie (KIT)
Jörg Henß	FZI Forschungszentrum Informatik
Angelika Kaplan	Karlsruher Institut für Technologie (KIT)
Heiko Klare	Karlsruher Institut für Technologie (KIT)
Sandro Koch	Karlsruher Institut für Technologie (KIT)
Nico Kopp	FZI Forschungszentrum Informatik
Sebastian Krach	FZI Forschungszentrum Informatik
Ilona-Dewi Kusardi	Karlsruher Institut für Technologie (KIT)
Martina Rapp	FZI Forschungszentrum Informatik
Frederik Reiche	Karlsruher Institut für Technologie (KIT)
Max Scheerer	FZI Forschungszentrum Informatik
Larissa Schmid	Karlsruher Institut für Technologie (KIT)
Maximilian Walter	Karlsruher Institut für Technologie (KIT)
Dominik Werle	Karlsruher Institut für Technologie (KIT)
Daniel Zimmermann	FZI Forschungszentrum Informatik

Track Chairs

Tamim Asfour	Karlsruher Institut für Technologie (KIT)
Ingmar Baumgart	FZI Forschungszentrum Informatik
Steffen Becker	Universität Stuttgart
Jürgen Beyerer	Karlsruher Institut für Technologie (KIT) & Fraunhofer IOSB
Klemens Böhm	Karlsruher Institut für Technologie (KIT)
Roger Gutbrod	SAP SE
Sami Haddadin	Technische Universität München (TUM)
Veit Hagenmeyer	Karlsruher Institut für Technologie (KIT)
Oliver Hillermeier	SAP SE
Stephan Jonas	Technische Universität München (TUM)
Hubert B. Keller	Karlsruher Institut für Technologie (KIT)
Kristian Kersting	Technische Universität Darmstadt
Birgitta König-Ries	Friedrich-Schiller-Universität Jena
Luise Kranich	FZI Forschungszentrum Informatik
Volkmar Lotz	SAP SE
Jörn Müller-Quade	Karlsruher Institut für Technologie (KIT) & FZI Forschungszentrum Informatik
Astrid Nieße	Carl von Ossietzky Universität Oldenburg
Thomas Schamm	Robert Bosch GmbH
Thomas Schlegel	Hochschule Karlsruhe
Stefan Schmid	Robert Bosch GmbH
Rainer Stiefelhagen	Karlsruher Institut für Technologie (KIT)
Ali Sunyaev	Karlsruher Institut für Technologie (KIT)
York Sure-Vetter	Karlsruher Institut für Technologie (KIT) & FZI Forschungszentrum Informatik
Thomas Usländer	Fraunhofer IOSB
Peter Vortisch	Karlsruher Institut für Technologie (KIT)
Dirk Weißer	INIT Innovation in Traffic Systems SE
Meike Zehlike	Max-Planck-Institut für Softwaresysteme
J. Marius Zöllner	Karlsruher Institut für Technologie (KIT) & FZI Forschungszentrum Informatik

Workshop-Organisation

Daniel F. Abawi	Hochschule für Technik und Wirtschaft des Saarlandes
Andreas Abecker	Disy Informationssysteme GmbH
Can Adam Albayrak	Hochschule Harz
Sören Auer	Technische Informationsbibliothek (TIB)
Gunnar Auth	Hochschule Meißen (FH) und Fortbildungszentrum
Grit Behrens	Fachhochschule Bielefeld
Jörg Benze	T-Systems Multimedia Solutions GmbH
Willi Bernhard	Fernfachhochschule Schweiz
Maximilian Blatt	accenture GmbH
Carsten Brockmann	Deloitte Consulting GmbH
Julian Bruns	Disy Informationssysteme GmbH
Manuel Burghardt	Universität Leipzig
Christian Czarnecki	Hochschule Hamm-Lippstadt
Anusch Daemi-Ahwazi	accenture GmbH
Jan deMeer	smartspacelab.eu GmbH
Hermann de Meer	Universität Passau
Julian Dierker	Universität Osnabrück
Ralf Dörner	Hochschule RheinMain
Alexander Dregger	FZI Forschungszentrum Informatik
Anna Fehrenbach	Universität Osnabrück
Michael Fellmann	Universität Rostock
Mario Gleirscher	Universität York
Matthias Goeken	Hochschule der Deutschen Bundesbank
Frauke Goll	FZI Forschungszentrum Informatik
Rüdiger Grimm	Fraunhofer SIT & Universität Koblenz-Landau
Mesut Güneş	Otto-von-Guericke-Universität Magdeburg
Hans-Joachim Hof	Technische Hochschule Ingolstadt
Gerrit Hornung	Universität Kassel
Reinhard Kahle	Universität Tübingen
Robert Kaiser	Hochschule RheinMain
Oliver Kamin	Fernfachhochschule Schweiz
Eva Kern	Leuphana Universität Lüneburg
Natalja Kleiner	FZI Forschungszentrum Informatik
Matthias König	Fachhochschule Bielefeld
Birger Lantow	Universität Rostock
Ralf Laue	Westsächsische Hochschule Zwickau
Sebastian Lehnhoff	Carl von Ossietzky Universität Oldenburg
Kerstin Lenk	Universität Tampere
Jens-Martin Loebel	Hochschule Magdeburg-Stendal
Ulrike Lucke	Universität Potsdam
Klaus Mainzer	Technische Universität München (TUM) & Universität Tübingen
Maximilian Marowsky	Universität Osnabrück
Ingo Mauser	EnBW Energie Baden-Württemberg AG

Bodo Möslein-Tröppner	Duale Hochschule Baden-Württemberg Ravensburg
Claudia Müller-Birn	Freie Universität Berlin
Stefan Naumann	Hochschule Trier
Simon Nestler	Technische Hochschule Ingolstadt
Astrid Nieße	Carl von Ossietzky Universität Oldenburg
Paul Ohm	Universität Osnabrück
Tom Petersen	Universität Hamburg
Axel Rennoch	Fraunhofer FOKUS
Jana-Rebecca Rehse	Universität Mannheim
Daniel Rost	DB Cargo AG
Thomas Schlegel	Hochschule Karlsruhe
Hartmut Schmeck	FZI Forschungszentrum Informatik
Christoph Sorge	Universität des Saarlandes
Indra Spiecker gen. Döhmann	Goethe-Universität Frankfurt am Main
Markus Stocker	Technische Informationsbibliothek (TIB)
Eldar Sultanow	Universität Potsdam
York Sure-Vetter	Karlsruher Institut für Technologie (KIT) & FZI Forschungszentrum Informatik
Waldemar Titov	Hochschule Karlsruhe
Mathias Trefzger	Hochschule Karlsruhe
Kristina Voigt	
Markus von der Heyde	vdH-IT
Karl Waedt	FRAMATOME GmbH
Anke Weidlich	Universität Freiburg
Doris Weßels	Fachhochschule Kiel
Axel Wiepke	Universität Potsdam
Volker Wohlgemuth	Hochschule für Technik und Wirtschaft Berlin
Sebastian Zug	Technische Universität Bergakademie Freiberg

Inhaltsverzeichnis

Fachliches Programm

Workshops

(Agiles) Enterprise Architecture Management in Forschung und Praxis

8. Workshop Umweltinformatik zwischen Nachhaltigkeit und Wandel

5th GI/ACM I4.0 Standardization Workshop on Industrial Automation and Control Systems

Künstliche Intelligenz für kleine und mittlere Unternehmen

Hochschule 2030

6. Workshop zum Stand, den Herausforderungen und Impulsen des Geschäftsprozessmanagements

Konzeptionelle Herausforderungen für die KI

Original oder Plagiat? – Der schnelle Weg zur wissenschaftlichen Arbeit im Zeitalter künstlicher Intelligenz(en)

Graph theory & ML with real-time IoT data

Herausforderungen zukünftiger cyber-physischer Energiesysteme

Innovativer Informatikunterricht – Theorie und Praxis

Recht und Technik – Datenschutz im Diskurs

3rd Workshop on Smart Systems for Better Living Environments

Künstliche Intelligenz in der Umweltinformatik

1st Workshop on Evaluating Intelligent and Ubiquitous Mobility Systems

Workshop on Tools and Concepts for Communication and Networked Systems

Methoden und Anwendungen der Computational Humanities

Autorenverzeichnis

Fachliches Programm

Software Engineering

INFORMATIK 2020 – Bericht der Session zum Thema Software Engineering

Steffen Becker [1], Hubert B. Keller [2]

Abstract: Dieser Artikel enthält einen kurzen Bericht über die drei Vorträge in der Session „Software Engineering“ der virtuellen INFORMATIK 2020 Tagung. Innerhalb der Session gab es drei Fachvorträge, die über ein Videokonferenzsystem vorgetragen wurden. Die Vorträge fanden interaktiv statt, Fragen konnten jederzeit in den Vortrag hereingegeben werden.

Keywords: Bericht; INFORMATIK 2020; Software Engineering; virtuelle Konferenz

1 Einleitung

Die Session „Software Engineering“ fand am Dienstag, 29.09.2020, von 12:45 - 14:15 Uhr mit 3 Vorträgen statt. Chairs der Sitzung waren Hubert B. Keller, Institute for Automation and Applied Informatics (IAI), Karlsruhe Institute of Technology (KIT), sowie Steffen Becker, Institute of Software Engineering, University of Stuttgart.

Die drei Vorträge waren

- Safe and Secure Software Engineering - the SPARK Approach” von Yannick Moy, AdaCore, Paris
- „Agile Organisation und Führung innerhalb der Softwareentwicklung: Wie geht das zusammen?“ von Andreas Störzbach, Trumpf, Deutschland und
- „Wie Cloud-native-Ansätze Architektur(arbeit) verändert“ von Thomas Franz, adesso.

2 „Safe and Secure Software Engineering - the SPARK Approach”, Yannick Moy, AdaCore, Paris

Yannick Moy ist Senior Software Engineer bei AdaCore und verantwortet die Entwicklung von SPARK. Als Co-Direktor leitet er das Labor ProofInUse. Er hat an der Entwicklung

[1] Softwarequalität und -architektur (SQA), Institut für Software Engineering, Universität Stuttgart, Universitätsstr. 38, 70569 Stuttgart, steffen.becker@iste.uni-stuttgart.de, https://orcid.org/0000-0002-4532-1460
[2] Fachgebiet Advanced Automation Technologies (A2T) und Arbeitsgruppe Reliability, Safety & Security of Software and Systems (RS4), Institute for Automation and Applied Informatics (IAI), Karlsruhe Institute of Technology (KIT), Hermann-von-Helmholtz-Platz 1, 76344 Eggenstein-Leopoldshafen, hubert.keller@kit.edu, https://orcid.org/0000-0001-9916-5283

doi:10.18420/inf2020_01

der Software-Quellcode-Analysatoren CodePeer, Frama-C und PolySpace Verifier C ++ mitgewirkt. Bei AdaCore arbeitet er an den Software-Quellcode-Analysatoren CodePeer und SPARK zur Erkennung von Fehlern, zur Überprüfung von Sicherheitseigenschaften und zu formalen Spezifikation. Daneben hat er zahlreiche Publikationen über formale Methoden und die SPARK Technologie veröffentlicht.

Yannick Moy ging zuerst auf den Zusammenhang zwischen Software ‚Qualität und Zuverlässigkeit ein. Je geringer die Wahrscheinlichkeit von Fehler in Software ist, um so höher ist die zu erwartende Zuverlässigkeit. Eigentlich ist schon Software mit nur einem Fehler unzuverlässig. Ziel ist es also so weit wie möglich in Richtung „zero-defect“ Software zu kommen. SPARK ist ein formaler Spezifikationsansatz aus den Bereichen Avonic, Militär und Schienensysteme, der mittlerweile seinen Weg auch in den Medizinbereich, in Automotive Software und insbesondere in den Security Bereich gefunden hat. SPARK ermöglicht über 5 Ebenen Security Anforderungen abzubilden und abzusichern. Neben der Spezifikationssprache existieren entsprechende Tools für den praktischen Einsatz. Yannick Moy zeigte wie durch formale Spezifikationen Seiteneffekte durch globale Variablen verhindert oder Vor- und Nachbedingungen angegeben werden können sowie welche Toolketten dazu existieren. SPARK baut auf der Sprache Ada mit den Kerneigenschaften der strengen Typisierung etc. auf und erweitert diese um formale Konstrukte, z.B. um Daten- und Flussabhängigkeiten zu analysieren, den Nachweis der Vermeidung von Laufzeitfehlern wie die Verletzung des Priority Ceiling Protocols usw. sicher zu stellen etc. Je nach Umfang der formalen Spezifikation werden 5 Software Assurance Levels erfüllt. Die Anwendung im industriellen Umfeld zeigte er an mehreren Beispielen mit der Angabe des Aufwandes diese Technologie in Projekten zu implementieren. Als Ergebnis des Aufwandes nannte er eine deutlich geringere Zahl an Software Fehlern, einen geringeren Aufwand beim Testen und letztlich eine enorme Einsparung im gesamten Projekt. SPARK 2014 erlaubt dabei ausführbare formale Spezifikationen und geht dabei weit über MISRA C hinaus. Für den Einsatz von SPARK sind viele Dokumentationen verfügbar (siehe www.adacore.com).

3 „Agile Organisation und Führung innerhalb der Softwareentwicklung: Wie geht das zusammen?“, Andreas Störzbach, Trumpf, Deutschland

Andreas Störzbach von der Trumpf GmbH erläuterte in seinem Vortrag das Spannungsfeld zwischen Softwareentwicklung für langlebige, qualitativ hochwertige Produkte (in Fall von Trumpf meist komplexe Werkzeugmaschinen) und agiler Softwareentwicklung. Insbesondere die Spannungen, die durch eine klassische Organisation und Führung eines führenden Maschinenbauunternehmens mit der flexiblen und agilen Struktur bei der modernen Softwareentwicklung entstehen, standen im Zentrum des Vortrags.

Dabei hat Herr Störzbach nicht nur das Problem gut verständlich präsentiert, sondern den anwesenden Zuhörern auch detaillierte Einblicke in die Lösungsansätze bei Trumpf im allgemeinen und seinem Team im Besonderen gegeben.

Das Interesse der Zuhörer galt dabei insbesondere den Fragen der Unternehmenskultur. Die Frage, wie es ein Unternehmen schafft, die agile Struktur und Denkweise seinen Mitarbeitern zu vermitteln und vorzuleben, stand im Kern vieler Fragen der Zuschauer. Im Detail waren Themen zur klassischen Verteilung der Organisationsaufgaben (z.B. Matrixorganisation vs. Agile Projektteams) von großem Interesse. Wir hatten den Eindruck, dass die Zuhörer insbesondere von den Erfahrungen, die Herr Störzbach zu schildern hatte, viel profitiert haben.

4 „Wie Cloud-native-Ansätze Architektur(arbeit) verändert“, Thomas Franz, adesso, Deutschland

Thomas Franz von der adesso SE hat aus seiner Praxis heraus dargestellt, wie sich die Arbeit der Architekten durch Cloud-native Ansätze verändert hat. Dazu hat er verglichen, wie früher klassische Software gebaut und entworfen wurden, gegenüber der Vorgehensweise bei Anwendungen, die für die Cloud und ihre Eigenheiten optimiert werden.

Insbesondere die Aufteilung in kleinere Services, die unabhängig entwickelt, skaliert und betrieben werden können, führt hier doch zu deutlichen Veränderungen. Ein spezieller Aspekt dabei für den Architekten ist der, dass nun auch die Infrastrukturaspekte mit bedacht werden müssen. Dabei kommt zum Tragen, dass Architekten Infrastruktur wie Code behandeln können und damit effizient Zugriff auf große Ressourcenpools haben. Der Betrieb erfolgt in Cloud-Umgebungen und daher muss die Anwendung nicht mehr beim Kunden installiert und betrieben werden.

Aus dem Publikum kamen denn auch neben einigen technischen Detailfragen die Frage auf, ob nicht auch das Geschäftsmodell von adesso sich dahingegen wandelt, dass sie direkt für den Kunden Lösungen anbieten und sogar betreiben können, was in Einzelfällen wohl auch schon der Realität entspricht.

5 Zusammenfassung

Die Session hat ausgehend von dem formalen Ansatz mit SPARK über die Problematik einer agilen Organisation der Softwareentwicklung über die neuen Ansätze des Cloud-basierten Service Angebots gezeigt, dass die Konzepte und Methoden der Softwareentwicklung nicht stehen bleiben. Diese Entwicklungen integriert zusammen zu bringen, um Software kostengünstig in komplexen Organisationen unter Nutzung von, auch Cloud-basierten, Services den Anforderungen von Security, Safety und hoher Zuverlässigkeit zu entwickeln, wird eine weitere Herausforderung der Zukunft sein.

Steffen Becker und Hubert B. Keller

Semantics and Knowledge Engineering

R. Reussner, A. Koziolek, R. Heinrich (Hrsg.): INFORMATIK 2020,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2021 41

Using Knowledge Graphs to Manage a Data Lake

Henrik Dibowski [1], Stefan Schmid[2]

Abstract: Knowledge graphs as fundamental pillar of artificial intelligence are experiencing a strong demand. In contrast to machine learning and deep learning, knowledge graphs do not require large amounts of (training) data and offer a bigger potential for a multitude of domains and problems. This article shows the application of knowledge graphs for the semantic description and management of data in a data lake, which improves the findability and reusability of data, and enables the automatic processing by algorithms. Since knowledge graphs contain both the data as well as its semantically described schema (ontology), they enable novel ontology-driven software architectures, in which the domain knowledge and business logic can completely reside on the knowledge graph level. This article further introduces such a use case: an ontology-driven frontend implementation, which is able to fully adapt itself based on the underlying knowledge graph schema and dynamically render information in the desired manner.

Keywords: Artificial Intelligence; Ontology; Knowledge Representation; Knowledge Graph; Semantic Data Lake; Data Catalog; Semantic Search; Semantic Layer; Ontology-Driven UIs

1 Introduction

Artificial intelligence (AI) is one of the biggest topics in computer science, and we have seen a big development of machine learning (ML) and deep learning (DL) during the past ten years. As many other enterprises in the world, Bosch is heavily investing in becoming one of the world-leading AI companies. The foundation of the Bosch Center for Artificial Intelligence in 2017 was an important milestone and meanwhile employs over a hundred AI experts. With ML and DL, data-driven subsymbolic approaches have dominated the past decade. Ontologies and knowledge graphs are another fundamental pillar of the AI landscape and they are emerging from the shadows. These approaches are fundamentally different, as knowledge is not represented by the weights of a neural network (black box), but in an explicit, symbolic way by means of logic languages with formal semantics (white box). Their inferences are deterministic and their results are traceable and explainable, which constitutes an essential requirement for a wide range of applications, such as autonomous driving. At Bosch, we are facing a high demand for knowledge graphs from various business units and are seeing a strong momentum. That is not much of a surprise, as Gartner recently rated knowledge graphs as one of the most promising emerging AI technologies of the next five to ten years [Br19]. We believe that knowledge graphs have a way bigger potential

[1] Robert Bosch GmbH, Corporate Research, 71272 Renningen, Germany, henrik.dibowski@de.bosch.com, https://orcid.org/0000-0002-9672-2387
[2] Robert Bosch GmbH, Corporate Research, 71272 Renningen, Germany, stefan.schmid@de.bosch.com

doi:10.18420/inf2020_02

for being a truly disruptive technology than ML and DL, because they do not depend on large amounts of available data for training the networks, and because they are applicable for representing knowledge of practically any domain, be it engineering, sciences, the humanities, medicine, etc.

The shift from subsymbolic to symbolic approaches goes along with a fundamental paradigm shift: from the use of raw data to the use of knowledge in AI. Purely subsymbolic approaches assume that with more and more data, AI can beat even the best algorithms. On the contrary, symbolic approaches emphasize that the quality of the data is more important than the quantity, and weigh knowledge over raw data. Data without description of its meaning is of low value, and it takes a lot of effort to turn such data into business value. Knowledge and semantically well-described information however is a key for many kinds of novel applications, machine understanding and truly smart AI.

2 Semantic Data Lakes

With the development of cloud computing, *data lakes* have emerged as new kind of cloud-based repositories for storing large amounts of data. Data lakes can store data in its native format in a flat architecture and run different types of analytics on the data. However, just "dumping" large amounts of data into a data lake does not provide value on its own. Without a contextual, semantic description of the data and without provenance information, the data stored in a data lake is unlikely to be usable by people and machines others than the ones that stored it and can still remember what was stored where. We reckon that this accounts for a majority of all existing data lakes, which are rather "*data swamps*" than data lakes. Data in a data lake is only (re)usable by people and machines, if it is semantically well described, and if the data provenance is clearly defined and tracked. *Semantic data lakes* are a specific form of data lakes in which a semantic layer on top enriches and connects the data semantically. The semantic layer overcomes data silos and enables semantic search across all data. This facilitates completely new, advanced use cases and analytics on the entirety of stored data. In our opinion, the best technology for realizing semantic data lakes are semantic technologies and knowledge graphs.

2.1 The role of data catalogs

In the era of big data, *data catalogs* emerged as the standard for metadata management. In the last few years, new application areas have appeared and the volume and richness of metadata required has grown significantly. Data lakes constitute one such important new application for data catalogs, besides warehouses, master data repositories, etc. According to Gartner, a data catalog "... maintains an inventory of data assets through the discovery, description, and organization of datasets. The catalog provides context to enable data analysts, data scientists, data stewards, and other data consumers to find and understand a relevant dataset for the purpose of extracting business value." [Ed17].

2.2 State of practice and criticism

Current vendors offer a wide range of commercial data catalog software. A sample of such vendors includes Alation Data Catalog, Atlan Enterprise Data Catalog, Talend Data Catalog, Collibra Data Catalog, Informatica Enterprise Data Catalog, Microsoft Azure Data Catalog, Oracle Cloud Infrastructure Data Catalog, and even Google is joining the market with its Google Data Catalog. To our knowledge, however, none of these data catalogs uses or supports standard semantic technologies (W3C recommendations), nor do they allow for using existing ontology vocabularies. Rather, they are closed, propriety systems with their own metadata languages and glossaries.

Our solution differs from existing solutions by proposing a semantic data lake architecture that incorporates a semantic data catalog, built with standard semantic technologies, and that addresses provenance and access control for resources in the data lake. This solution is described in detail in the following sections.

3 Semantic data lake catalog ontology

As a primary contribution, this section describes a data catalog ontology for semantic data lakes: the DCPAC ontology (Data Catalog, Provenance, and Access Control). The DCPAC ontology can be applied for adding a semantic layer to a data lake, which provides semantic description of the content, provenance, and access control permissions of the resources in a data lake. This ontology was created by combining several common, (predominantly) standardized ontology vocabularies and by aligning and extending them where necessary.

3.1 Ontology layer architecture

A big benefit of semantic technologies is the possibility to reuse, combine and extend existing ontology vocabularies, instead of reinventing the wheel. There is a large amount of open and commonly used ontologies available, which cover many domains and specify the domain knowledge and expertise of thousands of domain experts and knowledge engineers. One can only benefit from reusing such vocabularies, and further contribute and extend them. Building on open and standardized vocabularies leads to interoperable metadata representations and enables the exchange of information between different systems and applications. This is the approach that we followed, and the result can be seen in Fig. 1. The figure shows a layer architecture diagram of the DCPAC ontology. Boxes in this figure each represent an ontology, and the arrows between them define the vocabularies that are imported. The DCPAC ontology is shown at the bottom, and recursively imports all other ontologies. Additionally, it defines SHACL constraints for validating instance data (ABox).

In the following subsection, the primary ontologies utilized by DCPAC are introduced.

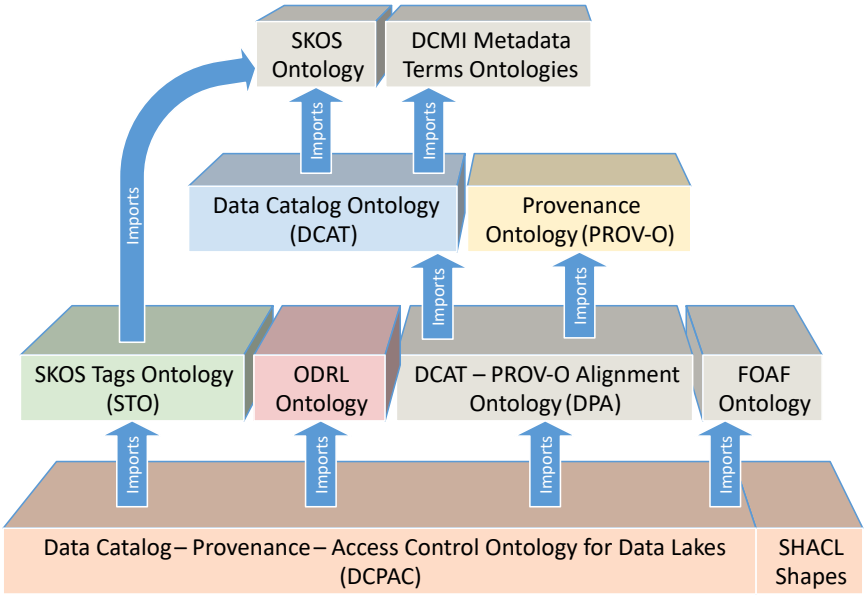


Fig. 1: Layer architecture of the semantic data lake catalog ontology

3.2 Utilized ontology vocabularies

The *Data Catalog (DCAT) ontology* (prefix: dcat) is a recent W3C recommendation and constitutes "... an RDF vocabulary designed to facilitate interoperability between data catalogs published on the Web." [Al20]. The DCAT ontology imports and uses the widely recognized SKOS [MB09] and DCMI Metadata Terms [Du20] ontologies. Its primary purpose in the context of the DCPAC ontology is the semantic description of the content of resources in a data lake.

The *Provenance Ontology (PROV-O)* (prefix: prov) is another W3C recommendation that "... provides a set of classes, properties, and restrictions that can be used to represent and interchange provenance information generated in different systems and under different contexts." [Le13]. Its purpose in the context of DCPAC is to describe the provenance of the data lake resources.

The *Open Digital Rights Language (ODRL) Ontology* (prefix: odrl) "... is a policy expression language from W3C that provides a flexible and interoperable information model ... for representing statements about the usage of content and services." [Ia17]. In our data lake scenario, ODRL is applied to defining access control permissions for the data lake resources.

The *DCAT – PROV-O Alignment (DPA) ontology* (prefix: dpa) [WW19] was created by the

W3C Dataset Exchange Working Group and contains alignment axioms between DCAT ontology and PROV-O. Thereby, it enhances the DCAT ontology with the ability to use PROV-O for expressing advanced provenance information.

The *Simple Knowledge Organization System (SKOS)* (prefix: skos) is "a common data model for sharing and linking knowledge organization systems" [MB09]. We use SKOS to define separate domain-specific *SKOS Tags Ontologies (STO)* that comprise sets of *semantic tags* and their semantic relationships. By assigning each dataset a set of relevant semantic tags, they provide semantic description of the content of data lake resources and enable semantic search on the resources in the semantic data lake.

3.3 Data Catalog – Provenance – Access Control (DCPAC) Ontology

The *Data Catalog – Provenance – Access Control (DCPAC) ontology* (prefix: dcpac) [Di20] is our primary contribution to the ontology layer architecture (see Fig. 1). It combines, aligns and extends the other ontologies. The ontology imports the ODRL, DPA, and FOAF ("Friend of a Friend") ontology [BM14] and optionally one or more STO ontologies, and recursively imports all other shown ontologies.

The DCPAC ontology aligns the DCAT ontology with the ODRL ontology by declaring the classes `dcat:Distribution` and `dcat:Resource` to be subclasses of `odrl:Asset`. This enables the definition of access control permissions for these DCAT classes and subclasses with the ODRL vocabulary. Another contribution is the alignment of PROV-O with the ODRL ontology by declaring the PROV-O class `prov:Agent` as subclass of `odrl:Party`, hence enabling all instances of `prov:Agent` to undertake roles in access control permissions. Additionally, the DCPAC ontology defines new subclasses of `dcat:Dataset` and `prov:Activity`, which allow for distinguishing different types of datasets and activities.

The DCPAC ontology is associated with a SHACL shapes definition file that defines a comprehensive set of SHACL constraints [KK17]. They can define cardinalities and type restrictions on properties, and regular expressions on the allowed values of string properties. Also complex constraints can be defined as SPARQL-based graph patterns. A SHACL engine can process the constraints and validate the consistency of the knowledge graph (ABox). That improves the integrity and quality and prevents issues.

3.4 The core vocabulary

This Section provides a short explanation of the core vocabulary of the DCPAC ontology and the primary imported vocabularies, which are explained in the previous sections. For the explanation, we refer to Fig. 2, which shows the main ontology classes as well as the most important object properties and datatype properties. The stereotypes shown for some of the classes in Fig. 2 contain their superclasses and hence their alignment to the other ontologies described in the previous sections.

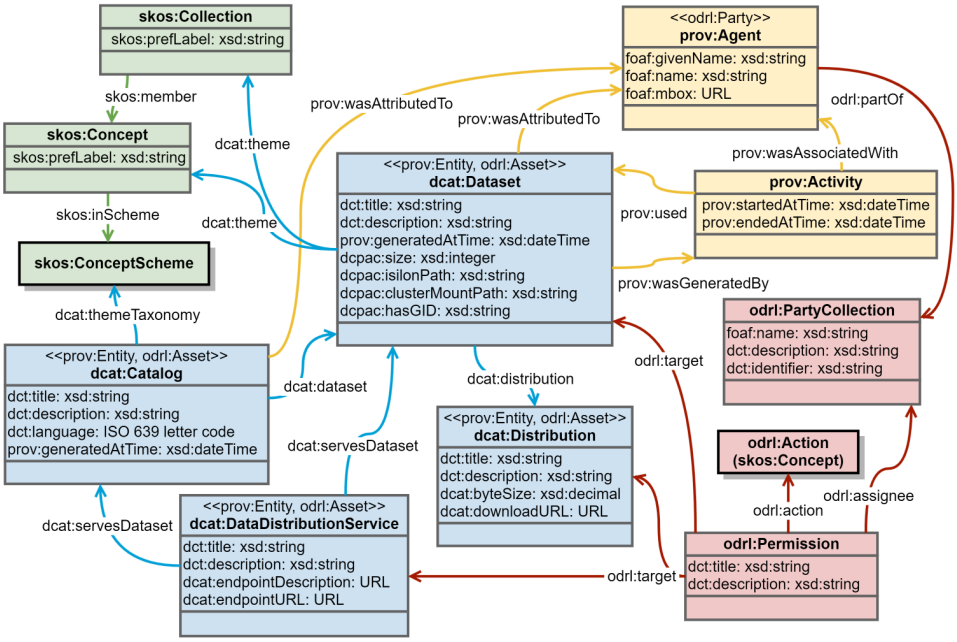


Fig. 2: The main classes and properties of the semantic data lake catalog ontology (TBox)

The *DCAT ontology classes* are shown in the center and bottom left of Fig. 2. The overall data catalog of the data lake is represented by one instance of class `dcat:Catalog`. It can contain many `dcat:Dataset` instances, one per resource in the data lake, e.g. raw data files, HBase or Hive tables, or RDF-based knowledge graphs. An instance of class `dcat:Distribution` models a specific representation of a dataset, comprising a specific serialization or schematic arrangement. Different distributions can exist for the same dataset, and are accessible via a URL (`dcat:downloadURL`). The data catalog and the datasets can each have several data distribution services (`dcat:DataDistributionService`), which are endpoints that provide access. They are accessible via an endpoint URL (`dcat:endpointURL`).

The *PROV-O classes and properties* shown in the top right part of Fig. 2 are used for modeling the provenance of the data catalog and its datasets (both declared as subclasses of `prov:Entity`), and for defining agents (e.g. person, software agent) they are attributed to (`prov:wasAttributedTo`) or that were involved in the activity of creating the dataset. Activities (`prov:Activity`) are initiated by agents (`prov:wasAssociatedWith`), create new datasets (`prov:wasGeneratedBy`), have an start and end time, and can use other datasets as input (`prov:used`).

Access control is defined by *classes and properties from the ODRL ontology*. An `odrl:Permission` can define an access rule for groups of agents (`odrl:Party-Collection`) to datasets, their distributions and/or data distribution services (`odrl:target`). The allowed

actions (`odrl:Action`), such as display, read, modify, delete, are defined as `skos:Concept` and attached via `odrl:action`.

SKOS finally is applied for defining the semantics of the content of a dataset. Therefore, the catalog refers to one or more sets of SKOS concepts (`skos:ConceptScheme`) that can be used for semantically tagging datasets. The defined SKOS tags can be either directly linked to a dataset (`dcat:theme`), or they can be bundled and linked as a collection (`skos:Collection`), which enables the reuse of (large) sets of SKOS tags.

The overall ontology vocabulary accomplished by the DCPAC ontology constitutes a vocabulary developed and harmonized by a broad community and standardized in most of the parts. It enables an interoperable representation and exchange of information beyond the limits of a specific data lake in place. The vocabulary is widely domain independent, with the only exception being the SKOS tags ontologies, which make it customizable towards particular domains of interest.

4 Semantic data lake for automotive data

At Bosch, we have built a semantic data lake for our automotive data as a centralized platform for the engineering and testing of our autonomous driving applications. The architecture of our data lake is shown in Fig. 3. It stores large amounts of data collected from test drives, which involves the logging of hundreds of sensor readings that are being logged each millisecond, but also includes the vehicle configuration, driving maneuvers, weather conditions and other related information. This data quickly accumulates to Petabytes of data, which is stored in large Hadoop-based data stores.

During the ingestion and processing of new incoming data, the data lake catalog population service is triggered, which automatically creates a semantic description and layer on top of the data. The semantic layer is stored and managed as knowledge graph in the semantic data lake catalog using the vocabulary defined by the semantic data lake catalog ontology (see Section 3). Therefore, the data lake catalog population service reads the available metadata on the ingested data assets and constructs the relevant semantic data by aligning, annotating and enriching the input data with DCPAC concepts. The resulting knowledge graph forms a semantic layer on the data lake assets and describes their content, provenance and access rights. This information plays a crucial role in the semantic data lake, as it semantically describes the stored data and hence turns it into valuable information and information assets. It allows an in-depth tracking of provenance related aspects of the data, and a definition of access rights on the data assets.

The semantic data lake catalog is a key-component of the data lake architecture, as it handles and controls the retrieval and access of information. It enables a semantic search of data and improves findability. Data can be found faster, better and automatically even by machines, which enables advanced data analytics use cases. The overall data reuse is much higher, which ultimately results in less test drives and related effort and cost.

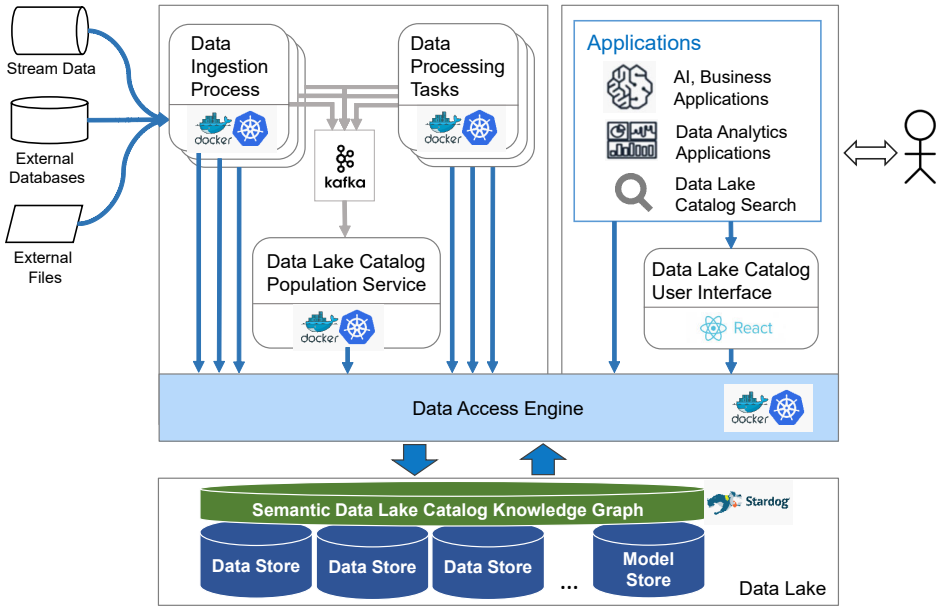


Fig. 3: Data lake architecture and role of semantic data lake catalog knowledge graph

5 Ontology-driven self-adaptive frontends

Now that more and more applications are using knowledge graphs as representation layer for data and information, new ontology-driven software architectures for self-adaptive frontends become feasible and are gaining momentum. This is doable due to knowledge graphs containing both the data as well as its rich semantically defined schema, which can describe domain knowledge, laws and constraints in a detailed way. Applications and frontends can utilize this knowledge, without the need of duplicating or repeating that inside their code. In an ideal scenario, business logic and expert knowledge can reside completely on the knowledge graph, and the applications and frontends can be independent of the model, domain or even use case.

As a first step into that direction, we have implemented self-adaptive frontends that can dynamically render information from a knowledge graph in a generic way. An example web frontend can be seen in Fig. 4, which shows information from the semantic data lake catalog of our semantic data lake (see Fig. 3). What information is to be shown where and how is defined by views on the knowledge graph. We followed the proposal of Tim Berners-Lee [Ti19] and used a combination of SHACL shapes and forms in order to define different views on a graph in a generic, expressive way. Dynamically at start up, the frontends request the view from the knowledge graph to be rendered. They receive all required information, such as the classes to be shown as tabs in the header and their labels in the required language, the

properties to be shown for each class, their type, cardinalities, datatypes, range, supported languages, their ordering in the frontend, as well as the properties that can be used for filtering the instances, along with their filter operators and possible values. But also, who has read access or who can change, create or delete instances is controlled by the knowledge graph.

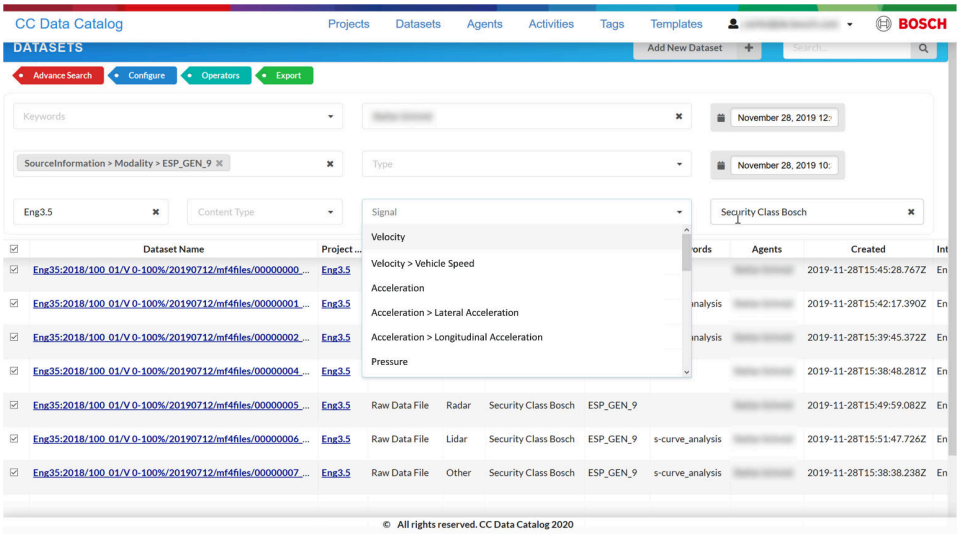


Fig. 4: Ontology-driven self-adaptive web frontend showing the semantic data lake catalog

Such ontology-driven self-adaptive frontends can dynamically render views on knowledge graphs. We believe that the initial overhead of realizing such next-generation ontology-driven software architectures and applications will pay off quickly, not only since changes of the underlying model do not require any changes at the frontend.

6 Conclusion

Semantic technologies and knowledge graphs are experiencing a high demand and are gaining strong momentum. The shift from ML and DL to symbolic approaches goes along with a fundamental paradigm shift: from the use of raw data to the use of knowledge in AI. The massive amounts of data in a data lake are only of value if the context of the data is semantically well described and its provenance is defined. We have realized a semantic data lake for automotive data, in which ontologies and knowledge graphs fulfil this role and are of utmost importance. Knowledge graphs enable us to search data based on rich semantics (context and provenance). Moreover, they also allow intelligent agents to automatically search and access the data, and thus, contribute to a significantly higher data reuse, which ultimately results in less test drives, and related efforts and costs.

With knowledge graphs as novel representation layer for data and information, new ontology-driven software architectures become feasible. We have implemented ontology-driven self-adaptive frontends that dynamically render information from a knowledge graph in a generic way. In an ideal scenario, both business logic and expert knowledge can reside completely in the knowledge graph, and the applications and frontends can be fully reused across domains and use cases. This will allow us to move from developing applications and frontends via coding (i.e. writing programs) towards modeling, where we define the views on the information and business logic in the knowledge graph itself.

Bibliography

[Al20] Albertoni, R.; Browning, D.; Cox, S., Beltran, A. G.; Perego, A.; Winstanley, P.: Data Catalog Vocabulary (DCAT) – Version 2. W3C Recommendation, https://www.w3.org/TR/vocab-dcat-2/, 2020, accessed: 14/10/2020.

[BM14] Brickley, D.; Miller, L.: FOAF Vocabulary Specification 0.99. http://xmlns.com/foaf/spec/, 2014, accessed: 14/10/2020.

[Br19] Brant, K.; Hare, J.; Sicular, S.: Hype Cycle for Artificial Intelligence. Gartner Research (ID: G00369840), 2019.

[Di20] Dibowski, H.; Schmid, S.; Svetashova, Y.; Henson, C.; Tran, Tuan: Using Semantic Technologies to Manage a Data Lake: Data Catalog, Provenance and Access Control. In: 13th International Workshop on Scalable Semantic Web Knowledge Base Systems (SSWS 2020), Athens, Greece, 2020.

[Du20] Dublin Core Metadata Initiative Board: DCMI Metadata Terms. DCMI Recommendation, https://www.dublincore.org/specifications/dublin-core/dcmi-terms/, 2020.

[Ed17] Edjlali, R.; Duncan, A. D.; De Simoni, G.; Zaidi, E.: Data Catalogs Are the New Black in Data Management and Analytics. Gartner Research, 2017.

[Ia17] Iannella, R. et. al.: ODRL Version 2.2 Ontology. W3C, https://www.w3.org/ns/odrl/2/, 2017, accessed: 14/10/2020.

[KK17] Knublauch, H.; Kontokostas, D.: Shapes Constraint Language (SHACL). W3C Recommendation, https://www.w3.org/TR/shacl/, 2017, accessed: 14/10/2020.

[Le13] Lebo, T.; Sahoo, S.; McGuinmess, D.: PROV-O: The PROV ontology. W3C Recommendation, http://www.w3.org/TR/prov-o/, 2013, accessed: 14/10/2020.

[MB09] Miles, A.; Bechhofer, S.: SKOS Simple Knowledge Organization System Reference. W3C Recommendation, https://www.w3.org/TR/skos-reference/, 2009.

[Ti19] Berners-Lee, T.: Linked Data Shapes, Forms and Footprints. White paper, https://www.w3.org/DesignIssues/Footprints.html, 2019, accessed: 14/10/2020.

[WW19] W3C Dataset Exchange Working Group (DXWG): DCAT-PROV alignment ontology. W3C, https://github.com/w3c/dxwg/blob/gh-pages/dcat/rdf/dcat-prov.ttl, 2019.

Industrie 4.0

Industrie 4.0 – Aktuelle Entwicklungen aus Sicht der Informatik

Zusammenfassung aus der Sitzung SES-06 am Mittwoch, 30.09.2020

Thomas Usländer [1]

Abstract: Der Artikel fasst die in der Sitzung SES-06 Industrie 4.0 vorgetragenen Beiträge zusammenfassen und setzt sie in den Kontext der aktuellen Entwicklung rund um die Initiative Industrie 4.0 und deren Vision 2030.

Keywords: Industrie 4.0; Datensouveränität; Geschäftsmodelle; Interoperabilität; Semantik; International Data Spaces; Smart Factory Web; Edge Computing

1 Motivation

Die Initiative Industrie 4.0 erfreut sich weiterhin einer ungebrochenen Dynamik und ist zu der entscheidenden Kraft für den Industriestandort Deutschland geworden, wenn es um die Digitalisierung der industriellen Produktion geht. Aus Sicht der Informatik geht es dabei weniger um wissenschaftlich bedeutsame Erkenntnisgewinne, sondern eher um eine profilierte Anwendung und Integration bekannter Konzepte der Informatik (Meta-Datenmodelle, SOA-Konzepte, semantische Annotation, IT-Sicherheit). Die Innovation liegt darin, diese Konzepte so maßzuschneidern und miteinander zu kombinieren, dass sie zusammen genommen helfen, die Ziele der Industrie 4.0 einer vernetzten, flexiblen Produktion in einer vertrauenswürdigen und modernden IT-Umgebung umzusetzen. Während anfangs in der konzeptionellen Gremienarbeit die Interoperabilität bei der Modellierung der Assets (u.a. Maschinen) und der Kommunikation und Interaktion zwischen den Assets im Vordergrund stand und bis heute immer noch steht, geht es in der Vision 2030 der Plattform Industrie 4.0 zunehmend um weitreichendere Fragen wie z.B. die Autonomie und Souveränität der Industrie 4.0 Systeme und Technologien sowie die Unterstützung der Nachhaltigkeit der industriellen Produktion. Parallel dazu finden die Ergebnisse der Gremienarbeit der Plattform Industrie 4.0 zunehmend Eingang in die IT-Strategien und Produkte der beteiligten Unternehmen sowie in die nationale und internationale Standardisierung. In der Sitzung SES-06 haben wir deshalb eine Mischung aus forschungs- und industrieorientierten Themen aufgegriffen wie z.B. Geschäftsmodelle, Semantische

[1] Fraunhofer IOSB, Informationsmanagement und Leittechnik (ILT), Fraunhoferstr. 1, 76131 Karlsruhe, thomas.uslaender@iosb.fraunhofer.de, https://orcid.org/0000-0001-6864-6631

doi:10.18420/inf2020_03

Interoperabilität, Edge Computing und Datensouveränität/International Data Spaces (IDS), die sich in den nachfolgend beschriebenen vier Vorträgen widerspiegeln

2 Vorträge

2.1 Transformation von Geschäftsprozessen mit Industrie 4.0 auf Basis von ERP und einer digitalen Supply Chain (Dominik Metzger, SAP)

Ausgehend von einer strategischen, markt- und werteorientierten Betrachtung von Industrie 4.0 stellt dieser Vortrag die SAP Strategie Industry 4.Now vor, die sich aus den drei Elementen „Center on Customers“, „Reinvent Production“ und „Connect the Entire Company“ zusammensetzt. Daraus ergibt sich die Industrie 4.0 Vision von SAP eines smarten Unternehmens („Intelligent Enterprise“): Maschinendaten vertikal integrieren zur Unterstützung von horizontalen Geschäftsprozessen auch über Unternehmensgrenzen hinaus.

Während Edge/Cloud Computing, IoT, 5G, Industrial Big Data Management und KI als grundlegend notwendige Technologien gesehen werden, sieht SAP Standardisierung als Schlüsselerfolgsfaktor für Industrie 4.0-basierte Geschäftsnetzwerke. Dazu gehören neben den SDOs (Standards Developing Organizations) wie z.B. die OPC Foundation, ECLASS und IEC, auch Industriekonsortien wie z.B. die PlattformIndustie 4.0, das Industrial Internet Consortium (IIC), die International Data Spaces Association (IDSA) und die GAIA-X Foundation AISBL. Zur Umsetzung der Strategie Industry 4.Now fördert SAP die Entwicklung von „intelligent products“, „intelligent factories“, „intelligent assets“ und „empowered people“.

2.2 Semantische Interoperabilität in cyber-physikalischen Produktionssystemen (Tizian Schröder, Chrtistian Diedrich, Otto Von Guericke Universität Magdeburg)

Die Vision des industriellen Internet of Things (IIoT) sieht vor, das Paradigma des Internet of Things (IoT) auf industrielle Anwendungsdomänen anzuwenden. Dies schließt Bereiche wie beispielsweise Städte, Gebäude, Fertigung, Transport, Landwirtschaft oder Medizin ein, da hier Informationen entlang einer Wertschöpfungskette ausgetauscht werden. Hierbei wird über eine Vielzahl dezentraler Systeme ein dezentral organisiertes Netzwerk von Systemen über das Internet gebildet, in dem autonome, intelligente Teilnehmer zum Zweck ihrer individuellen Zielerreichung miteinander interagieren. Als grundlegendste Motivation dieser Bemühung kann es angesehen werden, die schon heute enorme und zukünftig potentiell noch weiterwachsende Komplexität von Netzwerken von Systemen dadurch besser beherrschbar zu machen, dass von einem zentralistischen, auf der Automatisierungspyramide basierendem Ansatz zu einem dezentralen Ansatz übergegangen wird. Der Idee der cyber-physikalischen (Produktions-)Systeme (CP(P)S) folgend werden physische Systeme durch eine digitale Repräsentanz erweitert, um eine Integration der physischen Welt in die

Informationswelt vorzunehmen (Digitalisierung). Auf diesem Konzept aufbauend entsteht in der Digital Factory ein Netzwerk von Systemen, in dem diese als dezentrale Akteure nachrichtenbasiert miteinander interagieren. In derartigen verteilten Systemen müssen, insbesondere bei Domänen-übergreifender Kommunikation, mehrere Ebenen der Interoperabilität berücksichtigt werden, um Informationen austauschen zu können. Dem ETSI White Paper ‚Achieving technical interoperability' folgend, unterschieden wir bezüglich der Interoperabilität zwischen den folgenden vier Ebenen: technische, syntaktische, semantische und organisatorische Interoperabilität.

Während das ISO/OSI-Modell den Bereich der technischen Interoperabilität abdeckt, berühren Middleware-Systeme bereits die syntaktische und semantische Ebene. Semantische Interoperabilität gilt als einer der Schlüssel zur Umsetzung der Ideen hinter dem IIoT Paradigma. Damit heterogene Systeme über Anwendungsdomänen hinweg interoperabel kommunizieren können, wird zunächst eine präzise Definition des Begriffs ‚Semantik' benötigt. Denn nur auf einer formalen Basis lassen sich Informationen zwischen verteilten IIoT-Geräten korrekt austauschen. Dieser Beitrag fokussierte insbesondere auf die Ebene der semantischen Interoperabilität und lieferte eine formale Definition des Begriffs ‚Semantik'. Er hilft, Lösungsansätze für die semantische Interoperabilität innerhalb von Netzwerken von Systemen weiterzuentwickeln. Auf der Ebene des Verständnisses, zumindest aber was die Terminologie betrifft, ist diese Sichtweise auf den Begriff der Semantik nach unserer Erkenntnis derzeit nicht etabliert.

Stattdessen existiert eine Reihe von intuitiv motivierten Ansätzen, die beabsichtigen, den Begriff der Semantik zu präzisieren wie die Wissenspyramide oder das Levels of Conceptual Interoperability Model (LCIM). Der Beitrag stellte diese Ansätze vor und ordnete diese kritisch in den Kontext des vorgeschlagenen Modells ein. Zudem wurden weitere bestehende aber einen anderen Zweck adressierende Konzepte wie Ontologien, die Semantik von Datentypen und die axiomatische, denotationelle sowie operationelle Semantik von Programmen in Beziehung zum Inhalt des Beitrags gesetzt.

2.3 Strategische Rolle des Edge Computing für Industrie 4.0 Anwendungsszenarien (Sebastian Ritz, German Edge Cloud GmbH & Co. KG)

Edge-Computing in Industrie 4.0-Umgebungen wird immer wichtiger für die Lösung rechen- oder datenintensiver Probleme in der Optimierung der Gesamtanlageneffektivität (Overall Equipment Effectiveness OEE) von industriellen Produktionsumgebungen. Dies trifft insbesondere zu bei der Verarbeitung von Maschinenmassendaten, die ontologisch sauber aufbereitet werden müssen, und/oder bei vorgegebenen Echtzeitbeschränkungen. Außerdem kann man beobachten, dass sich unterschiedliche Edge-Computing-Architekturen entwickeln, die auf verschiedene Problemklassen der OEE-Optimierung und des Digital Twin-Managements zugeschnitten sind.

Im Vortrag wurde ein Überblick über Edge-Computing-Architekturen in Produktionswerkstätten (shopfloor) und ihre Relevanz für verschiedene Industrie 4.0-Anwendungsfälle gegeben und erläutert, warum die Edge-Cloud-Technologie im Shopfloor helfen kann, Hürden in digitalen Transformationsprogrammen zu überwinden.

2.4 International Data Spaces zur Datensouveränität in Industrie 4.0 (Ljiljana Stojanovic, Fraunhofer IOSB)

Die Arbeiten zu den International Data Spaces (IDS) sind mit den Zielsetzungen der Industrie 4.0 verzahnt in der sogenannten IDS-Industrial Community (IDS-I) unter Federführung des Fraunhofer IOSB. Beide Konzepte, also der IDS Connector einerseits und die Verwaltungsschale (VWS) andererseits, ergänzen einander. In diesem Vortrag werden mögliche Kombinationen von IDS und VWS diskutiert, um die Kommunikation mit einer VWS sicher und datensouverän zu gestalten. Die Idee besteht darin, weiterhin die Standardprotokolle und Nachrichtenformate der Verwaltungsschalten zu verwenden, sie aber mit IDS-Komponenten anzureichern, um das Vertrauensniveau zu erhöhen und so die Nutzungskontrolle der Daten zu steuern.

Der Vortrag stellte den vom Fraunhofer IOSB entwickelten IDS Factory Connector vor, der Industrie 4.0 Anwendungsfälle für sicheren, datensouveränen, semantischen und interoperablen Datenaustausch unterstützt und somit die Schnittstelle zwischen Industrie 4.0 und IDS darstellt. Abschließend wurden die Nutzung des Factory Connectors im Kontext des Smart Factory Web (https://www.smartfactoryweb.de) diskutiert, einer Blueprint-Architektur für resiliente und nachhaltige Produktionsökosysteme inkl. der Lieferketten. Eine Ausprägung des Smart Factory Web ist die prototypische Realisierung eines industriellen Marktplatzes für die industrielle Produktion diskutiert, der auch als Referenzanwendungsfall für die IDS-Industrial Community dient.

3 Ausblick

Diese Sitzung hat gezeigt, dass die Initiative Industrie 4.0 in einer entscheidenden Phase angekommen ist. Das Konzept der Verwaltungsschale (VWS) wird zunehmend gesehen als ein entscheidendes Modellierungselement zur Realisierung von interoperablen digitalen Zwillingen in einer Industrie 4.0 Systemumgebung. Interoperabilität, auch auf der semantischen Ebene, wird zum Schlüsselfaktor für ein effizientes Datenmanagement und darauf aufbauende neue Geschäftsmodelle. Die harmonische Integration von weiteren unverzichtbaren, nicht-funktionalen Anforderungen wie IT-Sicherheit und Datensouveränität sowie damit einhergehende Industrie 4.0 Infrastrukturdiensten werden weitere Bausteine für die Umsetzung von Industrie 4.0 in der geschäftlichen Praxis, der Abbildung eine GAIA-X Systemumgebung und in der internationalen Standardisierung.

Data Science

Data Science - more than just Machine Learning: A summary of the Data Science Session at INFORMATIK 2020

Birgitta König-Ries[1], Klemens Böhm[2]

Abstract: In this short article, we briefly summarize the Data Science session at INFORMATIK 2020. With three invited talks, the session focused on data-science challenges beyond the development of new machine learning models.

Keywords: Data Science; Data Management

1 Introduction

Too often, data science is reduced to machine learning (ML) and the development of new ML models. While the three talks that we had invited were very heterogeneous with respect to background (one talk from academia, two from industry) and thematic focus, they all agreed on one point: While the development of new ML models is obviously an important part of data science, it is just one part out of several important ones. The talks highlighted these other aspects very well, as discussed below.

2 Invited Talks

Benno Stein from Bauhaus University in Weimar introduced the Internet Archive hosted at Bauhaus University and its partner universities in Halle, Leipzig and Paderborn as a representative sample. On the one hand, this copy of a large subset of all web data is the basis for numerous research activities by his group. They range from work on argumentation and celebrity profiling to vandalism detection. On the other hand, to successfully build and maintain a suitable copy, specific research efforts are needed. Examples include work on Webpage segmentation and on the quality analysis of web crawling.

In the second talk of the session, **Torsten Grabs** from Snowflake argued that today, machine learning models are successfully being developed, but that their use in production environments often fails. He quoted a customer stating that ” we are good at building models that we don’t use”. Two of the main hurdles towards moving from development

[1] Friedrich Schiller University Jena, Heinz Nixdorf Chair for Distributed Information Systems, Jena, Germany, birgitta.koenig-ries@uni-jena.de
[2] Karlsruhe Institute of Technology (KIT), IPD, Karlsruhe, Germany, klemens.boehm@kit.edu

doi:10.18420/inf2020_04

to production are the lack of integration between data and ML platforms and the large heterogeneity of existing ML platforms; each platform is best suited for a certain subset of use cases. Torsten Grabs encouraged research on environments that facilitate declarative integration of data and analysis, analogously to what SQL and the relational algebra did for business intelligence.

The third and final talk of the session focused on a popular modeling paradigm to represent knowledge, namely Knowledge Graphs. The talk was given by **Steffen Lamparter** from Corporate Technology, Siemens AG. He argued against the development of a unified, company-wide Knowledge Graph and in favor of the ad hoc and use-case specific development of small Knowledge Graphs. The talk showed examples of a wide variety of use cases of different complexity supported by these graphs that integrate structured data (e.g., from DBMS), semi-structured data (e.g., from documents) and human knowledge. Siemens aims for comprehensive tool support that enables engineers (i.e., domain experts) to build knowledge graphs without the help of data scientists.

3 Discussion and Conclusion

A recurrent theme in the discussions during and after the session was the strong need in both industry and academia for highly qualified data scientists with a strong understanding of the (mathematical) foundations of machine learning. This understanding is a prerequisite for solving the challenges identified in the session.

R. Reussner, A. Koziolek, R. Heinrich (Hrsg.): INFORMATIK 2020,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2021 61

Web Archive Analytics*

Michael Völske[1] Janek Bevendorff[1] Johannes Kiesel[1] Benno Stein[1]
Maik Fröbe[2] Matthias Hagen[2] Martin Potthast[3]

Abstract: Web archive analytics is the exploitation of publicly accessible web pages and their evolution for research purposes—to the extent organizationally possible for researchers. In order to better understand the complexity of this task, the first part of this paper puts the entirety of the world's captured, created, and replicated data (the "Global Datasphere") in relation to other important data sets such as the public internet and its web pages, or what is preserved thereof by the Internet Archive. Recently, the Webis research group, a network of university chairs to which the authors belong, concluded an agreement with the Internet Archive to download a substantial part of its web archive for research purposes. The second part of the paper in hand describes our infrastructure for processing this data treasure: We will eventually host around 8 PB of web archive data from the Internet Archive and Common Crawl, with the goal of supplementing existing large scale web corpora and forming a non-biased subset of the 30 PB web archive at the Internet Archive.

Keywords: Big Data Analytics; Web Archive; Internet Archive; Webis Infrastructure

1 The Global Datasphere

In the few decades since its beginnings, the World Wide Web has embedded itself into the fabric of human society at a speed only matched by few technologies that came before it, and in the process, it has given an enormous boost to scientific endeavors across all disciplines. Particularly the discipline of web science—in which the web itself, as a socio-technical system, becomes the subject of inquiry—has proven a productive field of research [KG03, Be06]. It is well understood that the vast size and sheer chaotic structure of the web pose a challenge to those that seek to study it [He08]; yet its precise dimensions are not at all easily quantified.

In what follows, we derive an informed estimate of the web's size and conclude that it represents, in fact, only a small fraction of all existing data. To this end, we start by considering the broader question of how much data there is overall. This question can be asked in two ways: First, how much data is generated continuously as a result of human activity; and secondly, how much data is persistently stored? To answer the first question is to determine (according to the International Data Corporation) the size of

* An abstract of this paper has been published in the proceedings of the Open Search Symposium 2020.

[1] Bauhaus-Universität Weimar, Germany. <first>.<last>@uni-weimar.de
[2] Martin-Luther-Universität Halle-Wittenberg, Germany. <first>.<last>@informatik.uni-halle.de
[3] Leipzig University, Germany. martin.potthast@uni-leipzig.de

doi:10.18420/inf2020_05

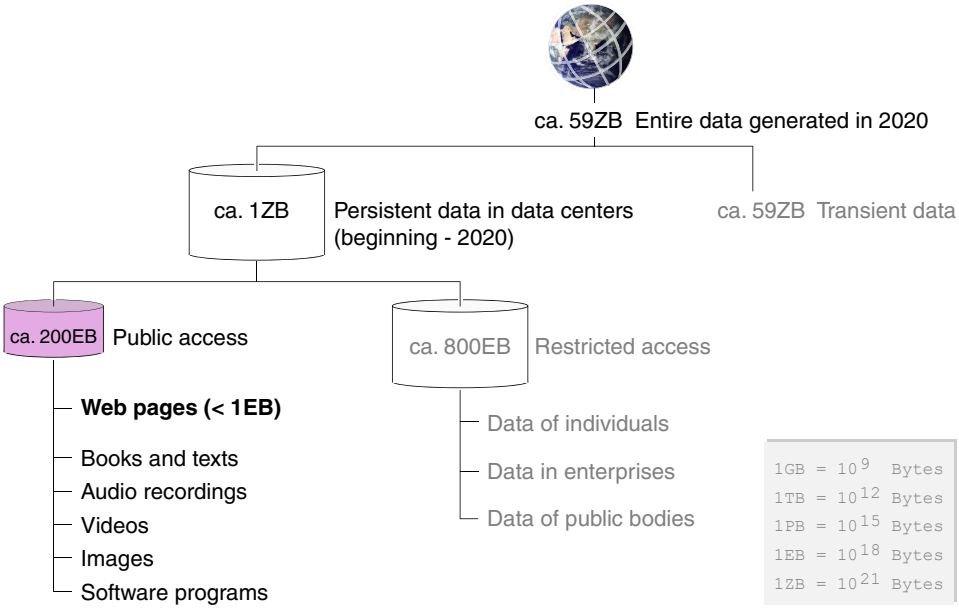


Fig. 1: Overview of the Global DataSphere in 2020. The shown numbers are estimates based on analyses and market forecasts by Cisco Systems, the International Data Corporation, and Statista Inc., among others.

the Global Datasphere, a "measure of all new data captured, created, and replicated in a single year." [RGR18]. Thanks to efforts at tracking the global market for computer storage devices, spearheaded by storage manufacturers and market researchers alike, we can make an educated guess at how much data that might be (see Figure 1 for an illustration).

Analyses like the "Data Never Sleeps" series [DI19, DI20] by business intelligence company Domo, Inc., provide pieces of the puzzle: While as recently as 2012, only a third of the world's population had internet access [DI17], that figure has risen to 59 %, or 4.5 billion people, today. Throughout 2018, the combined activity of internet users produced almost two petabytes per minute on average [DI19]. This year, in one minute of any given day, users upload 500 hours worth of video to YouTube, and more than a billion of them initiate voice or video calls [DI20]. This combined effort of internet users, however, is nowhere near the total volume of the Global Datasphere. In the same period, as automatic cameras record security footage and ATMs and stock markets transmit banking transactions, the ATLAS particle detector at CERN's Large Hadron Collider records a staggering amount of nearly four petabytes worth of particle collisions [Br11].

Past and current projections by the aforementioned IDC complete our picture of the Global Datasphere's size. Already in 2014, more than four zettabytes of data were created yearly,

and this amount has since increased at a rate of more than 150 % per year [Tu14]. The total figure had grown to 33 zettabytes by 2018 [RGR18] and was previously projected to reach 44 zettabytes by 2020, and between 163 and 175 zettabytes by 2025 [Tu14, RGR18, RGR20]. However, in the wake of the COVID-19 pandemic, the growth of the global datasphere has sharply accelerated and updated projections expect it to exceed 59 zettabytes (or $5.9 \cdot 10^{22}$ bytes) already by the end of 2020 [RRG20].

To put this number into context, consider the following thought experiment: The largest commercially available spinning hard disk drives at the time of writing hold 18 terabytes in a 3.5-inch form factor. To store the entirety of the Global Datasphere would require more than 3.2 billion such disks, which—densely packed together—would more than fill the volume of the Empire State Building and almost cost the equivalent of Apple's two-trillion-dollar market capitalization in retail price.[4] The densest SSD storage available today would reduce the volume requirements six-fold, while increasing the purchase cost by a factor of eight. However, even by the year 2025, four out of every five stored bytes are still expected to reside on rotating-platter hard drives [RGR20, RR20].

While all of the above makes the vastness of the Global Datasphere quite palpable, it leaves out one key point: Most of the 59 ZB is transient data that is never actually persistently stored. Once again, CERN's particle detectors quite impressively illustrate this: Almost all recorded particle collisions are discarded as uninteresting already in the detector, such that out of those four petabytes per minute, ATLAS outputs merely 19 gigabytes for further processing [Br11]. Analogously, most data generated by and about internet users does not become part of the permanent web, which in and of itself is only a fraction of what resides on the world's storage media [Tu14].

2 The Archived Web

Statista and Cisco Systems estimate that in the year 2020, about two zettabytes of installed storage capacity exist in datacenters across the globe [Ci20b]. Worldwide storage media shipments support this claim. About 6.5 ZB have been shipped cumulatively since 2010. Accounting for a yearly 3 % loss rate, as well as the fact that only a portion of all storage resides in enterprise facilities, leads to the same estimate of about 2–2.5 ZB installed capacity in 2020 [RGR20]. About half this capacity appears to be effectively utilized in datacenters, which means that the industry stores one zettabyte of data [Ci20a]. Adding to this the remaining 30 % share of data from consumer devices, we can estimate that the world keeps watch over a data treasure of roughly 1.43 zettabytes in total at the time of writing. Although the majority of these sextillion stored bytes is probably quite recent [Pe20], the figure does include all data stored (and retained) since the beginning of recorded history. The majority of persistent data is not publicly accessible, but hides in closed repositories belonging to individuals and private or statutory organizations. But even of what we call the "public

[4] We disregard issues like power, cooling, and redundancy, but also what should be a sizeable bulk discount.

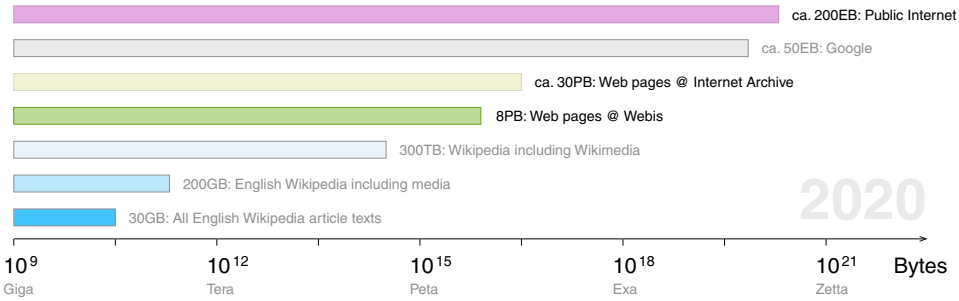


Fig. 2: The 2020 sizes of large, persistent data sets in comparison, illustrated on a logarithmic scale. The English Wikipedia is often compared to the Encyclopædia Britannica, but largely exceeds it by a factor of more than 80 at a combined size of only 30 GB. This is just over a one hundred millionth part of all textual web pages, whose total volume we estimate significantly smaller than 1 EB (cf. Fig. 1).

internet," only the small "surface web" portion—openly accessible and indexed web pages, books, audiovisual media, and software—is actually *publicly* accessible. What remains is called the "deep web," which includes email communications, instant messaging, restricted social media groups, video conferencing, online gaming, paid music and video-on-demand services, document cloud storage, productivity tools, and many more, but also the so-called "dark web." The size of the deep web is much harder to assess than that of both the surface web and the world's global storage capacity. Folklore estimates range from 20 to 550 times the size of the surface web, but, to the best of our knowledge, no reliable sources exist and previous attempts at measuring its size typically relied on estimating the population of publicly retrievable text document IDs [LL10]. In absence of reliable data, we assume its actual size to be on the larger end and take an educated guess of up to 200 exabytes. That is a fifth of all persistent data and less than half a percent of the Global Datasphere [RGR20].

The indexed surface web is substantially easier to quantify. In 2008, Google reported to have discovered one trillion distinct URLs [Go08], though only a portion of that was actually unique content worth indexing. Based on estimates of the current index sizes of the two largest web search engines [vBd16], there are approximately 60 billion pages (lower bound: 5.6 billion) in the indexed portion of the World Wide Web as of 2020.[5] The HTTP Archive,[6] which aims to track the evolution of the web from a metadata perspective, records (among other metrics) statistics about page weights, i.e., the total number of bytes transferred during page load. Based on the page weight percentiles given in their 2019 report [EH19], we estimate that the combined size of all indexed web pages is at least 1.8 petabytes counting their HTML payload alone. As a rather generous upper bound, we assert that the World Wide Web including all directly embedded style, script, and image assets is (not accounting for redundancy) most likely less than 200 petabytes and—considering the English-language

[5] https://www.worldwidewebsize.com
[6] https://httparchive.org

bias [vBd16] of the employed method—certainly less than one exabyte in size (Figure 1). As such, the surface web accounts for less than half a percent of the presumed 200 EB constituting the public internet. It is important to note that this figure only captures the *textual* part of the publicly indexable surface web. Publicly accessible on-demand video streaming, accounting for 15 % of all downstream internet traffic in 2020 [Sa20] (60 % including paid subscriptions) and necessarily taking a sizable portion of enterprises' mass storage, is not included.

Figure 2 compares the sizes of large, persistent data sets on the public internet, starting with an estimate for the total amount of data stored by Google. While Google does not publish any official figures on their data centers' storage capacity, they have maintained a public list of datacenter locations for at least the past eight years [Go13]. In 2013, thirteen datacenters were listed worldwide and contemporary estimates put the total amount of data stored by Google at 15 EB [Mu13]. By mid-2020, Google's list of datacenters had grown to 21 entries [Go20]. At the same time, the areal density of hard disk storage has approximately doubled since 2013 [Co20]. For the total amount of data stored by Google in 2020, we hence propose an updated estimate of 50 EB—this includes both public and non-public data.

The Internet Archive[7] is a large-scale, nonprofit digital library covering a wide range of media, including websites, books, audio, video, and software. In 2009, the Internet Archive's entire inventory comprised only one petabyte [JK09], but according to current statistics,[8] it has grown by a factor of almost 100 over the past decade. The Wayback Machine, an archive of the World Wide Web, forms a significant portion of the Internet Archive's collections. As of this writing, it comprises nearly 500 billion web page captures, which take up approximately 30 petabytes of storage space. In the Webis research group, we aim to store up to 8 PB of web archive data on our own premises, much of it originating from the Internet Archive, but also from other sources, such as the Common Crawl.

The Wikipedia is commonly used in large-scale text analytics tasks and thus we include it in the comparison in three ways: The combined size of all Wikipedias, including the files in the Wikimedia Commons project, comes at approximately 300 TB,[9] of which approximately 3 TB are actual wikitext.[10] The English Wikipedia including all media files is three orders of magnitude smaller, at about 200 GB.[11] Lastly, the plain wikitext content of the English Wikipedia is yet another order of magnitude smaller, adding up to merely 30 GB in 2020.[12]

Figure 3 is based on an analysis by IDC and Seagate [RGR18, RGR20] inquiring *where* the world's persistent data has been, and is going to be stored in the future: A decade ago, most of the data created across the world remained on largely disconnected consumer endpoint

[7] https://archive.org
[8] https://archive.org/~tracey/mrtg/du.html
[9] https://commons.wikimedia.org/wiki/Special:MediaStatistics
[10] https://stats.wikimedia.org/#/all-projects/content/net-bytes-difference/normal|bar|all|~total|monthly
[11] https://en.wikipedia.org/wiki/Special:MediaStatistics
[12] https://en.wikipedia.org/w/index.php?title=Wikipedia:Size_in_volumes&oldid=966973828

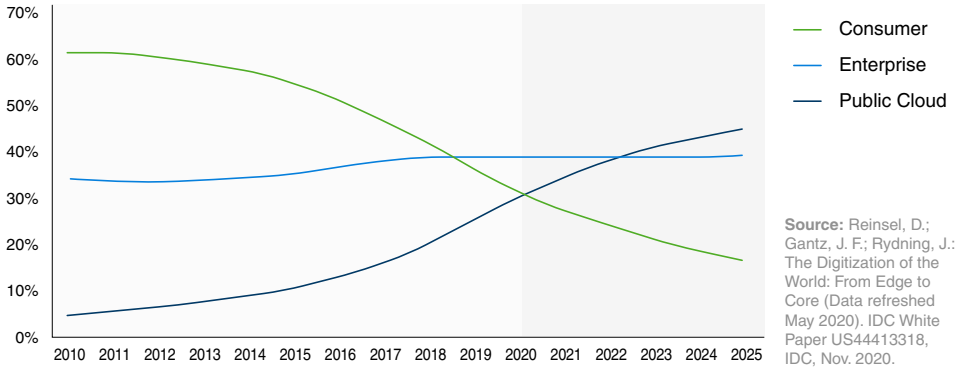


Fig. 3: As of 2020, IDC believes, more bytes are stored in the public cloud than on consumer devices. By 2022, more data is going to be stored in the cloud than in traditional datacenters.

devices, but that has steadily shifted towards public clouds. As of 2020, the share of data stored in public clouds has exceeded that stored on consumer devices, and within the next three years, the majority of persistent data is predicted to reside in public clouds.

The rapid growth in both size and publicness of web-based data sources drives interest in web science globally, not only in the Webis group. We have shown the persistent World Wide Web to be quite a bit smaller than one might expect in contrast with the totality of data created and stored; nevertheless, true "web archive analytics" tasks, which we still consider to involve datasets exceeding hundreds of terabytes, certainly do require a large and specialized hardware infrastructure to be feasible.

3 Web Archive Analytics at Webis

Figures 4 and 5 illustrate the software and infrastructure relevant to this effort: Two different computing clusters contribute to our web archive analytics efforts, out of five total currently operated by our research group. We have maintained cluster hardware in some form or another for more than a decade. Figure 4 outlines their specifications. In the following, we give a short overview of typical workloads and research areas each individual cluster is tasked with: α-web is made up of repurposed desktop hardware and enabled our earliest inroads into web mining and analytics; nowadays it still functions as a teaching and staging environment. β-web is currently our primary environment for web mining, batch and stream processing, data indexing, virtualization, and the provision of public web services. γ-web's primary purpose is the training of machine learning models, text synthesis, and language modeling. δ-web is focused on storage and serves as a scalable and redundant persistence backend for web archiving and other tasks. ε-web is used for indexing and argument search.

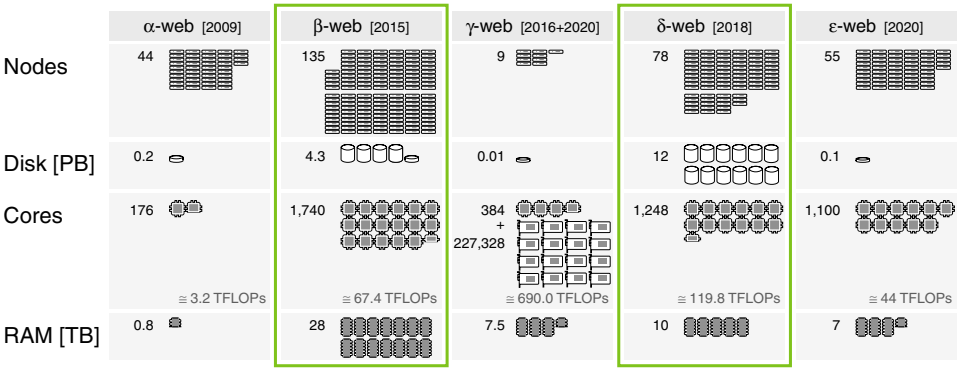


Fig. 4: The cluster hardware owned by the Webis research group, organized from left to right by acquisition date (shown in square brackets). β-web and δ-web have been specifically designed to handle large-scale web archive analytics tasks.

The β-web and δ-web clusters are the primary workhorses in our ongoing web archiving and analytics efforts. The clusters comprise 135 and 78 Dell PowerEdge servers, respectively, spread across two datacenters joined via a 400 GB/s interconnect. Each individual node is attached to one of two middle-of-row Cisco Nexus switches via a 10 GB/s link. The storage cluster maintains more than 12 PB of raw storage capacity across 1 248 physical spinning disks. The compute cluster possesses an additional 4.3 PB for serving large indexes and storing ephemeral intermediate results across 1 080 physical disks.

Our analytics stack is shown conceptually in Figure 5, beginning with the bottom-most data acquisition layer: Primary sources for data ingestion include web crawls and web archives, such as the aforementioned Internet Archive, the Common Crawl,[13] the older ClueWeb[14] datasets, as well as our own specialized crawls focused on individual research topics, such as argumentation and authorship analytics. Crowdsourcing services, such as Amazon Mechanical Turk, support and complement (distantly) supervised learning tasks.

Both the β-web and δ-web cluster are provisioned and orchestrated using the SaltStack configuration management and IT automation software, which manages the deployment and configuration of fundamental file-system- and infrastructure-level services on top of a network-booted Ubuntu Linux base image. We supervise the proper functioning of all system components with the help of a Consul cluster (for service discovery) and a redundant Prometheus deployment (for event monitoring and alerting).

The primary purpose of the δ-web cluster is to serve a distributed Ceph [We07] storage system with one object storage daemon (OSD) per physical disk, as well as five redundant

[13] https://commoncrawl.org
[14] https://lemurproject.org

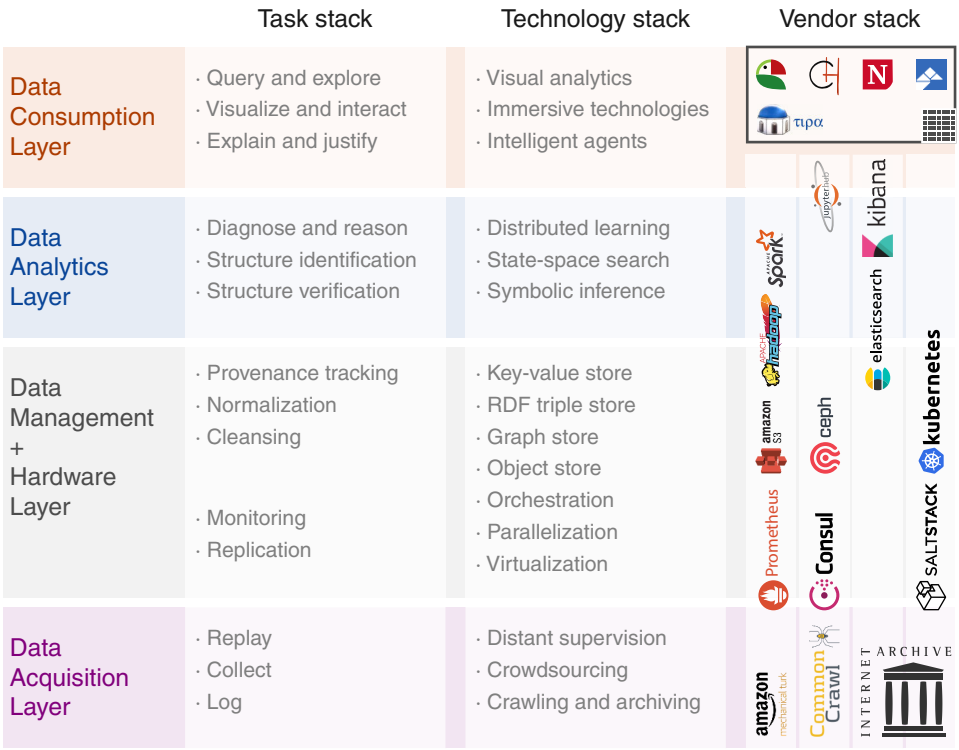


Fig. 5: The Webis web archive analytics infrastructure stack. The y-axis organizes (bottom to top) the processing pipeline from data acquisition to data consumption. The x-axis organizes (right to left) the employed tools (vendor stack), algorithms and methods (technology stack), and tackled problem classes (task stack). The box at the top of the vendor stack shows the resulting frontend services maintained by the Webis research group: args.me, ChatNoir, Netspeak, Picapica, and Tira.

Monitor (MON) / Manager (MGR) daemons. The bulk of the data payload is accessed via the RadosGW S3 API gateway service, which is provided by seven redundant RadosGW daemons and backed by an erasure-coded storage pool. Smaller and more ephemeral datasets live in a CephFS distributed file system with three active and four standby Metadata Servers (MDS).

On β-web, we maintain a Saltstack-provisioned Kubernetes cluster on top of which most of our internal and public-facing services are deployed. Kubernetes services are provided with persistent storage through a RADOS Block Device (RBD) pool in the Ceph cluster. A Hadoop Distributed File System (HDFS) holds temporary intermediate data.

Services operated on top of Kubernetes include a 130-node Elasticsearch deployment powering our search engines, as well as a suite of data analytics tools such as Hadoop Yarn, Spark, and JupyterHub. Our public-facing services are, among others, the search engines args.me [Wa17, Po19a] and ChatNoir [Be18], the Netspeak [SPT10, Po14] writing assistant, the Picapica [Ha17] text reuse detection system, and the Tira [GSB12, Po19b] evaluation-as-a-service platform. The web-archive-related services at archive.webis.de are still largely under construction. Ultimately, we envision also allowing external collaborators to process and analyze web archives on our infrastructure. As of October 2020, almost 2.3 PB of data—of which 560 TB stem from the Internet Archive and the rest from the Common Crawl—have been downloaded and are stored on our infrastructure.

4 Conclusion

The web and its ever-growing archives represent an unprecedented opportunity to study humanity's cultural output at scale. While in 2020 alone, a staggering 59 zettabytes of data will have been created throughout the year, most of that is transient, and the World Wide Web is nearly five orders of magnitude smaller. Nevertheless, the scale of the web is large enough to be challenging, and so web analytics requires potent and robust infrastructure.

The Webis cluster infrastructure has already enabled many research projects on a scale which typically only big industry players can afford due to the massive hardware requirements. This allows us to develop and maintain services such as the ChatNoir web search engine [Be18], the argument search engine args.me [Wa17], but also facilitates individual research projects which rely on utilizing the bundled computational power and storage capacity of the capable Hadoop, Kubernetes, and Ceph clusters. A few recent examples are the analysis of massive query logs [Bo20], a detailed inquiry into near-duplicates in web search [Fr20b, Fr20a], the preparation, parsing, and assembly of large NLP corpora [Be20, Ke20], and the exploration of large transformer models for argument retrieval [AP20].

With significant portions of the Internet Archive available and further hardware acquisitions in progress, many more of these research projects at a very large and truly representative scale will become feasible, including not only the evaluation, but also the efficient training of ever larger machine learning models.

Bibliography

[AP20] Akiki, Christopher; Potthast, Martin: Exploring Argument Retrieval with Transformers. In: Working Notes Papers of the CLEF 2020 Evaluation Labs. September 2020.

[Be06] Berners-Lee, Tim; Hall, Wendy; Hendler, James A.; O'Hara, Kieron; Shadbolt, Nigel; Weitzner, Daniel J.: A Framework for Web Science. Found. Trends Web Sci., 1(1):1–130, 2006.

[Be18] Bevendorff, Janek; Stein, Benno; Hagen, Matthias; Potthast, Martin: Elastic ChatNoir: Search Engine for the ClueWeb and the Common Crawl. In: 40th European Conference on IR Research (ECIR 2018). LNCS, Springer, Berlin Heidelberg New York, March 2018.

[Be20] Bevendorff, Janek; Al-Khatib, Khalid; Potthast, Martin; Stein, Benno: Crawling and Preprocessing Mailing Lists At Scale for Dialog Analysis. In: 58th Annual Meeting of the Association for Computational Linguistics. July 2020.

[Bo20] Bondarenko, Alexander; Braslavski, Pavel; Völske, Michael; Aly, Rami; Fröbe, Maik; Panchenko, Alexander; Biemann, Chris; Stein, Benno; Hagen, Matthias: Comparative Web Search Questions. In: 13th ACM International Conference on Web Search and Data Mining (WSDM 2020). February 2020.

[Br11] Brumfiel, Geoff: High-Energy Physics: Down the Petabyte Highway. Nature, 469(7330):282–283, January 2011.

[Ci20a] Cisco Systems: Amount of data actually stored in data centers worldwide from 2015 to 2021. Statistics, Statista Inc., New York, March 2020.

[Ci20b] Cisco Systems: Data center storage capacity worldwide from 2016 to 2021, by segment. Statistics, Statista Inc., New York, April 2020.

[Co20] Coughlin, Tom: HDD Market History and Projections. Report, May 2020. https://www.forbes.com/sites/tomcoughlin/2020/05/29/hdd-market-history-and-projections/.

[DI17] Domo Inc., American Fork, Utah: Data Never Sleeps 5.0. Leaflet, November 2017. https://www.domo.com/learn/data-never-sleeps-5.

[DI19] Domo Inc., American Fork, Utah: Data Never Sleeps 7.0. Leaflet, August 2019. https://www.domo.com/learn/data-never-sleeps-7.

[DI20] Domo Inc., American Fork, Utah: Data Never Sleeps 8.0. Leaflet, August 2020. https://www.domo.com/learn/data-never-sleeps-8.

[EH19] Everts, Tammy; Hempenius, Katie: The 2019 Web Almanac, Chapter 18: Page Weight. Technical report, The HTTP Archive, 2019. https://almanac.httparchive.org/en/2019/page-weight.

[Fr20a] Fröbe, Maik; Bevendorff, Janek; Reimer, Jan Heinrich; Potthast, Martin; Hagen, Matthias: Sampling Bias Due to Near-Duplicates in Learning to Rank. In: 43rd International ACM Conference on Research and Development in Information Retrieval (SIGIR 2020). July 2020.

[Fr20b] Fröbe, Maik; Bittner, Jan Philipp; Potthast, Martin; Hagen, Matthias: The Effect of Content-Equivalent Near-Duplicates on the Evaluation of Search Engines. In: Advances in Information Retrieval. 42nd European Conference on IR Research (ECIR 2020). volume 12036 of Lecture Notes in Computer Science, Springer, Berlin Heidelberg New York, April 2020.

[Go08] Google, LLC: We knew the web was big.... Web page, August 2008. https://web.archive.org/web/20200918111648/https://googleblog.blogspot.com/2008/07/we-knew-web-was-big.html.

[Go13] Google, LLC: Data Center Locations. Web page, August 2013. https://web.archive.org/web/20130801052830/http://www.google.com/about/datacenters/inside/locations/index.html.

[Go20] Google, LLC: Discover our Data Center Locations. Web page, July 2020. https://web.archive.org/web/20200718223350/https://www.google.com/about/datacenters/locations/index.html.

[GSB12] Gollub, Tim; Stein, Benno; Burrows, Steven: Ousting Ivory Tower Research: Towards a Web Framework for Providing Experiments as a Service. In: 35th International ACM Conference on Research and Development in Information Retrieval (SIGIR 2012). August 2012.

[Ha17] Hagen, Matthias; Potthast, Martin; Adineh, Payam; Fatehifar, Ehsan; Stein, Benno: Source Retrieval for Web-Scale Text Reuse Detection. In: 26th ACM International Conference on Information and Knowledge Management (CIKM 2017). November 2017.

[He08] Hendler, James A.; Shadbolt, Nigel; Hall, Wendy; Berners-Lee, Tim; Weitzner, Daniel J.: Web science: an interdisciplinary approach to understanding the web. Commun. ACM, 51(7):60–69, 2008.

[JK09] Jaffe, Elliot; Kirkpatrick, Scott: Architecture of the Internet Archive. In: Proceedings of SYSTOR 2009: The Israeli Experimental Systems Conference. SYSTOR '09, Association for Computing Machinery, New York, NY, USA, 2009.

[Ke20] Kestemont, Mike; Manjavacas, Enrique; Markov, Ilia; Bevendorff, Janek; Wiegmann, Matti; Stamatatos, Efstathios; Potthast, Martin; Stein, Benno: Overview of the Cross-Domain Authorship Verification Task at PAN 2020. In: Working Notes Papers of the CLEF 2020 Evaluation Labs. September 2020.

[KG03] Kilgarriff, Adam; Grefenstette, Gregory: Introduction to the Special Issue on the Web as Corpus. Computational Linguistics, 29(3):333–348, 2003.

[LL10] Lu, Jianguo; Li, Dingding: Estimating deep web data source size by capture-recapture method. Inf. Retr., 13(1):70–95, 2010.

[Mu13] Munroe, Randall: Google's Datacenters on Punch Cards. Web page, September 2013. https://web.archive.org/web/20130920004219/http://what-if.xkcd.com:80/63/.

[Pe20] Petrov, Christo: 25+ Impressive Big Data Statistics for 2020. Internet blog, techjury.net., September 2020. https://techjury.net/blog/big-data-statistics/.

[Po14] Potthast, Martin; Hagen, Matthias; Beyer, Anna; Stein, Benno: Improving Cloze Test Performance of Language Learners Using Web N-Grams. In: 25th International Conference on Computational Linguistics (COLING 2014). August 2014.

[Po19a] Potthast, Martin; Gienapp, Lukas; Euchner, Florian; Heilenkötter, Nick; Weidmann, Nico; Wachsmuth, Henning; Stein, Benno; Hagen, Matthias: Argument Search: Assessing Argument Relevance. In: 42nd International ACM Conference on Research and Development in Information Retrieval (SIGIR 2019). July 2019.

[Po19b] Potthast, Martin; Gollub, Tim; Wiegmann, Matti; Stein, Benno: TIRA Integrated Research Architecture. In: Information Retrieval Evaluation in a Changing World, The Information Retrieval Series. Springer, Berlin Heidelberg New York, September 2019.

[RGR18] Reinsel, David; Gantz, John F.; Rydning, John: The Digitization of the World: From Edge to Core. IDC White Paper US44413318, International Data Corporation (IDC), Framingham, Massachusetts, November 2018.

[RGR20] Reinsel, David; Gantz, John F.; Rydning, John: The Digitization of the World: From Edge to Core (Data refreshed May 2020). IDC White Paper US44413318, International Data Corporation (IDC), Framingham, Massachusetts, May 2020.

[RR20] Reinsel, David; Rydning, John: Worldwide Global StorageSphere Forecast, 2020–2024: Continuing to Store More in the Core. IDC Market Forecast US46224920, International Data Corporation (IDC), Framingham, Massachusetts, May 2020.

[RRG20] Reinsel, David; Rydning, John; Gantz, John F.: Worldwide Global StorageSphere Forecast, 2020–2024: The COVID-19 Data Bump and the Future of Data Growth. IDC Market Forecast US44797920, International Data Corporation (IDC), Framingham, Massachusetts, April 2020.

[Sa20] Sandvine: The Global Internet Phenomena Report 2020. COVID-19 Spotlight. Technical report, Sandvine Incorporated, May 2020.

[SPT10] Stein, Benno; Potthast, Martin; Trenkmann, Martin: Retrieving Customary Web Language to Assist Writers. In: Advances in Information Retrieval. 32nd European Conference on Information Retrieval (ECIR 2010). Springer, Berlin Heidelberg New York, March 2010.

[Tu14] Turner, Vernon; Gantz, John F.; Reinsel, David; Minton, Stephen: The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things. IDC White Paper 1672, International Data Corporation (IDC), Framingham, Massachusetts, April 2014.

[vBd16] van den Bosch, Antal; Bogers, Toine; de Kunder, Maurice: Estimating Search Engine Index Size Variability: A 9-Year Longitudinal Study. Scientometrics, 107(2):839–856, May 2016.

[Wa17] Wachsmuth, Henning; Potthast, Martin; Al-Khatib, Khalid; Ajjour, Yamen; Puschmann, Jana; Qu, Jiani; Dorsch, Jonas; Morari, Viorel; Bevendorff, Janek; Stein, Benno: Building an Argument Search Engine for the Web. In: 4th Workshop on Argument Mining (ArgMining 2017) at EMNLP. Association for Computational Linguistics, September 2017.

[We07] Weil, Sage A.; Leung, Andrew W.; Brandt, Scott A.; Maltzahn, Carlos: RADOS: a scalable, reliable storage service for petabyte-scale storage clusters. In: Proceedings of the 2nd International Petascale Data Storage Workshop (PDSW '07), November 11, 2007, Reno, Nevada, USA. pp. 35–44, 2007.

Mobilitätssysteme

Mobilitätssysteme – von Daten, Schnittstellen und Modellen

Politische Weichenstellungen für Digitalisierung in der Mobilität, Entwicklung zukünftiger ÖPNV-Systeme und Entwicklungen in der Simulation

Thomas Schlegel [1]

Abstract: Die Session Mobilitätssysteme auf der Informatik 2020 vereint drei Beiträge zum Themenfeld Mobilitätssysteme. Informationstechnologien, Systeme und mobilitätsspezifische Entwicklungen wurden im Hinblick auf Mobilitätsdaten, Schnittstellen in Mobilitätssystemen (im Schwerpunkt Öffentlicher Personenverkehr) und modellbasierte Simulation untersucht und vorgestellt. Dabei spielt der Systemgedanke eine Rolle, der Daten, Schnittstellen und Modelle für Systeme vereint. Dieser Beitrag gibt die Inhalte von Vortrag und Diskussion für die Fachwelt aber auch die interessierte Öffentlichkeit wieder.

Keywords: Mobilitätssysteme; Mobilitätsdaten; Datenakquise; Datenqualität; Datenplattformen; Baden-Württemberg; Verkehr; Schnittstellen; Öffentlicher Personenverkehr; Verkehrssimulation; Mobilitäts-Informationstechnologien

1 Einleitung

Die Session Mobilitätssysteme lieferte einen spannenden Überblick zu sowie detaillierte Aspekte aus der Praxis und Forschung zu Mobilitätssystemen in der Informatik. Der Schwerpunkt der Session lag dabei auf Informationstechnologien, Systemen und mobilitätsspezifischen Entwicklungen für den nicht-motorisierten und öffentlichen Verkehr und betrachtete das Thema der Mobilitätssysteme nicht isoliert an einzelnen Fahrzeugen sondern aus den Perspektiven von Nutzern, Entwicklern und Politik. Dabei spielen der Systemgedanke und damit Daten, Schnittstellen, Modelle und Systeme eine große Rolle. Drei Vorträge und die zugehörige Diskussion zeigten, dass Datenverfügbarkeit (Schnittstellen, Plattformen, Integration), domänenspezifische Daten und Systeme (hier im Öffentlichen Personenverkehr) und Verkehrssimulation Teile des komplexen Themenfelds der datengetriebenen Mobilitätssysteme sind: Informationen in und aus Mobilitätssystemen.

[1] Hochschule Karlsruhe, Institut für Ubiquitäre Mobilitätssysteme, Moltkestr. 30, 76133 Karlsruhe, Thomas.Schlegel@hs-karlsruhe.de, https://orcid.org/0000-0003-3339-3720

doi:10.18420/inf2020_06

2 Mobilitätsdaten: Mehrwert und Herausforderung

„Mobilitätsdaten zugänglich machen und sinnvoll nutzen: was kann ein Bundesland dafür tun?“ fragte Dr. Wolf Engelbach (Ministerium für Verkehr Baden-Württemberg). Sein Vortrag handelte von Intermodalität und Steuerung als Beitrag zum Klimaschutz im Verkehr und von verkehrspolitischen Themen aber auch von Informatik für kommunikative und informationstechnische Datenbündelung bis hin zu Data Governance und Datenqualität.

Um intelligente Mobilitätssysteme und neue Mobilitätsdienste zu entwickeln – aber auch um die Qualität zu erhöhen, sind Daten eine zwingende Voraussetzung. Dabei spielen zunächst vor allem technische Aspekte eine Rolle.

Jedoch rücken auch zunehmend die Abstimmung zwischen den Beteiligten, rechtliche Hintergründe und die Zugänglichkeit von Daten aus organisatorischer Sicht in den Fokus. Hier spielt gerade auch die Politik sowohl durch die Setzung von Standards und Vorschriften, als auch durch einen förderpolitischen Rahmen eine tragende Rolle.

Der diesen folgende, ebenso wichtige Punkt, ist die Nutzung dieser Daten. Gerade für Mobilitätsdaten existiert häufig keine zentrale Stelle, die diese erfassen und zur Verfügung stellen kann. Vielmehr erzeugen sowohl Verkehrsträger als auch mobile Menschen Daten dezentral in diversen Systemen und in unterschiedlichen Formaten und Qualitätsstufen.

Im Vortrag wurde gezeigt, was ein Bundesland hierfür tun kann. Verkehrsinformationen sind im Land bereits an vielen Stellen vorhanden. So sind Verkehrszustände wie Stau oder Baustellen verfügbar und werden bereits über eine App des Landes Baden Württemberg der Bevölkerung zur Verfügung gestellt.

Auch im Öffentlichen Verkehr werden bereits seit langer Zeit Daten zur Verfügung gestellt, seit 2019 auch Sollfahrplan-Daten aus ganz Baden-Württemberg als Open Data. Auch aktive Systeme wie der Ticketverkauf über E-Ticketing oder das European Train Control System (ETCS) tragen hierzu bei. Neue Systeme entstehen auch durch die Förderlinie MobiArch des Landes.

Gerade die öffentlichen Stellen, wie auch die Länder, haben Open Data mittlerweile häufig als Leitprinzip identifiziert. Dabei besteht die Hoffnung, wenn Daten offen zur Verfügung gestellt werden, die Entwicklung innovative Applikationen und Anwendungen durch andere Stakeholder, vor allem aus der Wirtschaft, zu ermöglichen. Hier kann die öffentliche Hand nicht flächendeckend für alle notwendigen Anwendungsfelder Applikationen schaffen.

Der Aufbau von MobiData BW als verkehrsträgerübergreifendes Portal für Daten und digitale Dienste im Mobilitätsbereich von Baden-Württemberg umfasst neben der IT-Infrastruktur und dem Betrieb durch die Nahverkehrsgesellschaft von Baden-Württemberg (NVBW) auch ein aktives Partnermanagement und Impulse für Innovationen.

Beispielhafte Anwendungsfälle umfassen

- Datensätze für lokales intermodales Routing
- Digitale Einbettung von Parkraum
- Mobilitätsdaten für Planung und Evaluation
- Lokale Innovationen
- Mobilitätsspezifischen aus der Sharing-Welt bündeln

Gerade Sharing-Angebote und deren Verknüpfung profitieren von Daten und Datenintegration.

Im öffentlichen Bereich sind von EU über Bund und Länder bis hin zu den Regionen und Kommunen unterschiedliche Stakeholder beteiligt, die für die Entwicklung und Etablierung Standards gut zusammenarbeiten müssen.

Trotzdem bleiben die Integration und die Verknüpfung von bestehenden Daten weiterhin eine große Herausforderung sowohl in der Forschung als auch in der praktischen Realisierung. Dabei helfen die Daten, den Verkehr besser zu organisieren und damit auch den Weg hin zu einer klimafreundlicheren Mobilität zu bereiten. Hierzu tragen auch Intermodalitäts-Aspekte, also die integrierte und verkettete Nutzung von unterschiedlichen Verkehrsmitteln, zu Erreichung des Ziels bei.

Diskutiert wurde hier im Rahmen der Session eine Incentivierung für die Bereitstellung von Daten und auch die Harmonisierung. So sollen Daten einfacher ausgetauscht werden können, indem diese in einem standardisierten Format und über eine Plattform bereitgestellt werden. Diese Plattform könnte dabei auch zum Teil Daten nur durchleiten und damit über eine einheitliche Schnittstelle zur Verfügung stellen, die eine dezentrale Bereitstellung für einen zentralen Zugriff öffnet.

Erforderlich ist eine faire Nutzung von offenen Daten, also eigene Daten im Sinne von Open Source und Open Data auch wieder offen zur Verfügung zu stellen, im Gegenzug für eine freie Nutzung fremder Daten. Dies stellt allerdings den Gegenpol zu „Daten als Öl des einundzwanzigsten Jahrhunderts“ dar und zeigt, dass es nicht nur um technische, sondern vielfach eher politische, rechtliche und gesellschaftliche Fragen geht.

Auch das Thema der Provenance, also der Verlässlichkeit von Daten, und auch die Sicherung von Datenqualität bleibt weiterhin eine Herausforderung, gerade in Zeiten von Open Data.

3 Herausforderungen bei der Einführung von Informationstechnologie im öffentlichen Verkehr

Dirk Weißer (INIT SE) legte den Schwerpunkt dann auf den öffentlichen Verkehr und die entsprechenden Systeme und Technologien. Er zeigte zum einen wie schwer es ist, aktuelle und verlässliche Daten aus den bestehenden Systemen zu erhalten, zum anderen aber auch praktische und aktuelle Entwicklungen in diesen Systemen von Leitstellen mit ITCS bis zu Bordsystemen.

Der öffentliche Verkehr ist eine etablierte und sicherheitsorientierte Branche, bei der Kontinuität eine tragende Rolle spielt. Dies hat seine Ursachen auch in der langen Nutzungszeit und Langlebigkeit von Systemen, wo beispielsweise Schienenfahrzeuge 30 Jahre und mehr eingesetzt werden. Fahrzeuge aus dem Jahre 1985 sind hier keine Seltenheit. Dies bedingt stärker evolutionäre denn revolutionäre Vorgehensweisen, auch im Bereich der damit verbundenen Informationstechnologie. Dabei ist der öffentliche Verkehr Teil der Daseinsvorsorge, d.h. dass öffentlicher Verkehr durch die öffentliche Hand ermöglicht wird und damit stabil funktioniert, gleichzeitig finanziell allerdings keinen Gewinn erzeugt. Damit sind zur Finanzierung kaum Möglichkeiten gegeben, größere Budgets für Technologiethemen und umfassende Neuerungen im Bereich der Systeme einzusetzen.

Historisch gewachsene und monolithische Systeme machen es dabei schwierig, neue Technologien, die häufig auch in kurzen Zyklen auftauchen, in die Landschaft zu integrieren. Eine Vielzahl von Leitungen, Protokollen und Komponenten unterschiedlicher Funktionsbereiche muss dabei integriert werden. Ein positives Beispiel stellt das nationale Forschungs- und Standardisierungsprojekt IP-KOM-ÖV dar, das moderne, IP-basierte Kommunikation und eine dienstbasierte Architektur für Fahrzeuge des öffentlichen Personenverkehrs standardisiert hat. Allerdings ist eine nachträgliche Verkabelung wirtschaftlich für Fahrzeuge kaum darstellbar. Daher wird für eine Migration häufig nach Lösungen basierend auf existierender Infrastruktur gesucht. Im Bereich der Busse sind die Zyklen deutlich kürzer und auch im Bahnbereich ist eine evolutionäre Migration zu beobachten. Allerdings sind hier nur erprobte Technologien akzeptiert, wobei gleichzeitig die Erprobung von der Möglichkeit des Einsatzes abhängt. Bei Bordsystemen wie CAN oder Funk-Systemen besteht zudem häufig nicht die Möglichkeit spezifische Systeme zu schaffen. Vielmehr muss Rücksicht auf existierende Technologien und Anwendungen in anderen Branchen genommen werden.

Im Bereich der Hintergrundsysteme ist allerdings aktuelle eine Entwicklung hin zu neuen und stärker modularen Architekturen und einer dynamischen Verknüpfung über Broker-Architekturen zu beobachten, die sich auch in manchen größeren Verbünden bereits entwickeln. Auch in diesem Bereich stellt sich wieder die Frage nach Schnittstellenstandardisierung und Datenaustausch bis hin zu Open Data oder zumindest der Integration von Datenflüssen über unterschiedliche Hersteller und Domänen hinweg.

Fragestellungen, die auch in der angewandten Forschung noch nicht vollständig beantwortet sind, beginnen bereits bei der Bezeichnung von Objekten und der Beschreibung von Inhalten.

Hier können semantische Technologien und die Standardisierung weiterer Beschreibungs- und Kommunikationsschichten Beiträge leisten.

Heute werden häufig Schnittstellen nur direkt zwischen zwei beteiligten Partnern spezifiziert, die auch nur das Wissen von heute abbilden. Auch bei Broker-Architekturen ist weiterhin der Datenkonsument verantwortlich für die Nutzbarmachung der Daten, obwohl eine Annotation an der Quelle deutlich mehr Möglichkeiten bietet. Wie immer bei der Standardisierung ist dabei die flächendeckende Nutzung bzw. die Akzeptanz entscheidend.

Eine Empfehlung an Entscheidungsträger und vor allem für Auftraggeber lautet hier eine nachhaltige und durchdachte Entwicklung voranzutreiben. Häufig ist zu beobachten, dass sehr kurzzyklische Programme gefahren werden, die pragmatisch entstandene Lösungen fördern, jedoch keine Harmonisierung und innovative Weiterentwicklung ermöglichen.

Ein in der Diskussion angesprochenes Thema ist die Einführung von 5G-Funktechnologien im öffentlichen Personenverkehr. Während dies auf Herstellerseite viele Probleme löst, ist aus den oben genannten Gründen eine schnelle und flächendeckende Umsetzung eher unwahrscheinlich. Durch schnelle Mobilfunktechnologien ist dabei zukünftige eine deutlich bessere Zeitauflösung der Daten und auch steigende Datenqualität zu erwarten, gegebenenfalls auch mit bisheriger Funktechnologie als Fallback-Lösung.

4 Erkenntnisse aus 50 Jahren Verkehrsfluss-Simulation, Entwicklungen aus Karlsruhe

In der Verkehrstechnik, d.h. bei der Planung von Straßen, Kreuzungen, Ampeln etc., wird schon seit Jahrzehnten die Simulation des Verkehrsablaufs eingesetzt, um zu überprüfen, ob die Verkehrsanlage wie geplant funktionieren wird. Die weltweit am häufigsten eingesetzte Software dafür kommt aus Karlsruhe und ihre wissenschaftlichen Wurzeln reichen 50 Jahre zurück. Prof. Dr. Peter Vortisch (KIT) gab daher entlang dieser historischen Entwicklung einen Einblick in die Modelle hinter der Verkehrssimulation und in die typischen und zukünftigen Anwendungsgebiete.

Wo für Fragestellungen zu künftigen Systemen oder mittels noch nicht verfügbarer Daten Entscheidungen getroffen werden müssen, kommt die Verkehrssimulation ins Spiel. Diese hat in Forschung und Praxis im Raum Karlsruhe eine lange Tradition.

Bei der Simulation von Knotenpunkten und räumlich eng begrenzten Fragestellungen wird häufig auf vergleichsweise detailreiche und wenig abstrakte Simulationen zurückgegriffen. Diese sind meist im Bereich von 10 bis 20 Millisekunden aufgelöst und ermöglichen daher eine kontinuierliche Betrachtung von Entwicklungen nahe an der Realität für die Verkehrstechnik.

Die Verkehrstechnik betrachtet dabei beispielsweise Signalanlagen und damit Kommunikationssituationen, beispielsweise an Kreuzungen. So können Leistungsfähigkeitsnachweise

bereits vorab in der Verkehrstechnik erbracht werden. Dies gilt beispielsweise in ähnlicher Weise auch für die Absicherung der verkehrstechnischen Gestaltung an Autobahnen.

Ein Anwendungsbereich ist auch die intermodale Verknüpfung, die sowohl im Bereich der Datenintegration, wie oben beschrieben, als auch für die konkrete Gestaltung wichtige Aussagen treffen kann. So werden beispielsweise Fußverkehrssimulationen für den öffentlichen Verkehr in großen Umsteigegebäuden als auch beispielsweise für Evakuierungen durchgeführt.

4.1 Verhaltensmodelle für die Verkehrssimulation

Im Bereich der Kraftfahrzeuge werden beispielsweise drei unterschiedliche Ebenen für das war verhalten betrachtet. Auf der **strategischen** Ebene geht es um die Routenwahl im gesamten Netz, wogegen auf der **taktischen** Ebene Fahrermanöver stattfinden und die Kooperation zwischen Verkehrsteilnehmern betrachtet wird, beispielsweise beim Einfahren auf eine gestaute Autobahn.

Operational werden beispielsweise Unfälle verhindert, indem Abstände eingehalten oder Fahrerstreifen gewechselt werden. Dies bezieht nur das direkte Umfeld und einen kleinen Zeitraum in Sekundenbereich ein. So wird Geschwindigkeit und Abstand für die Regelung eingesetzt: Zu geringer Abstand führt zu einer Gefahrenbremsung, passender Abstand wird beim Folgen eingehalten und dabei frei gefahren. Bei Annährung wird dann die Geschwindigkeit reduziert.

Wie die realen Menschen, passen auch die simulierten Menschen auf der taktischen Ebene ihre Routenwahl anhand von Erfahrungen an, sodass Reisezeiten etc. zu einer Anpassung über das Netz führen.

Beim Fußverkehr werden beispielsweise Social-Force-Modelle eingesetzt, die das Verhalten der einzelnen Fußgänger anhand von Abstoßung modellieren. Durch Versuche mit realen Fußgängern können Modelle anhand der Empirie dann kalibriert werden.

Modularisierung und nutzergenerierte Modelle führten Anfang der 1980er Jahre zu einer Verbreitung der Ansätze in der Praxis, auch über die Simulationssoftware der PTV mit VISSIM Anfang der 1990er Jahre.

Heute entwickeln sich die Simulationswerkzeug immer stärker in Richtung realem Verhalten im Verkehr, die je nach Kultur auch unterschiedliche Modelle erfordern. Diese integrieren beispielsweise anderes Längs- und Querverhalten, wie geringere Spur-Orientierung in Ländern wie Indien, stärker. Auch die Kopplung der Simulation mit aktuellen Fahrzeugtechnologien wie Elektroantrieben „in-the-loop“ zeigt Entwicklungspfade auf.

Akzeptanzmodelle ermöglichen auch „echtes“ Verhalten zu modellieren, beispielsweise anhand bekannter Einhaltungsquoten von dynamischen Tempolimits oder Wegeempfehlungen, so dass die Simulation immer näher an die reale Praxis rückt.

5 Fazit

Es zeigt sich, dass beim Thema Daten und Modelle in Mobilitätsystemen Forschung, Politik und Unternehmen sowohl voneinander profitieren können als auch voneinander abhängig sind. Das Thema Interoperabilität im Datenbereich spielt hier bei allen Stakeholdern eine große Rolle und wird auch aus dem Dreiklang der Vorträge und in der Diskussion deutlich: Simulation „in-the-loop“ – also mit Echtzeitdaten verknüpfte Simulation ist schon heute in manchen Bereichen möglich, während gleichzeitig noch immer die Datenintegration selbst in einzelnen Domänen der Mobilitätssysteme eine Herausforderung für Forschung und Praxis darstellt. Offene Daten, interoperable Systeme und besser Modelle zeigen hier Entwicklungspfade in Forschung und Praxis der Mobilitätsinformatik und der Mobilitätsysteme.

MobiData BW:

Mobilitätsdaten für Baden-Württemberg[1]

Dr. Wolf Engelbach[2], Dr. Christian Förster[2]

Abstract: MobiData BW bündelt als Plattform Daten und digitale Dienste für die verkehrsträgerübergreifende Mobilität in Baden-Württemberg. Städte, Gemeinden und Landkreise aber auch privatwirtschaftliche Akteure, etwa aus der Parkraumbewirtschaftung oder dem Feld der Sharing-Dienste, können sowohl Datengeber als auch Datennehmer sein und damit aktiv zur Gestaltung nachhaltiger Mobilität beitragen. Mobilitätsdaten auf MobiData BW stehen unter einer offenen Lizenz für kommerzielle wie auch nicht-kommerzielle Anschlussanwendungen zur Verfügung. Initiator der Plattform ist das Ministerium für Verkehr Baden-Württemberg, der Betrieb liegt bei der Nahverkehrsgesellschaft Baden-Württemberg NVBW.

Keywords: MobiData BW; Mobilitätsdaten; Open Data; NVBW; ÖPNV; Sharing-Dienste

1 Digitale Daten als Schlüssel zur Gestaltung neuer Mobilität

Von öffentlich zugänglichen Mobilitätsdaten sollen in Baden-Württemberg nach dem Willen der Landesregierung künftig noch weit mehr als bislang Verwaltungen, private NutzerInnen und Unternehmen profitieren. MobiData BW liefert einen weiteren Beitrag zur Transformation des Verkehrssystems im urbanen und ländlichen Raum. Diese Plattform für Mobilitätsdaten ist intermodal ausgerichtet. Sie orientiert sich am Paradigma der nachhaltigen Mobilität und macht sich das Open Data Prinzip zur Handlungsgrundlage.

Mobilität ist wichtig für eine funktionierende Gesellschaft, wird aber durch ihre Menge zu einer Herausforderung für lebenswerte Städte und Landkreise. Nicht zuletzt erfordern gesamtgesellschaftliche Aufgaben wie der Klimaschutz eine nachhaltige Gestaltung des Verkehrs durch eine Reduktion von verkehrsinduzierten Treibhausgasen. Die geforderte Transformation des Verkehrssystems ist sowohl in globalen Klimaabkommen (z.B. Klimaankommen von Paris 2016) als auch im Klimaschutzplan 2050 der Bundesregierung (BMU (Hrsg.), 2019) festgehalten.

[1] Die Autoren danken Christoph Meider sowie den Kolleginnen und Kollegen bei der Nahverkehrsgesellschaft Baden-Württemberg für die Unterstützung bei der Erstellung dieses Textes.

[2] Ministerium für Verkehr Baden-Württemberg, Dorotheenstraße 8, 70173 Stuttgart, wolf.engelbach@vm.bwl.de, christian.foerster@vm.bwl.de

doi:10.18420/inf2020_07

Sie wird im Klimaschutzgesetz und dem Integrierten Energie- und Klimaschutzkonzeptes für das Land Baden-Württemberg konkretisiert und hierin wird dem gezielten Umgang mit Mobilitätsdaten eine wichtige Rolle zugeschrieben.

Richtig ausgestaltet kann also die Digitalisierung dazu beitragen, Mobilitätsbedürfnisse mit weniger Verkehr zu befriedigen. Die neue Broschüre „Digitale Mobilität" stellt Anwendungen und aktuelle Projekte vor, welche die digitale und nachhaltige Mobilität in Baden-Württemberg entwickeln und umsetzen (Ministerium für Verkehr Baden-Württemberg, 2020).[3] Die öffentliche Hand kann auf Basis von Daten den Verkehr gezielt nach gemeinwohlorientierten Zielvorstellungen beeinflussen. Datengestützte Anwendungen können im Bereich des öffentlichen Verkehrs durch eine verbesserte Ticketbuchung und Reiseauskunft den Nutzungskomfort für Fahrgäste erhöhen und zu einem Umstieg auf öffentliche Verkehrsmittel motivieren. (Staatsministerium Baden-Württemberg, 2019). Zudem verbinden sich mit offenen Mobilitätsdaten Chancen für Innovationen in der Gründungsökonomie, aber auch Vorteile für Forschung, Analyse und Verkehrsplanung.

Dank einer zunehmenden Vernetzung und der damit verbundenen Möglichkeit zur schnellen und permanenten Datenverarbeitung, gerade auch im Verkehrssektor, stellt die Digitalisierung die Grundlage für eine neue Generation von Mobilität dar. Für eine flexible, nachhaltige und intelligent vernetzte Mobilität ist die Zusammenführung und Bereitstellung von mobilitätsrelevanten Daten des Straßenverkehrs mit Informationen aus dem Bereich des öffentlichen Verkehrs und der neuen Mobilität (z.B. Sharing-Dienste) wesentlich, um Mobilitätsdaten zu einer Basis für die Erreichung kommunaler Verkehrs- und Klimaschutzziele werden zu lassen.

Im Auftrag des Verkehrsministeriums betreibt die NVBW (Nahverkehrsgesellschaft Baden-Württemberg mbH) daher zusammen mit Partnern wie der Landesstelle für Straßentechnik seit September 2020 die Mobilitätsdatenplattform MobiData BW (www.mobidata-bw.de). Bereits im Rahmen des im Frühjahr 2019 beendeten Innovationsprojektes moveBW[4] hatte ein Konsortium aus verschiedenen Industriepartnern ein Zusammenspiel von Daten und IT-Komponenten entworfen, das die verkehrsträgerübergreifende Bereitstellung von Mobilitätsdaten ermöglicht. Mit MobiData BW setzt nun das Land diesen Weg fort. Als Bestandteil dieser umfassenden Mobilitätsdatenarchitektur dient MobiData BW als Instanz zur Bündelung und Verfügbarmachung mobilitätsbezogener Daten.[5]

2 Ausrichtung von MobiData BW

Die fortschreitende Digitalisierung produziert große Datenmengen, die auch im Mobilitätsbereich oft in Datensilos vorliegen, allerdings weisen sie bislang mangelnde Interoperabilität

[3] https://vm.baden-wuerttemberg.de/fileadmin/redaktion/m-mvi/intern/Dateien/Broschüren/200625_MfV_Bro_Digitale_Mobilität_A4_56S_WEB.pdf
[4] https://vm.baden-wuerttemberg.de/de/politik-zukunft/zukunftskonzepte/digitale-mobilitaet/movebw/
[5] https://www.nvbw.de/aufgaben/digitale-mobilitaet/mobidata-bw/

zueinander auf. Mobilitätsanbieter wie Sharing-Dienstleister und Verkehrsverbünde verfügen beispielsweise über Daten zu Standorten, Fahrplan- und/oder Echtzeitdaten sowie dem Nutzungsverhalten durch Mobilitätskunden. Öffentliche Institutionen wie Städte, Gemeinden und Landkreise, bieten zum Beispiel relevante Datensätze in Bezug auf Einflussfaktoren des Verkehrs (z.B. Baustellen, Umleitungen, Messquerschnitte der Straßen), Infrastruktur (z.B. Radwege, Straßenqualität, Abstellplätze, barrierefreie Zugänge) sowie zu Stellplätzen und zentralen Orten einer Kommune (z.B. öffentliche Points of Interest) (BMVI (Hrsg.), 2017, S. 41ff).

Die bestehende Fragmentierung der Mobilitätsdatenbestände führt dazu, dass Chancen der Digitalisierung ungenutzt bleiben. Um die Digitalisierung der Mobilität nicht als Selbstzweck zu verstehen, sondern als Mittel, um Mobilität nachhaltig zu verbessern, ist die einfache Verfügbarkeit sowie die gemeinsame Nutzbarkeit der Datenbestände zentral. Um eine flexible und passgenau auf die Kriterien einer nachhaltigen, intelligent vernetzten Mobilität der Zukunft zu gewährleisten, spielt die Bündelung der Daten in öffentlicher Hand eine Schlüsselrolle (Staatsministerium Baden-Württemberg, 2019, S. 47; Ministerium für Verkehr Baden-Württemberg, 2018, S. 46f).

Datengeber sind Landeseinrichtungen, Städte, Gemeinden und Landkreise mit ihren jeweiligen Betrieben sowie private Mobilitätsdienstleister. MobiData BW bündelt solche Informationen, die dann an einem Ort zugänglich sind. Die Mobilitätsdatenbündelung bietet eine Basis zum Aufbau nutzerzentrierter Services und innovativer Anschlussanwendungen, welche durch ein bei der NVBW angesiedeltes Transferzentrum unterstützt werden. Die Daten und die Services können u.a. zur Verkehrssteuerung von Kommunen und Mobilitätsdienstleistern sowie zur Information von MitarbeiterInnen und BesucherInnen durch viele Institutionen genutzt werden.

Als Datenabnehmer bzw. Anwender von MobiData BW werden daher neben eigenen Landesanwendungen (z.B. Auskunfts-Apps) in erster Linie institutionelle Partner aus dem öffentlichen und privaten Bereich adressiert. So gibt es im Rahmen der Usergroup des bundesweiten Mobilitätsdatenmarktplatzes Gespräche mit Navigationsdienstleistern. MobiData BW folgt damit dem aus der Innovationsförderung bekannten Gedanken eines „Business Ecosystems“, bezieht diesen aber auf nachhaltige Mobilität.

Auch Start-ups und zivilgesellschaftliche Akteure aus der Open Data Szene zählen zum Adressatenkreis von MobiData BW. Bereits 2018 richtete das VM die Hackathon-Veranstaltung „Digital Mobility Hack BW“ aus, in deren Rahmen EntwicklerInnen Prototypen für die digitale Mobilität entwickelten und erfolgversprechende Lösungen zusammen mit Mobilitätsanbietern weiterverfolgten. Vom 27. bis zum 29. November 2020 schließt sich daran der komplett als Online-Event stattfindende MobiData BW Hackathon an.[6]

[6] https://vm.baden-wuerttemberg.de/de/politik-zukunft/zukunftskonzepte/mobidata-bw-hackathon/

3 Technischer sowie lizenzrechtlicher Rahmen

MobiData BW besteht aus drei technischen Komponenten. In einer ersten Ausbaustufe ist hiervon derzeit das Herzstück des Angebots, die Datenintegrationsplattform, verfügbar. Die Softwarekomponenten werden durch von der NVBW als Betreibergesellschaft beauftragte Technologielieferanten entwickelt und zur Verfügung gestellt. Tabelle 1 bietet hierzu einen Überblick:

Tab. 1: Technische Komponenten von MobiData BW

	Datenintegrations-plattform	**Intermodaler Routing-dienst**	**Widget-Builder**
Einsatzzweck	Datenabnehmer können an einer zentralen Stelle Mobilitätsdaten des Landes und von Dritten über einheitliche und standardisierte Schnittstellen abfragen.	Intermodale Routingauskunft wird als offener Service zur Anschlussanwendung in landeseigenen Apps aber auch für Anwendungen von privaten und öffentlichen Partnern zur Verfügung gestellt.	Website-Betreiber können sich auf Basis des intermodalen Routingdienstes Mobilitätsauskünfte für User-Oberflächen generieren lassen und einfach auf ihre Websites einbetten.
Techn. Spezifikation	Datenbereitstellung derzeit via WMS, WFS, DORA-POI, GBFS, GTFS	Datenbereitstellung via DORA-Trip, OmniTrip, NeTEX	Datenzugriff via HTTP
Umsetzung	In Betrieb seit September 2020	Geplant 2021	Geplant 2021

Zu den genannten Komponenten hinzu kommt, dass die NVBW in Kooperation mit den baden-württembergischen Verkehrsverbünden die Elektronische Fahrplanauskunft EFA-BW betreibt. Über deren TRIAS-Schnittstelle und GTFS-Dateien werden Fahrplandaten und Echtzeitinformationen in das Gesamtkonstrukt integriert. Im breiten Verständnis von MobiData BW als Dachmarke für Mobilitätsdaten in Baden-Württemberg ist auch die EFA-BW eine wichtige Säule. Im Oktober 2020 sind auf MobiData BW die in Tabelle 2 dargestellten Datensätze verfügbar.

Der Datenbestand der Plattform wird in Partnerschaft mit Datenlieferanten aus der öffentlichen Hand (Behörden), Forschung und Privatwirtschaft weiter ausgebaut. Da derzeit kaum Regularien zur Offenlegung von Mobilitätsdaten aus den verschiedenen Quellen bestehen, sind freiwillige Vereinbarungen die Basis für den weiteren Ausbau.

Im Prozess der Datenerschließung wird von der NVBW als Betreiberorganisation auch eine lizenzrechtliche Homogenisierung durchgeführt. Daten können unter der „Datenlizenz Deutschland – Namensnennung – Version 2.0“ zur kommerziellen und/oder nicht-

Tab. 2: Verfügbare Datensätze

	Anbieter	**Art der Daten**	**Schnittstellen (Export)**
ÖPNV	EFA-NVBW	ÖPNV-Verbindungen (Soll-Fahrplan) und Fußgängerrouting	WMS, WFS, GTFS, DORA-POI
Parken	Verband Region Stuttgart	209 P+R-Areale in der Region Stuttgart (Stand 2019 statisch)	WMS, WFS, DORA-POI
Parken	Parkraumbewirtschaftung Baden-Württemberg (PBW)	Stuttgart (10 Objekte) + Karlsruhe (2 Objekte), dynamisch	WMS, WFS, DORA-POI
Parken	DB BahnPark GmbH	52 Parkflächen der DB BahnPark GmbH Statische POIs, teilweise dynamisch	WMS, WFS, DORA-POI
Bike-Sharing	Call a Bike/RegioRad Stuttgart	131 Standorte von Leihstationen in Stuttgart/-Ganz BW technisch angebunden	WMS, WFS, GBFS, DORA-POI
Car-Sharing	Flinkster, in Vorbereitung: Deer Carsharing	Verfügbarkeit und Standorte BW	WMS, WFS, DORA-POI

kommerziellen Anschlussanwendung verwendet werden.[7] Der Datenabruf erfolgt über das Portal www.mobidata-bw.de. Das Open Data-Portal wurde auf Basis der führenden Open-Source Datenkatalog-Software CKAN aufgebaut.

Literaturverzeichnis

[BMU19] Bundesministerium für Umwelt, Naturschutz und nukleare Sicherheit. (2019). *Klimaschutzplan 2050. Klimaschutzpolitische Grundsätze und Ziele der Bundesregierung (2. Auflage).* Berlin.

[BMVI17] Bundesministerium für Verkehr und digitale Infrastruktur. (2017). *"Eigentumsordnung" für Mobilitätsdaten? Eine Studie aus technischer, ökonomischer und rechtlicher Perspektive.* Berlin.

[IMBW17] Ministerium für Inneres, Digitalisierung und Migration Baden-Württemberg (2017). *Digitalisierungsstrategie der Landesregierung Baden-Württemberg.* Stuttgart.

[VMBW17] Ministerium für Verkehr Baden-Württemberg. (2017). *Intelligente Mobilität der Zukunft: Digitalisierung in der Schlüsselrolle.* Stuttgart.

[7] https://www.mobidata-bw.de/pages/lizenz

[VMBW18] Ministerium für Verkehr Baden-Württemberg. (2018). *Verwegen. Ideenschmiede für Digitale Mobilität. Impulse und Empfehlungen aus der Ideenschmiede.* Stuttgart.

[VMBW20] Ministerium für Verkehr Baden-Württemberg. (2020). *Digitale Mobilität: Nachhaltig und digital unterwegs in Baden-Württemberg.* Stuttgart

[SMBW19a] Staatsministerium Baden-Württemberg. (2019). Maßnahmen zur Förderung digitalisierter und nachhaltiger Mobilität der Zukunft. Abgerufen am 18. 11 2019 von https://vm.baden-wuerttemberg.de/de/service/presse/pressemitteilung/pid/massnahmen-zur-foerderung-digitalisierter-und-nachhaltiger-mobilitaet-der-zukunft/

[SMBW19b] Staatsministerium Baden-Württemberg. (2019). *Zweiter Fortschrittsbericht Strategiedialog Automobilwirtschaft BW.* Stuttgart.

Künstliche Intelligenz

Künstliche Intelligenz – Die dritte Welle

Ute Schmid[1] Voker Tresp[2] Matthias Bethge[3] Kristian Kersting[4] Rainer Stiefelhagen[5]

Abstract: Aktuelle Forschungsarbeiten aus dem Bereich Künstlichen Intelligenz werden vorgestellt. Dabei werden drei Perspektiven auf das Gebiet Maschinelles Lernen präsentiert, die über rein datenintensive Blackbox-Verfahren hinausgehen: Es werden Methoden vorgestellt, mit denen Erklärungen für die Entscheidungen von KI-Systemen generiert werden, aktuelle neurowissenschaftlich Ansätze zum maschinellen Sehen gezeigt und eine Möglichkeit Vorwissen in den Prozess des machinellen Lernens einzubringen aufgezeigt.

Keywords: Erklärbare KI; Wissensgraphen; Computer Sehen

1 Einführung

Das Thema Künstliche Intelligenz erfährt seit einigen Jahren sehr viel Aufmerksamkeit. Insbesondere besteht großes Interesse, KI-Technologie für viele Anwendungsbereiche – von Industrie 4.0 über Medizin bis hin zu Bildung – zu erschließen. Zunehmend zeigt sich, dass hier datenintensive Black-Box-Ansätze des maschinellen Lernens alleine nicht geeignet sind. Für einen robsten und transparenten Ensatz zu ermöglichen, müssen Klassifikationsentscheidungen adaptiv und kontextsensitiv sein, sich an menschliche Anforderungen anpassen und sollten vorhandenes breichsspezifisches Wissen berücksichtigen können. Drei Perspektiven auf entsprechende Weiterentwicklungen von Methoden des maschinellen Lernens werden im Folgenden präsentiert.

Die kurzen Texte beruhen auf teilweise überarbeiteten Kurzfassungen der drei eingeladenen Vorträge von Ute Schmid, Matthias Bethge und Volker Tresp im Rahmen einer von Kristian Kersting und Rainer Stiefelhagen organisierten Session zum Thema Künstliche Intelligenz.

[1] Universität Bamberg, Kognitive Systeme, ute.schmid@uni-bamberg.de
[2] Ludwig Maximilian Universität München & Distinguished Research Scientist at Siemens AG, Corporate Technology, volker.tresp@siemens.com
[3] Universität Tübingen & Amazon Scholar, Computational Neuroscience & Machine Learning, matthias.bethge@bethgelab.org
[4] Technische Universität Darmstadt, Künstliche Intelligenz und Maschinelles Lernen, kersting@cs.tu-darmstadt.de
[5] Karlsruhe Institute of Technology, Computer Vision for Human-Computer Interaction, rainer.stiefelhagen@kit.edu

doi:10.18420/inf2020_08

2 Die Dritte Welle der KI – Vom rein datengetriebenem Blackbox Lernen zu interaktiven und erklärbaren Ansätzen

Maschinelles Lernen wird als eine der wichtigsten Zukunftstechnologien für viele Bereiche der Wirtschaft und der Gesellschaft angesehen. Insbesondere Erfolge von Ansätzen der tiefen neuronalen Netzen auf Bilddaten versprechen, dass Modelle für komplexe Entscheidungsszenarien direkt aus Rohdaten gelernt werden können. Zunehmend zeigt sich allerdings, dass rein datengetriebene Ansätze in vielen Bereichen nicht umsetzbar sind: Zum einen können die hohen Anforderungen an die Menge und die Qualität an Daten, die hier für benötigt werden, häufig nicht oder nur mit sehr hohem Aufwand generiert werden. Zum anderen sind Entscheidungen von Blackbox-Modellen in vielen Bereichen rechtlich und ethisch unzulässig. Entsprechend wird aktuell die sogenannte *3rd Wave of AI* ausgerufen, nach der nun Erklärbarkeit und Parterschaftlichkeit die Ansätze des rein datengetreibenen maschinellen Lernens ablösen [TK19]. Maschinelles Lernen bietet eine Fülle verschiedener Ansätze. Je nach Problembereich können häufig auch direkt interpretierbare Ansätze eingesetzt werden [Mu18; Ru19].

Erklärungen können in verschiedenen Modalitäten erfolgen – insbesondere visuell und verbal. Für komplexe Entscheidungen sind visuelle Erklärungen häufig nicht ausdrucksstark genug. Beispielsweise kann ein Hervorheben desjenigen Bereiches im Bild eines Gewebeschnitts, auf dem die Entscheidung eines gelernten Modells maßgeblich beruht, häufig nur als Plausibilitätscheck dienen – etwa, dass tatsächlich der Bereich, in dem sich Tumorgewebe befindet, beachtet wird. Verbale Erklärungen können dagegen auf relevante Merkmalsausprägungen (‘Das Tumorgewebe hat einen Durchmesser größer 2 mm’), auf

Abb. 1: Beispielhafte Umsetzung eines Ansatzes zum erklärenden interaktiven maschinellen Lernen im Bereich Medizin.

quantifierte Aussagen (‘Alle identifizierten Metastasen sind kleiner als 1 mm’) und auf Relationen (‘Das Tumorgewebe berührt das Fettgewebe’) verweisen [SF20]. In einer beispielhaften Umsetzung für die digitale Gewebepathologie (siehe Abb.1) wird demonstriert, wie verbale Erklärungen erzeugt und von Domänenexperten korrigiert werden können. Durch interaktives Lernen wird es dadurch möglich, gelernte Modelle durch Einbringen von Expertenwissen inkrementell zu verbessern.

3 Maschinelles Lernen mit Wissensgraphen

Maschinelles Lernen mit Wissensgraphen findet zunehmend Interesse, sowohl im akademischen als auch im industriellen Umfeld [Ni15]. Die Knoten in Wissensgraphen sind Konzepte (Entitäten, Klassen, Attribute,. . .), die anhand ihrer semantischen Eigenschaften und ihrer Beziehungen zueinander beschrieben werden. Wissensgraphen können über Ihre Adjazenztensoren dargestellt werden, aus denen Tensormodelle abgeleitet werden können, die es dann erlauben, neue Fakten abzuleiten [NTK11]. In unserem Vortrag haben wir aufgezeigt, wie maschinelles Lernen mit Wissensgraphen in industriellen Anwendungen [Hi18] (Abbildung 2) und zur Unterstützung klinischer Entscheidungen. verwendet werden kann. Wichtige Probleme, mit denen wir uns im klinischen Umfeld befassen, sind fehlende Daten, Erklärbarkeit und Bewertung von Ansätzen zur klinischen Entscheidungsunterstützung [Wu20] . Weitere Schwerpunkte unserer Arbeit sind Szenengraphen in der Bilderkennung [BMT17] und die Untersuchung von Bezügen zu Wahrnehmung und Gedächtnis [Tr15].

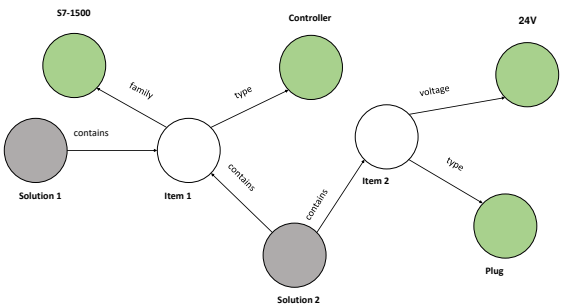


Abb. 2: Ein Wissensgraph beschreibt industrielle Komponenten, deren Eigenschaften und Bezüge zu Kundenlösungen.

4 Das sehen Maschinen aber anders

Maschinen werden immer besser darin, Wahrnehmungsaufgaben wie Objekt- oder Spracherkennung zu lösen. Die Eigenschaften, die von Maschinen verwendet werden, um zum Beispiel eine Katze von einem Hund zu unterscheiden, sind jedoch ganz anders als die Merkmale, die Menschen verwenden [Ge20]. Entsprechend vorsichtig sollte man sein, wenn man Maschinen kognitive Fähigkeiten, wie Objekterkennung oder Verstehen von Szenen zuschreibt. Psychologische Erkenntnisse und Untersuchungen können dazu beitragen, die Unterschiede zwischen menschlichem und maschinellem Sehen zu charakterisieren und zu verringern.

5 Abschließende Bewertung

Alle drei Beiträge haben den Fokus auf maschinellem Lernen als dem aktuell am meisten beachteten Bereich der Künstlichen Intelligenz Forschung. Jedoch zeigen die Beiträge deutlich auf, wo die aktuellen datenintensiven Ansätze ihre Grenzen haben. Die dritte Welle der KI Forschung erweitert den Blick über statistisches maschinelles Lernen hinaus hin zu hybriden Ansätzen, in denen klassische KI-Methoden aufgegriffen und weiterentwickelt werden. Zudem besteht neues Interesse an interdisziplinären Fragestellungen, insbesondere um Mensch-KI Interaktion partnerschaftlich zu gestalten.

Literatur

[BMT17] Baier, S.; Ma, Y.; Tresp, V.: Improving visual relationship detection using semantic modeling. In: ISWC. Springer, 2017.

[Ge20] Geirhos, R.; Jacobsen, J.-H.; Michaelis, C.; Zemel, R.; Brendel, W.; Bethge, M.; Wichmann, F. A.: Unintended cue learning: Lessons for deep learning from experimental psychology. Journal of Vision 20/11, S. 652–652, 2020.

[Hi18] Hildebrandt, M.; Sunder, S. S.; Mogoreanu, S.; Thon, I.; Tresp, V.; Runkler, T.: Configuration of industrial automation solutions using multi-relational recommender systems. In: Joint European Conference on Machine Learning and Knowledge Discovery in Databases. Springer, S. 271–287, 2018.

[Mu18] Muggleton, S. H.; Schmid, U.; Zeller, C.; Tamaddoni-Nezhad, A.; Besold, T.: Ultra-Strong Machine Learning: comprehensibility of programs learned with ILP. Machine Learning 107/7, S. 1119–1140, 2018.

[Ni15] Nickel, M.; Murphy, K.; Tresp, V.; Gabrilovich, E.: A Review of Relational Machine Learning for Knowledge Graphs. Proceedings of the IEEE/, 2015.

[NTK11] Nickel, M.; Tresp, V.; Kriegel, H.-P.: A three-way model for collective learning on multi-relational data. In: Proceedings of the 28th International Conference on International Conference on Machine Learning. S. 809–816, 2011.

[Ru19] Rudin, C.: Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence 1/5, S. 206–215, 2019.

[SF20] Schmid, U.; Finzel, B.: Mutual Explanations for Cooperative Decision Making in Medicine. KI-Künstliche Intelligenz 34/2, S. 227–233, 2020.

[TK19] Teso, S.; Kersting, K.: Explanatory interactive machine learning. In: Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society. S. 239–245, 2019.

[Tr15] Tresp, V.; Esteban, C.; Yang, Y.; Baier, S.; Krompaß, D.: Learning with memory embeddings. NIPS 2015 Workshop on Nonparametric Methods for Large Scale Representation Learning/, 2015.

[Wu20] Wu, Z.; Yang, Y.; Ma, Y.; Liu, Y.; Zhao, R.; Moor, M.; Tresp, V.: Learning Individualized Treatment Rules with Estimated Translated Inverse Propensity Score. arXiv preprint arXiv:2007.01083/, 2020.

Ethik und KI

R. Reussner, A. Koziolek, R. Heinrich (Hrsg.): INFORMATIK 2020,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2021 99

Towards a Flexible Framework for Algorithmic Fairness

Philipp Hacker[1], Emil Wiedemann[2], Meike Zehlike[3]

Abstract: Increasingly, scholars seek to integrate legal and technological insights to combat bias in AI systems. In recent years, many different definitions for ensuring non-discrimination in algorithmic decision systems have been put forward. In this paper, we first briefly describe the EU law framework covering cases of algorithmic discrimination. Second, we present an algorithm that harnesses optimal transport to provide a flexible framework to interpolate between different fairness definitions. Third, we show that important normative and legal challenges remain for the implementation of algorithmic fairness interventions in real-world scenarios. Overall, the paper seeks to contribute to the quest for flexible technical frameworks that can be adapted to varying legal and normative fairness constraints.

Keywords: algorithmic fairness; optimal transport; discrimination; algorithmic affirmative action; EU law

1 Introduction

Discrimination by machine learning applications has emerged as a major challenge for the widespread deployment of AI technology. In recent years, it has not only sparked a vivid academic debate in computer science and law [DH+12; FF+15; BS16; Ha18; ZH+20; WM+20; He20], but has also garnered significant attention from the media and policymakers. Importantly, bias in AI systems poses not merely a theoretical problem; rather, it is documented in an increasing number of reports and empirical studies. For example, Amazon had to discard its AI recruitment tool because it persistently discriminated against women [Re18]. Similarly, a study recently found that the machine learned model managing health populations in US hospitals exhibited racial bias [OP+19]. In fact, language itself encodes potentially discriminatory patterns, which surface in AI-based translation applications [CB+17]. Hence, algorithmic discrimination remains a pressing issue for both technology and policy, despite the research efforts of the past years.

Against this background, this short overview paper will approach the subject in three steps: first, it briefly covers core legal areas applying to algorithmic discrimination under EU law. Second, it explains how different technical strategies for remedying algorithmic

[1] European University Viadrina, European New School of Digital Studies, Chair for Law and Ethics of the Digital Society, Große Scharrnstraße 59, 15230 Frankfurt (Oder), Germany, hacker@europa-uni.de
[2] University of Ulm, Institute for Applied Analysis, Helmholtzstr. 18, 89081 Ulm, Germany, emil.wiedemann@uni-ulm.de
[3] Max Planck Institute for Software Systems, Campus E1 5, 66123 Saarbrücken, Germany, meikezehlike@mpi-sws.de

doi:10.18420/inf2020_09

discrimination can be combined using algorithmic fairness metrics and optimal transport. Third, it highlights that important normative challenges, both ethical and legal, remain for implementing technical solutions to bias in AI systems. The paper chiefly draws on longer articles in which the described algorithm, its functioning and legal implications were explained in greater detail [Ha18; ZH+20].

2 The Law of Algorithmic Discrimination

In EU law, there are two main fields that deal with discrimination in machine learning contexts: antidiscrimination law and data protection law. Both areas are tightly regulated at the EU level. The following paragraphs can only present a succinct summary of the application of that body of law to algorithmic discrimination.

2.1 Antidiscrimination Law

Concerning antidiscrimination law, it should first be noted that, in our view, it clearly covers instances of automated decision making. The scope of antidiscrimination legislation in the EU is not limited to activities mainly or exclusively undertaken by human beings. Rather, the provisions are formulated in a technologically neutral manner. Perhaps even more importantly, from a teleological perspective, it seems clear that the objective of antidiscrimination law – to prevent material and immaterial harm stemming from discriminatory practices – demands its application to AI systems as well. Generally, intentional practices will qualify as direct discrimination (e.g., masking bias in data sets). By contrast, most instances of discrimination by AI tools will amount to indirect discrimination, particularly if they are a mere side effect of machine learning optimization brought about unintentionally.

Importantly, however, instances of discrimination are not prohibited per se under EU law. Rather, they can generally be justified if the decision maker pursues a legitimate aim and the discriminatory practice is proportionate to that aim. Courts will undertake a comprehensive analysis of all circumstances to evaluate the merits of a claim for justification. As we have explained in greater detail elsewhere [Ha18], however, the chances of justifying algorithmic discrimination are crucially influenced by the source of the bias. If the discriminatory outcome is an artifact of "biased training" (such as biased training data or inadequate feature/target selection), it will be quite hard to justify the outcome. Decision makers must bear reasonable costs to avoid biased training and modeling processes. If, however, bias arises because the targeted qualities are, in reality, unequally distributed between different protected groups, one might speak of 'unequal ground truth'. In these cases, jurisprudence by the Court of Justice of the European Union (CJEU) suggests that it will be easier to meet the burden of justification if the disparate outcome reflects a 'real difference' between the groups that is relevant for the decision at hand.[4]

[4] See, e.g., CJEU, Case C-96/80, *Jenkins*, EU:C:1981:80, para 12; Case C-170/84, *Bilka-Kaufhaus*, EU:C:1986:204, para 36.

For potential victims of discrimination, the importance of the source of bias for justification raises an important barrier to justice, however. To differentiate between the different sources, and to gauge potential costs of litigation, plaintiffs would need access to the data and the model. Under discrimination law, however, they generally lack access rights if they merely suspect that the decision might be driven by discrimination.[5] Moreover, they are often not even in a position to establish a prima facie case of statistically significant differential treatment without access to data and the model.

2.2 Data Protection Law

At this stage, data protection law, and particularly the GDPR, enter the scene. Not only does it comprise access rights (Art. 15 GDPR), but it also contains public enforcement tools which might remedy the enforcement deficiencies of antidiscrimination law. However, for these GDPR tools to be available, algorithmic discrimination would have to qualify as a breach of data protection law. There are two main routes to bring about this result.

First, Article 5(1)(a) GDPR contains the principles of accuracy and of fair data processing. Arguably, these principles are breached in cases of unjustified discrimination [A18]. This entails an integrated understanding of data protection and antidiscrimination law. Second, in cases of automated decision making, Article 22(3) GDPR mandates that the controllers must take measures to safeguard the rights and freedoms of data subjects. Scholars have rightly argued that bias reduction mechanisms must count among those necessary safeguards [see ref. in Ha18, p. 1177].

Such potential violations of the GDPR open up the toolbox of GDPR enforcement. For example, data protection authorities may impose heavy fines (Art. 83 GDPR), and may perform algorithmic audits (Art. 58(1)(b) GDPR). Hence, the conceptual convergence of antidiscrimination and data protection law facilitates a much more robust enforcement mechanism for cases of AI bias under EU law.

However, even such an integrated understanding of EU law is unlikely to overcome the challenges of algorithmic discrimination on its own. Two main problems remain with this approach. First, data protection authorities, upon which much of the enforcement burden would rest, are notoriously under-resourced and, in some cases, may lack the technical competences to rigorously audit complex AI systems. Second, many of the instruments of data protection law are ex post correction strategies. This means, however, that the harm has already occurred, which may be quite severe in cases of discrimination.

[5] See CJEU, Case C-415/10, *Meister*, EU:C:2012:217, para. 46.

3 Matching Code and Law: Algorithmic Fairness and Optimal Transport

In view of the mentioned shortcomings of a purely legalistic approach, it seems attractive to look for strategies of preventing discrimination ex ante by implementing non-discrimination principles directly at the code level. This is precisely the object of an ever-growing research effort conducted around the world under the banner of algorithmic fairness. While an overview of the many fairness definitions employed in the computer science literature transcends the scope of this paper [DL19; PS20], most definitions can be categorized into two main groups: individual fairness and group fairness [DH+12; FS+16].

3.1 The Divide between Individual and Group Fairness

Individual fairness compares attributes and outcomes for single individuals. More specifically, it usually demands that two individuals that are similar in terms of their attributes (with the exception of protected attributes) are mapped, by the algorithmic process, onto a similar output [DH+12]. This corresponds to the old Aristotelian notion of 'treating likes alike'. Intriguingly, such an understanding also matches the definition of equality before the law (Art. 20 of the Charter of Fundamental Rights of the EU) in the jurisprudence of the CJEU.[6] Group fairness, by contrast, compares outcomes at the group level [DH+12; PS20]. One common metric, statistical parity, demands that the same proportion of individuals must be positively selected from each protected group [YS+18; PS20]. This corresponds to a more outcome-egalitarian concept of social justice [B18]. Moreover, since the groups are treated in a statistically equivalent way, it is very hard to find indirect discrimination once statistical parity is fulfilled.

Importantly, however, there is generally a trade-off between group fairness and individual fairness if the true score distributions between the protected groups differ at the outset [FS+16; ZH+20; but see also B20; ZC20]. In such a case, enforcing group fairness implies that some individuals from the discriminated group will be positively selected while similarly qualified individuals from the privileged group will be rejected, breaching individual fairness. Therefore, the discussion around the correct fairness metrics reproduces long-standing philosophical debates about meritocratic vs. egalitarian concepts of social justice [B18].

3.2 Bridging the Divide

In a model described in greater detail elsewhere [ZH+20], we propose to bridge this development by continuously interpolating between measures of individual and group fairness. To this end, we define a mapping from individuals' raw scores (the outcome

[6] See, e.g., CJEU, Case C-149/10, *Chatzi*, ECLI:EU:C:2010:534, para. 64.

of some ML process) to fair scores. Hence, our fairness tool generally functions as a post-processing approach, but it may theoretically also be used as a pre-processing tool by applying it to the target value of the training data points. In the mapping, we introduce a parameter (θ) which allows to fine-tune the degree to which the raw score distributions of different protected groups are approximated. More precisely, we calculate a barycenter of the different group raw score distributions. The barycenter represents an intermediate distribution with minimal distance from the various raw group score distributions in a least square sense.[7] The parameter θ, which runs from 0 to 1, determines the degree to which each raw score distribution is shifted toward the barycenter. If θ equals 0, the raw score distributions are left unchanged and individual differences between the scored individuals are fully preserved. This minimizes what we call an individual fairness error.[8] In the other extreme, if θ is set to 1, distributions are fully matched onto the barycenter, achieving statistical parity. While our mapping guarantees monotonicity within groups, the ranking is usually changed, under high θ values, for individuals who belong to different groups. Hence, the individual fairness error rises. Moreover, to the extent that the raw scores correctly represent the target qualities, the enforcement of group fairness breaches similarity-based definitions of individual fairness if the raw scores were unequally distributed between the different groups.

Importantly, we use optimal transport theory to minimize the information loss of the decision maker in the mapping process. Not only is the barycenter calculated through optimal transport, but the mapping of each individual group toward the barycenter, to the degree defined by θ, also follows optimal transport. Under the assumption that the raw scores correctly reflect the target qualities, the mapping therefore maximizes decision maker utility under varying flexible fairness constraints defined by the choice of the θ value.

3.3 Advantages and Limitations of Our Model

Empirical evaluations of the model on synthetic data and on the LSAT data set show that the choice of θ indeed directly determines the mentioned individual and group fairness values. The validation points to significant advantages, but also limitations of our model.

On the positive side, first, the model can match legal standards by choosing θ in such a way that the approximation of the group distributions prevents the finding of indirect discrimination. As mentioned, indirect discrimination presupposes a statistical disparity between the positive selection probabilities of the different groups. Hence, our model can be used as an important building block for a compliance model for AI systems. Second, the brief legal discussion has shown that differential outcomes between protected groups may be justified depending on the circumstances of the case. Our model provides a flexible framework with which decision makers, and regulators, may adapt the outcome of algorithmic processes

[7] See Theorem 2.4 in [ZH+20] for details.
[8] See Equation 2.9 in [ZH+20] for details.

to various situations in which different degrees of individual or group fairness may be warranted. In each of these situations, optimal transport guarantees that decision maker utility is maximized under the varying fairness constraints. Third, by choosing different θ values for different groups, the model allows to consciously push certain particularly disadvantaged groups. This allows us to handle problems of intersectional discrimination in which certain subgroups are disadvantaged because of multiple protected attributes. Fourth, finally, the model works equally for one-dimensional and multi-dimensional scores. This may be important if a scoring process contains assessments from different sources on different scales.

The use of optimal transport also limits the model to a certain extent, however. First, to function well, it necessitates group sizes of at least several hundred individuals. Hence, the model is particularly well-suited for Big Data applications, but less so for individual scenarios with few candidates. Second, it presupposes that the raw scores, while potentially imperfect, represent a useful approximation of the true target scores. Otherwise, the impact of the mapping on decision maker utility cannot be guaranteed. However, this constraint seems rather weak. If the scoring process results in unreliable raw scores, the developer must modify the scoring procedure. To the extent that the decision maker intends to use the scores, the assumption that the scores are meaningful does not seem far-fetched.

4 Normative Challenges for Algorithmic Fairness

Technical solutions can help to mitigate bias in decision making. The possibility to consciously choose certain fairness parameters constitutes a significant advantage of AI versus human decision making. Nevertheless, important normative challenges remain.

4.1 The Choice of θ

In the context of our model, one obvious question relates to the choice of θ. Institutionally, decision makers could be granted leeway to choose θ as they please (within the constraints of antidiscrimination law). However, in certain areas of pronounced societal importance (such as education, housing etc.) one may imagine that the legislator, regulatory agencies, or the courts define specific thresholds or even concrete values for the choice of θ [see also VB17].

Overall, that choice will reflect the trade-off between more individualistic, meritocratic approaches to social justice on the one hand and more outcome-egalitarian ones on the other. In our view, one possibility might be to draw on a capabilitarian approach [S09]. From this perspective, a low θ value could be chosen if two conditions are cumulatively met [ZH+20]. First, there is high confidence in the correctness of the raw scores. Second, a meritocratic allocation regime is normatively desired because the decision does not affect

basic capabilities or resources. Conversely, a higher θ value could be selected if decision makers only have an intermediate confidence in the correctness of the raw scores i.e., the scores have some information utility for the decision maker, but may to a certain extent be affected by bias (e.g., biased training). A higher θ value would preserve the in-group rankings, and hence the information encoded in the scores, while mitigating the effect of bias between the groups. Moreover, even if the raw scores are likely to be correct, high θ values might be justified for normative reasons if the allocation of basic resources and capabilities is at stake.

4.2 Legal Constraints

Finally, the enforcement of group fairness may engender conflicts with affirmative action law [Ha18; He20; Be20; ZH+20]. In the EU, as in other jurisdictions, it is difficult to draw the line between remedying unjustified discrimination and breaching the right to equality of those individuals negatively affected by affirmative action. This trade-off precisely mirrors the fundamental difference between individual and group fairness discussed above. In our view, under EU law, it must generally be possible to modify a selection procedure, including the raw scores, if the certification of the procedure would lead to unjustified discrimination (e.g., biased training). Otherwise, the restraints of affirmative action law would force the decision maker to violate basic non-discrimination standards.

What remains doubtful, however, is the extent to which decision makers can engage in affirmative action if the original outcome or procedure would have been legally justified (e.g., unequal ground truth). In the EU, standards differ based on whether an affirmative action policy (like the fairness intervention) is enacted before or after a first ranking of the candidates has been conducted. While criteria are more lenient before that first selection (Badeck case),[9] the jurisprudence of the CJEU is quite restrictive after it. In the Marschall case, the Court ruled that a re-ranking based on protected attributes must not be absolute and unconditional.[10] Rather, it may only be undertaken on the basis of an objective assessment taking into account all specific criteria of the affected individuals. In this way, the Court seeks to ensure that individual characteristics speaking against downgrading an applicant (e.g., specific individual hardship for single parents) are not blindly overridden by affirmative action preference.

Overall, the distinction between the Badeck and the Marschall case suggests a dividing line between post-processing approaches on the one hand and pre-/in-processing approaches on the other: the latter are easier to justify, under EU affirmative action law, than the former. Similarly, under US law, re-ranking based on 'race-norming' individual test results is prohibited in employment contexts [GJ96], making post-processing approaches difficult to justify in this domain [RB+20]. More generally, in the Ricci case,[11] the US Supreme Court

[9] CJEU, Case C-158/97, *Badeck*, EU:C:2000:163, paras. 55 and 63 (concerning selection for training and interview).
[10] CJEU, Case C-409/95, Marschall, EU:C:1997:533, para. 33.
[11] Ricci *v.* DeStefano, 557 U.S. 557, 563 (2009).

seemed to adopt a more lenient stance toward interventions at the test-design stage, rather than post-processing modifications [KH+17; K17; He20].

It seems unclear, however, whether such a distinction between the different interventions is normatively and legally justified. For a start, post-processing approaches afford the advantage that the re-ranking result is precisely defined, while in- and pre-processing approaches risk "overshooting", which would be worse for the privileged groups. In our view, it might be worth considering an 'attenuated Marschall standard', under which algorithmic affirmative action is possible, irrespective of the stage of the intervention, as long as meaningful human scrutiny is applied. However, such scrutiny need not accompany each and every case of re-ranking, but might be restricted to those of specific legal interest (e.g., particular individual hardship). In this way, the consideration of important individual criteria in specific cases can be combined with a broad application of algorithmic fairness to large data sets in which human scrutiny of each individual re-ranking decision would often be prohibitively costly.

5 Conclusion

Despite significant research efforts in law and computer science over the last years, a consensus on fairness metrics for the purposes of preventing discriminatory outcomes in machine learning contexts has not emerged yet. Arguably, this testifies both to the vagueness of legal standards and to the significantly diverging factual circumstances of the various areas in which AI models are deployed. Therefore, it seems important to develop technical tools that match the necessary openness and flexibility of the legal provisions and the facts of the case. The model described in this paper seeks to contribute to this endeavor.

Bibliography

[A18] Article 29 Working Party, Guidelines on automated individual decision-making and profiling for the purposes of Regulation 2016/679, WP251rev.01, 2018.

[B18] Binns, R. Fairness in machine learning: Lessons from political philosophy. In *Conference on Fairness, Accountability and Transparency*, 149-159, 2018.

[B20] Binns, R. On the apparent conflict between individual and group fairness. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 514-524, 2020.

[Be20] Bent, J. R., Is algorithmic affirmative action legal, 108(4) *Geo. L.J.*, 803-853, 2020.

[BS16] S. Barocas and A. Selbst. Big data's disparate impact. *California Law Review*, 104:671–732, 2016.

[CB+17] Caliskan, A., Bryson, J. J., & Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science*, *356*(6334), 183-186.

[DH+12] C. Dwork, M. Hardt, T. Pitassi, O. Reingold, and R.S. Zemel. Fairness through awareness. In *Proceedings of the 3rd Conference on Innovations in Theoretical Computer Science*, 214–226, 2012.

[DL19] J. Dunkelau and M. Leuschel. Fairness-aware machine learning. 2019. Working Paper, https://www.phil-fak.uni-duesseldorf.de/fileadmin/Redaktion/Institute/Sozialwissenschaften/Kommunikations-_und_Medienwissenschaft/KMW_I/Working_Paper/Dunkelau___Leuschel__2019__Fairness-Aware_Machine_Learning.pdf.

[FF+15] Feldman, M., Friedler, S. A., Moeller, J., Scheidegger, C., & Venkatasubramanian, S. Certifying and removing disparate impact. In *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining*, 259-268, 2015.

[FS+16] Friedler, S. A., Scheidegger, C., & Venkatasubramanian, S. On the (im)possibility of fairness. *arXiv preprint arXiv:1609.07236*, Working Paper, 2016.

[GJ96] Greenlaw, P. S., & Jensen, S. S. (1996). Race-norming and the Civil Rights Act of 1991. *Public personnel management*, *25*(1), 13-24.

[Ha18] P. Hacker. Teaching fairness to artificial intelligence: Existing and novel strategies against algorithmic discrimination under EU law. *Common Market Law Review*, 55:1143–1186, 2018.

[He20] D. Hellman. Measuring algorithmic fairness. *Va. L. Rev.*, 106:811–866, 2020.

[K17] Kim, P. T. Auditing algorithms for discrimination. *U. Pa. L. Rev. Online*, *166*, 189-203, 2017.

[KH+17] J.A. Kroll, Y. Huey, S. Barocas, E.W. Felten, J.R. Reidenberg, D.G. Robinson, and H. Yu. Accountable algorithms. *U. Pa. L. Rev.*, 165, 633-705, 2017.

[OP+19] Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, *366*(6464), 447-453.

[PS20] D. Pessach and E. Shmueli. Algorithmic fairness. *arXiv preprint arXiv:2001.09784*, Working Paper, 2020.

[RB+20] M. Raghavan, S. Barocas, J.M. Kleinberg, and K. Levy. Mitigating bias in algorithmic hiring: Evaluating claims and practices. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, pages 469–481, 2020.

[Re18] Reuters, Amazon ditched AI recruiting tool that favored men for technical jobs, *The Guardian*, October 11, 2018, https://www.theguardian.com/technology/2018/oct/10/amazon-hiring-ai-gender-bias-recruiting-engine

[VB17] Veale, M., & Binns, R. (2017). Fairer machine learning in the real world: Mitigating discrimination without collecting sensitive data. *Big Data & Society*, *4*(2), 2053951717743530.

[WM+20] S. Wachter, B. Mittelstadt, and C. Russell. Why fairness cannot be automated: Bridging the gap between EU non-discrimination law and AI. *Available at SSRN: https://ssrn.com/abstract=3547922*, 2020.

[YS+18] Yang, K., Stoyanovich, J., Asudeh, A., Howe, B., Jagadish, H. V., & Miklau, G. A nutritional label for rankings. In *Proceedings of the 2018 international conference on management of data*, 1773-1776, 2018.

[ZC20] Zehlike, M. and Castillo, C. Reducing disparate exposure in ranking: A learning to rank approach. In Proceedings of The Web Conference 2020, 2849-2855, 2020.

[ZH+20] M. Zehlike, P. Hacker, and E. Wiedemann. Matching code and law: Achieving algorithmic fairness with optimal transport. *Data Mining and Knowledge Discovery*, 34:163–200, 2020.

Sichere und zuverlässliche Systeme: Datensouveränität

SES-14: Sichere und zuverlässliche Systeme: Datensouveränität

Jörn Müller-Quade[1] Roger Gutbrod[2] Volkmar Lotz[3] Fabian Biegel[4] Benny Fuhry[5]
Jeremias Mechler[1]

Datensouveränität, also die Fähigkeit, informiert und selbstbestimmt zu entscheiden, wie und von wem Informationen über die eigene Person oder Institution, eigene Handlungen oder Produkte erhoben, verarbeitet und weitergegeben werden, ist eine Grundvoraussetzung für eine freiheitliche Gesellschaft, für eine funktionierende Wirtschaft und für einen unabhängigen Staat. Es bedarf einer großen Anstrengung, um digitale Souveränität im Allgemeinen und Datensouveränität im Speziellen zu erreichen. Dazu ist eine große konzertierte Aktion von Forschung, Wirtschaft und Politik nötig. Die Handlungsfelder umfassen sichere und zuverlässliche Systeme, aber auch Software, Prüf- und Zertifizierungsverfahren, sowie staatliche Regulierung und neue Geschäftsmodelle für Internetdienste.

Im Rahmen der GI-Jahrestagung 2020 wurden in der Session „Sichere und zuverlässliche Systeme: Datensouveränität“ aktuelle Entwicklungen zu diesem Thema vorgestellt.

Sichere Mehrparteienberechnungen für die Praxis: Hardware-Vertrauensannahmen – Jeremias Mechler, KIT Das Gebiet der *sicheren Mehrparteienberechnungen* befasst sich mit dem Szenario, dass n sich gegenseitig misstrauende Parteien gemeinsam eine Funktion f auf ihren geheimen Eingaben auswerten wollen. Dabei sollen gewisse Eigenschaften wie *Korrektheit* oder *Vertraulichkeit* der geheimen Eingaben sichergestellt sein. Seit den 80er Jahren sind Protokolle zur sicheren Mehrparteienberechnung von jeder effizient berechenbaren Funktion f bekannt, deren Sicherheit mit einem (gedanklichen) Szenario vergleichbar ist, in dem eine dritte vertrauenswürdige Partei die Auswertung von f auf den geheimen Eingaben übernimmt und nichts als das Ergebnis verrät. Protokolle mit solch starken Sicherheitsgarantien können einen wichtigen Beitrag zur Datensouveränität leisten.

Leider sind klassische Ansätze sehr ineffizient, weshalb es in der Praxis oftmals andere Lösungen bedarf. Helfen kann hier sichere Hardware wie beispielsweise Trusted Execution Environments. Diese versprechen, die Funktion der angenommenen vertrauenswürdigen

[1] Karlsruher Institut für Technologie (KIT), D-76131 Karlsruhe, crypto-info@iti.kit.edu
[2] SAP SE, D-69190 Walldorf, roger.gutbrod@sap.com
[3] SAP Labs France, F-06254 Mougins Cedex, volkmar.lotz@sap.com
[4] SAP SE, D-69190 Walldorf, fabian.biegel@sap.com
[5] SAP SE, D-69190 Walldorf, benny.fuhry@sap.com

doi:10.18420/inf2020_10

Partei ohne Effizienzverlust im Vergleich zu einer normalen CPU zu erfüllen. Dies setzt jedoch ein in der Praxis nicht gerechtfertigtes Vertrauen voraus – so ist im Allgemeinen nämlich nicht überprüfbar, ob Geheimnisse nicht doch beispielsweise an den Hersteller oder an staatliche Stellen abfließen. Schlimmer noch: Bekannte Implementierungen wie Intel SGX weisen eine Reihe von gravierenden Sicherheitslücken auf. Es gilt daher, durch die geschickte Kombination von klassischen kryptographischen Techniken Lösungen zu entwickeln, die effizient sind und kein übermäßiges Vertrauen voraussetzen – beispielsweise weil die sichere Hardware nur verschlüsselte Geheimnisse sieht. Forschung ist nicht nur bei diesem Entwurf der Protokolle notwendig, sondern auch bei passenden und in der Praxis aussagekräftigen Sicherheitsbegriffen.

Design einer effizienten, verschlüsselten und komprimierten In-Memory-Datenbank mit Hilfe von Intel SGX – Benny Fuhry, SAP Datensouveränität ist wichtig für jede einzelne Privatperson. Gleichzeitig forscht SAP an diesem Thema, um ihren Kunden größtmögliche Datensouveränität bei der Nutzung der SAP-Cloud zu gewähren. In den vergangenen Jahren lagern immer mehr Unternehmen ihre Datenspeicherung und -verarbeitung zu Cloud-Anbietern aus. Erfolgreiche Angriffe auf ausgelagerte Daten zeigen jedoch, dass Sicherheit noch immer ein großes Problem darstellt. In vielen Fällen werden die Daten während der Speicherung und der Übertragung verschlüsselt, aber die Verarbeitung finden auf Klartext-Daten statt. Aus diesem Grund können Angreifer mit Speicherzugriff an die sensitiven Daten gelangen. Im zweiten Vortrag der Session stellte SAP Security Research ein Design einer sicheren, effizienten, verschlüsselten und komprimierten In-Memory-Datenbank vor. Für den Schutz der Daten während der Verarbeitung wurde Intel SGX, ein Trusted Execution Environment, genutzt.

GAIA-X – Ein aktueller Stand zur europäischen Dateninfrastruktur – Fabian Biegel, SAP Datensouveränität hat nicht nur eine technologische, sondern auch eine politische und wirtschaftliche Dimension, die von Vereinbarungen oder Richtlinien zur Datenportabilität bis hin zur Standardisierung von Schnittstellen und Protokollen zur Gewährleistung der Interoperabilität alle Aspekte einer Daten- und Cloud-Infrastruktur umfasst, die einen reibungslosen und sicheren Betrieb ermöglicht, in dem die Kunden die Kontrolle über ihre Daten behalten. Nachdem die ersten beiden Vorträge der Session technische Forschungsergebnisse und Innovationen vorgestellt haben, die Datensouveränität in einer Cloud-Umgebung erst ermöglichen, stellte der abschließende Vortrag „GAIA-X – Ein aktueller Stand zur europäischen Dateninfrastruktur“ die europäische Initiative zur Realisierung der Datensouveränität vor. Mit GAIA-X entwickeln Vertreter aus Politik, Wirtschaft und Wissenschaft gemeinsam einen Vorschlag zur Gestaltung der nächsten Generation einer Dateninfrastruktur für Europa. Ziel ist eine sichere und vernetzte Dateninfrastruktur, die den höchsten Ansprüchen an digitale Souveränität genügt und Innovationen fördert. In einem offenen und transparenten digitalen Ökosystem sollen Daten und Dienste verfügbar gemacht, zusammengeführt und vertrauensvoll geteilt werden können. Fabian Biegel, Repräsentant

von SAP in der europäischen GAIA-X Foundation, diskutierte Chancen und Herausforderungen der digitalen Transformation und stellte das Konzept von GAIA-X als Bindeglied zweier Ökosysteme – des Daten- und des Infrastruktur-Ökosystems – dar, das durch föderierte Dienste zum Identitäts- und Vertrauensmanagement, zum sicheren und vom Anwender kontrollierten Datenaustausch, zur Compliance und weiteren gebildet wird. GAIA-X hat einen ersten Meilenstein mit der Gründung der europäischen GAIA-X Foundation erreicht und erwartet Mitte 2021 erste (Teil-)Lösungen, die konform zum technischen Konzept, das in zweiter Version bis Ende des Jahres vorliegen wird, sind.

Den Zuhörerinnen und Zuhörern konnte im Rahmen der drei Vorträge und Diskussionen ein interessanter Einblick in die aktuellen Entwicklungen in der theoretischen und angewandten Forschung sowie der Wirtschaftspolitik gegeben werden.

Energie

Energieinformatik – Von der Forschung in die Umsetzung

Ein Rückblick auf den Track „Energie“ auf der INFORMATIK2020

Astrid Nieße [1], Veit Hagenmeyer[2]

Abstract: Die Energieinformatik widmet sich den Herausforderungen zukünftiger postfossiler Energiesysteme aus Sicht der Digitalisierung. Für den Track „Energie“ der INFORMATIK2020 wird in diesem Beitrag ein kurzer Rückblick gegeben. Dabei wird eingegangen auf die drei entsprechenden Vorträge: Im ersten Vortrag stellte Frau Prof. Dr. Anke Weidlich von der Universität Freiburg aktuelle Entwicklungen im Bereich der Flexibilitätsmodellierung zur Diskussion. Im zweiten Vortrag zeigte Herr Dr. Postina von der EWE AG —dem fünftgrößten Energieversorgungsunternehmen in Deutschland— in vielfältigen Beispielen auf, wie in der Praxis im Rahmen der Digital Factory der EWE AG der Entwicklungsprozess von der Ideenfindung bis zur Umsetzung der Idee als Geschäftsmodell oder als Produkt gestaltet wird. Im dritten Vortrag widmete sich schließlich Frau Stefanie Mollemeier von der Mettenmeier GmbH in Paderborn der Frage, wie konsortiale Softwareentwicklung im regulierten Umfeld der Energiesysteme erfolgreich sein kann.

Keywords: Energieinformatik; Energiesysteme; Flexibilitätsmodellierung; Digital Factory; konsortiale Softwareentwicklung

Zusammenfassung

Die Herausforderungen, die sich aus der Transformation der Energiesysteme hin zu postfossilen und u.a. in diesem Sinne nachhaltigen Energiesystemen stellen, sind sehr unterschiedlicher Natur. Nicht zuletzt beschäftigt die Gesellschaft viele politische Fragestellungen, z.B. in der Diskussion um neue Kernkraftwerke in den Niederlanden, oder – ebenfalls hochaktuell – der Bau von Gaspipelines. Eng verflochten mit dem Erfolg unterschiedlicher Entwürfe ist aber die Beantwortung technischer Fragestellungen, denen wir uns in cyberphysischen Energiesystemen widmen, mit nicht unerheblichen Rückwirkungen auf die politische Notwendigkeit und gesellschaftliche Akzeptanz.

Die technologischen Fragestellungen betreffen zunehmend Fragestellungen und Methoden der Informatik – sei es für die Entwicklung und Bewertung neuartiger Steuerungskonzepte, die Entwicklung von neuen Geschäftsmodellen oder aber die Software-Entwicklung, die zur Umsetzung in die Praxis alles in Einklang zu bringen hat.

[1] Fk. II – Department für Informatik/Carl von Ossietzky Universität Oldenburg, Abteilung Digitalisierte Energiesysteme, 26111 Oldenburg, astrid.niesse@uni-oldenburg.de, https://orcid.org/0000-0003-1881-9172
[2] Karlsruhe Institute of Technology (KIT), Institute for Automation and Applied Informatics, 76344 Eggenstein-Leopoldshafen, veit.hagenmeyer@kit.edu

doi:10.18420/inf2020_11

Zu allen drei Bereichen ist es uns gelungen, im Track Energie der INFORMATIK2020 renommierte Vortragende zu begeistern, uns einen Einblick in ihre aktuelle Arbeit zu geben.

Organisiert wurde dieser Track u.a. durch die Fachgruppe Energieinformatik der Gesellschaft für Informatik, die sich u.a. dem Schnittbereich von Energietechnik, Informatik, Elektro- und Automatisierungstechnik widmet und dort zahlreiche Fachleute aus Wirtschaft und Wissenschaft zusammenbringt.

Im ersten Vortrag stellte Frau Prof. Dr. Anke Weidlich von der Universität Freiburg aktuelle Entwicklungen im Bereich der Flexibilitätsmodellierung zur Diskussion: Für die effiziente Einbindung variabler Stromerzeugung aus erneuerbaren Energien ist die Einbindung flexibler Erzeuger und Verbraucher erforderlich. Flexibilität kann auf vielfältige Weise bereitgestellt werden. Jede Flexibilitätsoption hat eigene Einsatzmöglichkeiten und technische Restriktionen. Um Flexibilität optimal einsetzen und als Produkt vermarkten zu können, werden Modelle benötigt, mit deren Hilfe man Flexibilität beschreiben und quantifizieren kann. In ihrem Vortrag stellte Frau Weidliche unterschiedliche Modellierungsansätze vor, wie sie in den letzten Jahren u.a. im SINTEG-Vorhaben C/sells erarbeitet wurden. In der nachfolgenden Diskussion wurden Vor- und Nachteile intensiv diskutiert.

Der Frage, wie solche Ansätze effektiv in die Praxis umgesetzt werden können, widmete sich Herr Dr. Postina von der EWE AG in seinem Vortrag. Als Head of DataScience bei der EWE AG ist er für die Entwicklung datenbasierter Geschäftsmodelle verantwortlich. In vielfältigen Beispielen zeigte Herr Dr. Postina auf, wie im Rahmen der Digital Factory der EWE AG der Entwicklungsprozess von der Ideenfindung bis zur Umsetzung der Idee als Geschäftsmodell oder als Produkt gestaltet wird. Insbesondere das methodische Vorgehen der Potentialanalyse mittels sogenannter "BrainWaves" wurde vorgestellt und in der nachfolgenden Diskussion intensiv anhand der vorgestellten Beispielprodukte diskutiert.

Schließlich widmete sich der letzte Beitrag der Session der Frage, wie konsortiale Softwareentwicklung im regulierten Umfeld der Energiesysteme erfolgreich sein kann. Frau Stefanie Mollemeier von der Mettenmeier GmbH in Paderborn stellt dazu aktuelle Erfahrungen aus dem Softwareentwicklungsalltag vor, die der Besonderheit dieses Umfeldes Rechnung tragen: Die IT-Landschaft der Netzbetreiber ist von Systemvielfalt, Herstellerbindung, Datenredundanzen und komplexe Schnittstellen gekennzeichnet. Um die oft unerwünschten Effekte insbesondere der Schnittstelleninkompatibilitäten und Herstellerbindung zu vermeiden, wurde die openKONSEQUENZ Genossenschaft ins Leben gerufen. Der Zusammenschluss von Netzbetreibern, Software-Herstellern, Beratungsunternehmen und Forschungseinrichtungen entwickelt und betreibt gemeinsam unter Open-Source-Lizenz offene, modulare und sichere Software für den Einsatz in der Energie- und Wasserversorgung. Im Vortrag erläuterte Frau Mollemeier die Grundsätze der Zusammenarbeit in der Genossenschaft und die regelmäßigen Herausforderungen der Genossenschafts-Mitglieder aus Sicht eines Service-Providers. Auch an diesen Vortrag schloss sich eine intensive Diskussion an, in der Frau Mollemeier die Herausforderungen und Chancen des gewählten Modells der konsortialen Softwareentwicklung überzeugend darstellen konnte.

Der gewählte inhaltliche Bogen von der Erstellung wissenschaftlicher Modelle über die Entwicklung von Geschäftsmodellen für neue Technologien bis hin zur Softwareentwicklung, –sämtlich im regulierten und dynamischen Umfeld cyberphysischer Energiesysteme– führte zu einer anregenden Diskussion. Wir bedanken uns ganz herzlich bei den Organisatorinnen und Organisatoren der diesjährigen GI-Jahrestagung, bei den Vortragenden sowie den zahlreichen Teilnehmerinnen und Teilnehmern der Session.

Workshops

(Agiles) Enterprise Architecture Management in Forschung und Praxis

EAM im Spannungsfeld von Agilität und Digitalisierung

Carsten Brockmann[1] Christian Czarnecki[2] Eldar Sultanow[3]

Keywords: Enterprise Architecture Management; Agilität; Digitalisierung

Dieses Jahr findet der Workshop „(Agiles) Enterprise Architecture Management in Forschung und Praxis“ zum fünften Mal in Folge statt. Ziel des Workshops ist eine übergreifende Sichtweise auf das Thema EAM, welches aus unterschiedlichen Industrien und methodischen Perspektiven diskutiert wird. Dabei steht die anwendungsorientierte Forschung im Hinblick auf konkrete Praxisbeispiele und Handlungsempfehlungen im Vordergrund.

Auch wenn das Konzept der Agilität nicht neu ist, sondern auf die System- und Organisationstheorie sowie auf das Wissensmanagement zurückführt, haben Themen rund um die Agilität in den letzten Jahren deutlich an Bedeutung gewonnen. Insofern haben wir den Begriff „Agilität“ dieses Jahr als einen Schwerpunkt des Workshops im Titel ergänzt. Aus Sicht des EAM lässt sich die Frage der Agilität auf die folgenden beiden konträren Positionen zuspitzen:

1. Die Existenzberechtigung von Architekten und zentralen Architekturmaßnahmen werden komplett infrage gestellt, agile Methoden werden als Gegenthese zum EAM gesehen und sollen dieses ersetzen.
2. Agilität und EAM werden als sich ergänzende Konzepte verstanden, Agilität wird mit architektonischen Mitteln skaliert, und EAM wird durch agile Methoden flexibler.

Genau diese zweite Position gewinnt im Kontext der zunehmenden Digitalisierung – also die stetige Anpassung von Strukturen auf technologische Innovationen – zunehmend an Bedeutung. In diesem Spannungsfeld aus Agilität und Digitalisierung sind die diesjährigen Workshopbeiträge angesiedelt. So werden praktische Empfehlungen für die Digitalisierung des EAM, Auswirkungen Künstlicher Intelligenz auf Architekturen sowie neue Referenzmodelle des Technologiemanagements diskutiert. Die praktischen Beispiele demonstrieren EAM im Cloudumfeld, in Bezug auf Mobilitätsdienste sowie das Krisenmanagement in der Luftfahrt.

[1] Deloitte Consulting GmbH, Kurfürstendamm 23, 10719 Berlin, cbrockmann@deloitte.de
[2] Hochschule Hamm-Lippstadt, Marker Allee 76–78, 59063 Hamm, christian.czarnecki@hshl.de
[3] Universität Potsdam, Karl-Marx-Straße 67, 14482 Potsdam, eldar.sultanow@wi.uni-potsdam.de

doi:10.18420/inf2020_12

Digitalisation by Enterprise Architecture Management: Practical Recommendations

Benjamin Warnecke[1]

Abstract: This research in progress paper describes how enterprise architecture management can support digitalisation. It is based on an initial literature review and provides practical recommendations. The revealed key strategic recommendations are: (1) use capability-based planning, (2) include product IT in EAM and (3) give EAM-practices the required power in the organisation.

Keywords: enterprise architecture management; digitalisation; digital transformation

1 Introduction

Enterprise architecture (EA) can support transformations (e.g. process reengineering) by impact analysis and providing implications for capability changes in the area of business, data, application and technology [WF06]. Despite this promising possibility the topic enterprise architecture management (EAM), is perceived by the author, as discussed academically. There are recently signs to change the discussion towards a more practical one, like the proposal from market researcher Gartner to renew enterprise architecture programs [Om20]. For this reason, the focus of this research in progress paper is to give practical recommendations on how digitalisation can be enabled by enterprise architecture management. Thus, the target group is practitioners and scientists. The applied method is an initial literature review.

The terms digitalisation and EAM have several definitions. We support the definition of digitalisation as an organisational- or process-related change, which is induced and enabled by technology. This often includes a shift from physical to digital value creation [KSS18, PH15].

Enterprise architecture management is understood as the fundamental organisation of a corporation and "the principles governing its design and evolution" [WF06]. The primary domains in an organisation according to TOGAF are business, data, application and technology and their relationship to each other [To18]. Furthermore, EAM can be used to create and visualise reference architectures, like the ones for the aviation- [Su16] and pharma industry [Su18].

[1] FOM University of Applied Sciences, Agrippinawerft 4, 50678 Cologne, benjamin.warnecke@fom-net.de

doi:10.18420/inf2020_13

2 Related work

The literature shows the current level of research on the topic being discussed. The section "Practical recommendations" below presents further developments based on previous research.

The International Telecommunication Union provides guidance on the evaluation of the maturity of an organisation regarding digital transformation. Examples of organisation's maturity are "digital laggard" or "digital leader" [ITU19]. Kaidalova et al. [KSS18] enhanced the discussion of digitalisation regarding the consideration of not only corporate IT, but also product IT (e.g. smart devices). EAM has also been explored in prior studies by Babar & Yu [BY15]. Due to the authors, EAM must widen its scope and handle differently multiple levels of dynamics (e.g. operational- and innovation processes) in organisations [BY15].

More than 50 frameworks are available in enterprise architecture managements [Ma11]. One of the most used is "The Open Group Architecture Framework" (TOGAF). The latter was the basis for an approach to move organisations towards digitalisation [HA16].

Korhonen & Halén suggest that digitalisation should be fully integrated in every hierarchical level and function of an organisation for its full benefit. Digitalisation is currently often run as a program in organisations [KH17].

There are also related approaches to transform organisations, e.g. Architecture of Integrated Information Systems (ARIS) [ÖB05]. Further details will be provided in the section "Practical recommendations".

A recent study from 2019 with more than 1,500 respondents shows some insights for enabling digitalisation. It was conducted by the company "LeanIX GmbH", that provides an EAM-tool with the same name [En19].

Wißotzki and Sandkuhl suggest a process for the digitalisation and transformation of organisations. It includes mainly such disciplines as business model management, capability management and enterprise architecture management [WS17].

3 Practical recommendations

In this section we give practical recommendations to drive a company towards digitalisation with the help of EAM. We propose to split recommendations into strategic and supportive. The latter assists the strategic recommendations from EAM-related areas.

3.1 Strategic recommendations

First of all, an organisation should conduct a maturity evaluation regarding digitalisation. This can be a starting point for roadmap creation towards digitalisation.

There are several pathways to move an organisation to the desired position, e.g. digital leader [ITU19].

Furthermore, a capability-based planning should be conducted. This is a technique for planning of investments in capabilities (e.g. finance) [KH17]. Business capability maps in general strengthen and improve the communication of business and IT and thus business IT-alignment. Examples are the alignment on business demands and prioritizing of the support per capability. However, one main challenge of business capability maps is a lack of understanding followed by a high creation effort. These findings were gathered by Khosroshahi et al. in expert interviews [Kh18]. Considering the benefits mentioned above, it can be helpful to invest effectively in digitalisation. That is why a diagram showing the relation between capabilities and their supported applications is needed. The latter can help to identify white spots and do business modelling. By creating such a diagram organisations support digitalisation by verifying and defining the current and future business model.

We strongly suggest that an EAM-practice should govern, design and evaluate the full enterprise architecture. EAM's scope is currently focused on corporate IT and often disregards the product IT. IT components, which are built into the products are presently understood under the term product IT [KSS18]. An example can be a fingerprint sensor, which uses machine learning servers. With this information in an EAM-tool one can optimise the workload of the server with other IT-services from the corporate IT. And there are more examples where product IT with its internet of things devices can leverage more potential in combination with the corporate IT.

The multiple level of dynamics can be available in organisations, e.g. regarding approaches of software development (agile vs. traditional). It can be that one team creates microservices in an agile way and another team works with the traditional waterfall model. Both must be considered in EAM, but in a different way regarding EAM-governance and autonomy. This will help organisations to keep up with today's highly dynamic environment [BY15].

In order to digitalise an organisation two things must be considered from organisational point of view. Firstly, EAM-departments should be located close to the top in the IT organisation. By such a way it is possible to align directions directly with CIO/CDO and EAM has the power to lead the required changes of digitalisation. Secondly, digitalisation should be integrated in every function of an organisation in order to fully leverage the potential of digitalisation [KH17]. Thus these functions must align strongly with EAM to meet the desired strategy.

TOGAF's core is the Architecture Development Method (ADM). This approach enables organisations to create an enterprise architecture as well as to "shape and govern business transformation and implementation projects" [To18]. For this reason TOGAF ADM can be used as a driver for digitalisation. Similar research proposes to focus an EAM-practice on four areas: aligning a unified view for all stakeholders, creating the architecture vision, operating an architecture repository and conducting stakeholder management [HA16]. It

depends on the project and organisation, which deliverables per phase are required. An overview of them is given in the TOGAF standard [To18].

If an organisation wants to use an existing process (e.g. [WS17]) to conduct the digitalisation or adapt a process, needs to be decided upfront. The mentioned process requires knowledge in the areas of business model management, capability management and enterprise architecture management [WS17]. In large-scale enterprises this requirement can be matched, but small and medium-sized enterprises likely have a knowledge gap in this matter. External consultants might be hired to fill the latter.

Furthermore, we suggest that EAM should challenge programs and projects in order to fit to the future state of the organisation. This could be done via deliverables and reviews at specific milestones. In such a way the alignment of programs and projects towards the strategy can be improved.

The above-mentioned cross-industry study, conducted by the provider of the EAM-tool "LeanIX" with more than 1,500 participants, indicates that the top three obstacles for digitalisation are: (1) legacy applications, (2) IT security concerns and (3) application integration challenges [En19]. These obstacles can be overcome in an EAM-practice by analysing the as-is situation, getting buy-in from stakeholders and creating an EAM-roadmap. The study also revealed that a high EA maturity is a strong driver for having a high digital maturity, e.g. being a digital leader [En19]. That is why we strongly suggest ensuring that the maturity of your EAM-practice corresponds with your prospective digital maturity.

The figure (Fig. 1) below gives a summary of the strategic recommendations, which are surrounded by the supportive recommendations. Both can enable an organisation to become a digitalised one.

Fig. 1: Summary of strategic recommendations

3.2 Supportive recommendations

The strategic recommendations can be assisted with supportive ones. These are from EAM-related areas.

The domain business engineering is linked to the business architecture of EAM. It has the target to deliver innovative business models in a systematic approach. There are several approaches in business engineering, notable are: Multiperspective Company Modelling (dt. Multiperspektivische Unternehmensmodellierung [MEMO]), Business Engineering St. Gallen (dt. St. Galler Ansatz des Business Engineering), Architecture of Integrated Information Systems (ARIS) and TOGAF. The mentioned approaches can help to drive the transformation of digitalisation [ÖB05].

As stakeholder commitment is required to change an organisation towards a digital one, it is recommended to manage the changes in a change management model. Several models are reported in the literature to address this need [WW17]. Worth a mention are the Prosci ADKAR Model, the 3-Phase Model of Kurt Lewin and Kotter's 8-Step-Change Model.

4 Results & Discussion

In this paper we have given some practical recommendations to use EAM for driving digitalisation. We have suggested to: (1) use capability-based planning, (2) include product IT in EAM and (3) give EAM-practices the required power in the organisation. As a supportive recommendation it is worth to mention that EAM-related areas like business engineering and change management can support the digitalisation program.

Future research to give a complete overview of the topic might extend the practical recommendations and is appreciated. It will be important that future research completes the literature review in this research in progress paper. Furthermore, a case study in the industry could be conducted to evaluate the proposed recommendations.

Bibliography

[BY15] Z. Babar and E. Yu, 'Enterprise Architecture in the Age of Digital Transformation', in Advanced Information Systems Engineering Workshops, vol. 215, A. Persson and J. Stirna, Eds. Cham: Springer International Publishing, 2015, pp. 438–443.

[En19] LeanIX GmbH, 'Enterprise Architecture Insights Report 2019'. Available: https://www.leanix.net/en/download/enterprise-architecture-insights-report-2019 (accessed May 21, 2020).

[HA16] M. Hafsi and S. Assar, 'What Enterprise Architecture Can Bring for Digital Transformation: An Exploratory Study', in 2016 IEEE 18th Conference on Business Informatics (CBI), Paris, France, Aug. 2016, pp. 83–89, doi: 10.1109/CBI.2016.55.

[ITU19] International Telecommuncation Union, 'Digital transformation and the role of enterprise architecture'. Available: https://www.itu.int/pub/D-STR-DIG_TRANSF-2019. 2019 (accessed May 17, 2020)

[KH17] J. J. Korhonen and M. Halén, 'Enterprise Architecture for Digital Transformation', in 2017 IEEE 19th Conference on Business Informatics (CBI), Thessaloniki, Greece, Jul. 2017, pp. 349–358, doi: 10.1109/CBI.2017.45.

[Kh18] P. Aleatrati Khosroshahi, M. Hauder, S. Volkert, F. Matthes, and M. Gernegroß, 'Business Capability Maps: Current Practices and Use Cases for Enterprise Architecture Management', presented at the Hawaii International Conference on System Sciences, 2018, pp. 4603–4612, doi: 10.24251/HICSS.2018.581.

[KSS18] J. Kaidalova, K. Sandkuhl and U. Seigerroth, 'How Digital Transformation affects Enterprise Architecture Management – a case study', IJISPM - International Journal of Information Systems and Project Management, no. 06, pp. 5–18, 2018, doi: 10.12821/ijispm060301.

[Ma11] D. Matthes, Enterprise-Architecture-Frameworks-Kompendium: über 50 Rahmenwerke für das IT-Management. Berlin Heidelberg Dordrecht London New York: Springer, 2011.

[ÖB05] H. Österle and D. Blessing, 'Ansätze des Business Engineering', HMD Prax. der Wirtschaftsinformatik, vol. 241, pp. 7–17, 2005.

[Om20] G. Omale, '8 Steps for a High-Impact Enterprise Architecture Program'. Available: https://www.gartner.com/smarterwithgartner/8-steps-for-a-high-impact-enterprise-architecture-program/ (accessed May 30, 2020).

[PH15] M. E. Porter and J. E. Heppelmann, 'How Smart, Connected Products Are Transforming Companies', Harvard Business Review, October 2015, Oct. 01, 2015.

[Su16] Sultanow, E., Brockmann, C., Schroeder, K. and Breithaupt, C.: Lufthansa Aviation Standard: Developing an Open Group Reference Architecture for the Aviation Industry. In: Mayr, H. C. & Pinzger, M. (Eds.), INFORMATIK 2016, Klagenfurt, Gesellschaft für Informatik e.V. (p. 825-836), 2016.

[Su18] Sultanow, E., Chircu, A. & Schroeder, K. amd Kern, S.: A Reference Architecture for Pharma, Healthcare & Life Sciences. In: Czarnecki, C., Brockmann, C., Sultanow, E., Koschmider, A. and Selzer, A. (Eds.), Workshops der INFORMATIK 2018, Berlin, Architekturen, Prozesse, Sicherheit und Nachhaltigkeit, Gesellschaft für Informatik e.V. (p. 25-40), 2018.

[To18] The Open Group, 'The TOGAF® Standard, Version 9.2' 2018 Available: https://pubs.opengroup.org/architecture/togaf9-doc/arch/ (accessed May 23, 2020).

[WF06] R. Winter and R. Fischer, 'Essential Layers, Artifacts, and Dependencies of Enterprise Architecture', in 2006 10th IEEE International Enterprise Distributed Object Computing Conference Workshops (EDOCW'06), Hong Kong, China, 2006, pp. 1–12, doi: 10.1109/EDOCW.2006.33.

[WS17] M. Wißotzki and K. Sandkuhl, 'The Digital Business Architect – Towards Method Support for Digital Innovation and Transformation', in The Practice of Enterprise Modeling, vol. 305, G. Poels, F. Gailly, E. Serral Asensio, and M. Snoeck, Eds. Cham: Springer International Publishing, 2017, pp. 352–362.

[WW17] M. R. Wolf and U. Wiese, 'Change Management für Informatiker: Theorie und Praxis für erfolgreiche Projekte', in INFORMATIK 2017 Chemnitz, Germany, Sep. 2017, M. Eibl and M. Gaedke Eds., 2017, pp. 2595–2597, doi: 10.18420/IN2017_265.

The Enterprise Architect as a Crisis Manager: Insights from Lufthansa

Carsten Breithaupt[1], Jonas Vieracker[2], Alina Chircu[3], Sean Cox[4], Eldar Sultanow[5]

Abstract: In this paper we argue that the Enterprise Architect (EA) should be considered as a crisis manager. In times of a crisis organizations must create a holistic view on the situation and evaluate the proposed measures. Evaluation will be based on experience, data, upfront-created scenarios and assessments of risks, cost and benefits. The EA is already engaged in processes such as Continuity Management, Risk Management, IT Security, Business Process Management, IT Strategy, and others. Therefore, we propose that the EA is one good candidate (not the only one) for handling organizational crises. This paper presents a model of overall crisis management that incorporates an enterprise architecture view as well as related dimensions of crisis management: the ability to control a crisis, the level of communication during and after a crisis, the type of change brought by a crisis, the crisis' outcomes, its distribution within an enterprise, and an organization's criticality of business functions. Finally, we highlight how the EA role supports crisis management at Lufthansa, a top German aviation company.

Keywords: Crisis Management; Enterprise Architecture; Lufthansa

1 Introduction

Major, large-scale incidents such as the COVID-19 pandemic, the WannaCry cyberattack, Target's consumer data breach, the 9/11 terrorist attacks, Volkswagen's emissions scandal, or the Exxon Valdez and BP oil spills have repeatedly demonstrated that organizations are prone to significant internal and external threats as well as risks resulting from mismanagement of how threats are addressed [Bu17, Ko20, PC98]. These events can all be characterized as an organizational crisis – *"an event perceived by managers and stakeholders as highly salient, unexpected and potentially disruptive"* [Bu17]. For the most part, companies do not anticipate these kinds of crises and suffer severe losses, during and even after the crises – perhaps because each predicament is different and needs special treatment in order to prevent, detect and control the impact, avoid damage and accelerate recovery. Thus, the better a risky situation is analyzed and the more sophisticated the crisis management capability of an enterprise is, the better an organization can handle crisis situations. Thus,

[1] Deutsche Lufthansa AG, Von-Gablenz-Straße 2-6, 50679 Köln, Germany, carsten.breithaupt@dlh.de
[2] Capgemini, Bahnhofstraße 30, 90402 Nuremberg, Germany, jonas.vieracker@capgemini.com
[3] Bentley University, 175 Forest Street, Waltham, MA, USA, achircu@bentley.edu
[4] RatPacDune Entertainment, Los Angeles, CA, USA, sean.cox@ratpacent.com
[5] Capgemini, Bahnhofstraße 30, 90402 Nuremberg, Germany, eldar.sultanow@capgemini.com

doi:10.18420/inf2020_14

crisis management is a strategic business matter that affects a company's performance in achieving its goals [Ah12, p. 63].

The crisis management principles and strategies championed by an enterprise's senior executives (CxOs) need to be operationalized in order to take effect within the entire organization. A key way to integrate strategy into a company's organizational design, processes, and overall operations is Enterprise Architecture Management (EAM). EAM is a management philosophy that embraces holistic and sustainable change and provides an opportunity to drive strategic and operational changes while considering existing and required future business capabilities and assets [Ah12, p. 57]. An Enterprise Architect (EA) is an experienced business and technology professional, usually reporting directly to a CxO, who is responsible for EAM implementation, specifically for the alignment of the business objectives with information technology (IT) strategy and principles.

In this paper, we analyze the role of EAM and the EA from the perspective of crisis management, and propose that an EA's duties should include building and maintaining crisis management capabilities, which should be incorporated in the EAM architecture. An EA's skill set and the related EAM practices provide the structure with which all relevant activities are managed to conceptualize, implement and execute enterprise strategies. As a result, EA can also design and implement crisis management to respond to an unforeseen threat [Ah12, p. 61]. We build on this idea by proposing a holistic approach to manage crises. The goal is to provide a model with which a crisis management capability can be created as a part of EAM, turning the EA into the lead crisis manager of the organization.

2 Understanding the Role of EAM and EAs in Crisis Management

2.1 Crises and crisis management

Over the past few decades, many scholars have studied organizational crises and crisis management from a variety of disciplinary perspectives [Bu17, KW17, PC98]. They define a crisis as *"low-probability, high-impact event that threatens the viability of the organization and is characterized by ambiguity of cause, effect, and means of resolution, as well as by a belief that decisions must be made swiftly"* [PC98]. Crisis management, in turn, is *"a systematic attempt by organizational members with external stakeholders to avert crises and to effectively manage those that occur"* [PC98]. Crises can be understood from several different perspectives, including internal organizational preparedness and external stakeholder relationships [Bu17], or psychological, social-political and technology-structural perspectives [PC98]. Crises also share similar stages – pre-crisis prevention, ongoing management, and post-crisis outcomes [Bu17, Ko20].

Many typologies of crises exist. Crises can occur due to internal organizational failures, external undermining threats, or mismanagement by crisis responders [Ko20]. Crises are also characterized by their controllability, severity, undesirability and intentionality [Bu17].

From a responsibility perspective, crises affect organizations as victims, accidents, or preventable situations [Bu17]. Crisis types include armed conflict/humanitarian aid, business, climate/environment, pollution/toxic effects, natural disasters, critical infrastructure, health, ICT/cyber, riots/crowds, and terrorism, and crisis concerns include risk, preparedness, political leadership, crisis communication, decision making, organizing for safety, community resilience, crisis transboundary, and aftermath [KW17], as well as crisis detection, containment, business resumption, learning, reputation, resource availability an decision making [PC98].

Researchers have also posited that the success or failure of crisis management depends on how well an organization can minimize potential risks before a triggering event, mobilize key stakeholders and support sense-making in response to a triggering event, and re-adjusting assumptions and responses for recovery after a triggering event – all of which are affected by the executive mindset, the adoption of organizational practices, and the environmental context [PC98].

2.2 EAM, EAs, and crisis management

EAM is usually implemented using frameworks - with TOGAF being a widely-known one [Th18]. It provides methods, models and patterns to create, maintain and improve enterprise architecture capabilities. It distinguishes between four architectural layers – from the business architecture that supports an agreed upon architectural vision, down to the data and application architecture to finally defining the technology of the architecture. This enables an end-to-end alignment and integration of business and IT objectives. EAM enables EAs to make sense of the situation and construct shared meaning with organizational stakeholders, design organizational structures and coordinate complex systems - which are exactly the foundation of crisis management [Bu17, PC98]. The following subsections describe how each one of the architectural layers relates to crisis, and how EAs can perform crisis management activities in each case.

Business and strategy

According to TOGAF 9.2 [Th18, p. 407] the business architecture defines the business strategy, governance, organization and key business processes. It further represents holistic, multi-dimensional views of capabilities, end-to-end value delivery, information, and organizational structure as well as the relationships between these business views and strategies, products, policies, initiatives and stakeholders. TOGAF applies the business architecture to an entire organization. Although crisis management can be a part of the domain of governance it should be considered as unique architecture capability, which consists of strategy, governance and key business processes that define how the crisis management is set up within an enterprise. During a crisis, EAs can focus on business continuity by using the architectural model to answer business and strategy key questions such as:

- Can we continue generating revenue with our current business model?
- Can the crisis be overcome by focusing the business in other regions?
- Which measures can we take to stabilize our market?
- Are there possibilities to ensure liquidity during the time of crisis?
- How can we cover our fixed cost?

Data

The data architecture describes the structure of an organization's logical and physical data assets and data management resources and how they interact [Th18, p. 409]. The data structure and data management are defined by the capabilities laid out within the business architecture. However, the data architecture also influences the strategy made on the business level of crisis management. By leveraging the power of data analysis, EA can deploy the right crisis management strategies and methods during a crisis. EAs can also use the extracted information to prevent future incidents. Therefore, the data layer is closely linked to the ability to control (which is described later in this paper).

The better the overall data models are, the more implication can be drawn from them. For instance, data can support the prevention of crisis situations that arise from inside of the company (e.g. drop in demands, quality issues of products). Further, with reliable data, crisis situations can be simulated in order to take the most effective measures during a crisis. This keeps costs low and can speed up the recovery from a crisis. Another important aspect of preventing crisis on a data layer is taking measures to circumvent any kind of data loss or fraud. Data loss and data fraud are not limited to external attacks. Data loss is also caused through the failure of systems as well as infrastructure, and companies are continuously struggling with its prevention. By implementing standardized methods regarding information security, a huge step can be taken to prevent potential data risks.

Applications

The application architecture provides a blueprint to deploy the individual applications and to manage their interactions and their relationships to core business processes of the organization. In other words, applications support the business architecture to deliver key business functions and manage data assets [Th18, p. 21]. Throughout the history of big organizations, the application portfolio grows in order to cope with imminent business needs, and oftentimes applications are built on a heterogenous technology landscape with multiple applications serving the same or similar purposes and resulting in cost increases. When demand is low, but applications still need to be maintained, a heterogenous and "none-lean" application portfolio wastes money that is needed to ensure business continuity. Therefore, EAs need to evaluate the application portfolio frequently. By leveraging the foundation implemented within the data layer, the number of applications can be reduced to the ones needed for business delivery. Further, in times of crisis a reliable data model helps

EAs pinpoint the applications critical to business continuity, allowing the focus to be on these applications.

Technology

According to TOGAF [Th18, p. 12], the technology architecture describes the logical software and hardware capabilities that are required to support the deployment of business, data and application services; this includes IT infrastructure, middleware, networks, communications, processing, standards, etc. Analog to the application portfolio, technologies used in infrastructure can be heterogenous, resulting in huge maintenance costs. As a best practice, an EA should homogenize the technology portfolio to save cost. In addition, workflows in IT infrastructure are still characterized by manual tasks. Although the focus is shifting to platform strategies with a high amount of standardization and automation, not all units of an enterprise are implementing these measures. An important solution in this area is cloud computing. Cloud computing can increase the resilience of the infrastructure to handle crises (i.e. avoid outages) and can decrease infrastructure costs when properly managed. During a crisis, additional cloud capacity can be used to meet high workload demands. This results in lower cost, as an organization does not need to purchase and run the entire infrastructure necessary to handle peak demand situations in times of crisis.

2.3 Developing EA crisis management capabilities through EAM

EAM frameworks such as TOGAF describe the development of architecture capabilities and provide related methods, best practices and standards. An EA can use these tools to develop much-needed crisis management capabilities as well, by combining top-down and a bottom-up approaches. In the top-down approach, the EA develops and maintains a crisis management capability hierarchically, starting with the business and strategy layer. The directives set on this layer define the development of crisis management on the data, application and technology levels. The EA can use the bottom-up approach to gather data and information from underlying layers in order to adapt or influence the decisions made on business and strategy level. Figure 1 depicts this closed loop of developing a crisis management capability (see Figure 1).

Furthermore, by integrating the information gathered on lower levels, such as usage data of applications, health data of the IT infrastructure or general environmental data a continuous improvement of crisis management capability can be derived. Therefore, it can be stated that data is the crucial factor of crisis management and the data layer needs to be adapted to leverage the data.

Several crisis management dimensions are useful to further guide the development of crisis management capabilities by EAs, as described in the next paragraphs.

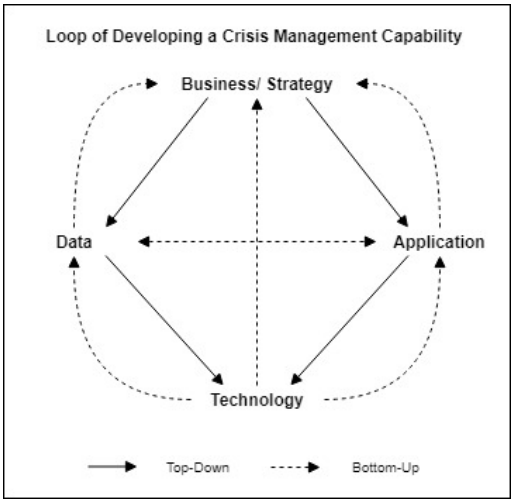


Fig. 1: Developing a Crisis Management Capability (Source: authors' own representation)

Type of change

According to Kovoor-Misra [Ko20] there are four types of changes possible in a crisis:

- **Unintentional change** occurs when the crisis changes are undesirable and are not intentional, and crate disruptions that need to be addresses.
- **Mindfully reactive change** comprises changes made by the organization after a crisis occurs in order to contain and recover from it.
- **Intentional and transformational change** is intentionally undertaken by the organization in response to or in preparation for a crisis, and may include changes in organizational structures, systems, norms and behaviors.
- **Proactive change** involves measures to prevent or address a potential crisis.

Ability to control

The ability to control describes the timing and extent of controlling and respectively manipulating (in the sense of "changing") the impacts to the organization, which increases with rising applicability of transparency [Su15, p. 9]. Ability to control is one way to characterize a crisis [Bu17]. Several types of control exist:

- **Post-preventive (corrective)** control includes measures taken after a crisis occurs. This ability does not require the detection or anticipation of influences related to the crisis. However, it demands resistance against the influences and a repository for storing the measures taken to cope with the influences.

- **Intervening (detecting)** control takes measures against influences when they occur. It requires immediate detection of influences throughout the management of the crisis.
- **Preventive** control involves taking measures for avoiding potential future influences. This indicates a need to anticipate the influences in order to take appropriate precautions.
- **Continuous** control involves a continuous effect across a time continuum. The ability of control is continuous if it is in an internal state or is natural.

Crisis outcomes

This dimension describes how enterprises emerge from a crisis [Au00]:

- **Loss through crisis** occurs when a company's business is negatively affected, for example through loses sales, orders or customers. Examples include the negative impact on airline companies from terrorist attacks or the worldwide COVID-19 pandemic.
- **Benefits from crisis** occur when the crisis positively affects the organization by creating opportunities for growth. For instance, in the case of Covid-19, the manufacturers of personal protective equipment (PPE) have benefited significantly from the increased demand for their products.

According to Watters, the focus on crisis outcomes rather than causes is important because "*there are infinite possible causes. A real risk is that you can spend all your life doing risk assessments and are not ready for what comes along. A better approach is to focus on the impacts and how they manifest themselves in terms of outcomes. It makes much more sense to do this because, for an infinite number of potential causes, the possible impacts boil down into only five outcome scenarios. That means you focus first on preparing the basic elements of business continuity so that you can survive the five possible outcome scenarios*" [Wa14, p. 8].

Level of communication

According to Laverdet et al. [La18, p. 150], effective crisis management requires organizations to communicate with stakeholders in a timely manner. This includes using the preferred communication channels of the organization to explain and provide regular updates for the current situation (to help organizational stakeholders understand the kind of crisis they are dealing with and how it is evolving) and provide instructions on how to deal with the evolving situation (at both organizational and individual level). After a crisis has ended, crisis managers should enter a learning phase and receive feedback from other members of the organization about the effectiveness of their approach in order to improve crisis management capabilities [Bu17, La18]. This approach can be adopted by an EA assuming crisis manager responsibilities. The EA can develop a crisis communication

plan together with relevant departments, such as corporate communication, orchestrate communication among different stakeholders, and filter the information available during a crisis in order to transmit the correct level of information to different groups and individuals in an organization, for example on a social intranet. Note that in our model, the dimension of communication is considered as a transversal dimension across all other dimensions, as the entire model influences communication during crisis situations.

Affected areas

According to Kovoor-Misra [Ko20] crisis can affect various areas, as follows:

- **Natural environmental** area includes the air, water, land and other natural resources associated with an organization.
- **Technical** area includes resources associated with computer, electrical, chemical and mechanical technologies, as well as raw materials used in manufacturing.
- **Economic** area includes the company's financial resources and the related management mechanisms (i.e. financial reporting and monitoring systems, etc.).
- **Human and social** area includes the physical, psychological and social aspects of the organization (i.e. individual health and behavior and organizational culture, identity and human resource structures and processes).
- **Political/reputational** area refers to the power and influence of the organization on its stakeholders, including perceptions of brand and reputation.
- **Legal** area includes the laws and regulations the organizations is subject to and the parts of the organization involved in compliance efforts.
- **Ethical** area includes the moral conduct principles and standards accepted within the organization.

Criticality of an organization's business functions

In times of crisis, an organization's priority is to ensure continuing core activities – i.e. transforming input to output at levels that satisfy the needs of key customers, with minimal interruptions or delays [PC98]. As a result, the importance of business functions within an enterprise changes during a crisis. The criticality of an organization's sub-units can be defined by using several categories [Hi00, Wa14]:

- **Vital/mission critical** business functions are necessary for survival – i.e. basic operations requirements, minimum acceptable work or service levels, or required legal aspects.

- **Essential** business functions are necessary for normal business activity – a reduction or delay in these functions will not cause an immediate negative impact, but they contribute to maintaining business continuity during the crisis.
- **Important** business functions are needed for normal operations but only have a small impact if disrupted during a crises. They support the first two types of functions, but are lower in importance.
- **Noncritical/not important** business functions have low/zero demand during a crisis, and can be put "on the back burner" in order to decrease costs and make scarce resources available for other functions.

Thus, in times of crisis the focus on an organization should be on ensuring the continuity of vital/mission critical and essential business functions. Furthermore, an organization should minimize the efforts for noncritical business functions and stabilize the performance of important functions.

Distribution of crisis management

This dimension describes the degree of centralization of the crisis management processes the success of crisis management in global organizations. We differentiate in:

- **Centralized** crisis management involves using a central point (such as the company's headquarters, or a centralized disaster management location) for all crisis management decisions. Especially when it comes to a global crisis, a standard in handling the crisis must be established centrally. While centralization ensures consistency, it can create unacceptable delays in responses to a crisis. A certain degree of freedom for subsidiaries and local managers may be needed to adjust the common crisis response to local conditions and reduce delays.
- **Decentralized** crisis management consists of local control over responses to a crisis. This implies that subsidiaries, business units, and local managers can come up with their own solutions. Decentralized crisis management is of high importance for companies with multiple business units and locations that differ in significant ways, such as due to different geographical, cultural, or economic factors. Giving these units the freedom to act and respond independently increases the chance of act quickly and achieve successful (but perhaps different) outcomes in each unit, but it can also create a disorganized, less effective and more resource-intensive response.

2.4 Relationship to other IT disciplines

Other IT management frameworks, also cover some, but not all aspects of crisis management. For example, COBIT (The Control OBjectives for Information Technology), includes IT

Risk Management AP012 and IT Continuity Management (DSS04) in its model. These areas provide processes on how to define business requirements with regards to the need for a continuous IT operations and potential risks for the IT as experienced and assessed by the business users. In addition, it is worth mentioning the "BSI IT-Grundschutz Handbuch". It is published by the German Federal Office for Information Security and defines a method to assess the protection level of an application, implicating how to deal with those applications before, during and after crises [Bu].

Evaluating other frameworks while developing an architectural approach on dealing with crisis is important as these frameworks exhibits that IT should be driven by business and its requirements. Knowing how to define such primarily non-functional requirements is integral for assessing means for crisis management.

3 Applying the model – insights from Lufthansa

To demonstrate the usefulness of our model, we applied it to a top European aviation company based in Germany – Lufthansa. Figure 2 shows the model instantiated with the Lufthansa details – including the architectural model and the six crisis management dimensions discussed previously (see Figure 2). The following sections describe relevant company practices for each component of the model.

3.1 Type of change

Lufthansa is primarily focusing on change that is initiated by the company – mindfully reactive, intentional, transformative and proactive change. The practices that are supporting the initiation of necessary change to the organization and its processes and governance include:

- **IT Strategy** is the key driver for change. By evaluating internal and external factors for its IT strategy, Lufthansa can anticipate crises and initiate change in time. For instance, Lufthansa transformed its IT operations towards being more agile in order to adjust quicker to changes in demand and other external factors.
- **IT and Business Continuity Management (ITCM/ BCM)** assesses, amongst others, the criticality of business processes and IT applications. This supports crisis management in clustering processes and applications and initiating change for the protection of the most critical systems. This is closely linked to the model's dimension on the criticality of business functions. The more critical systems are, the higher their need for continuity is. Lufthansa assesses the criticality of processes and systems on a regular basis.

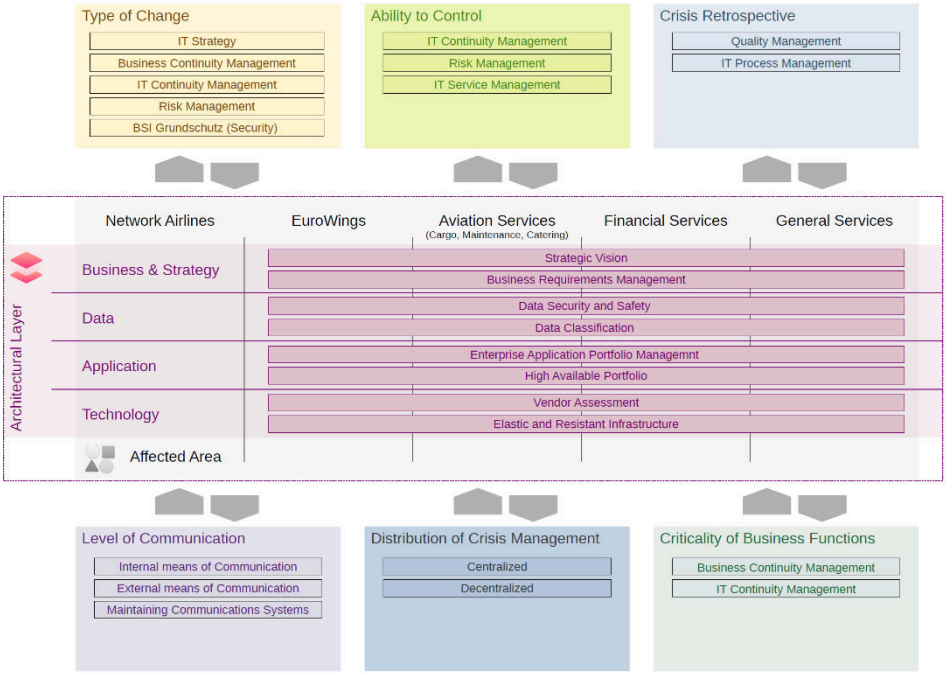


Fig. 2: Overall model applied to Lufthansa (Source: authors' own representation)

- **Risk Management** identifies threats for business and IT, as well as possible mitigation mechanisms. Being aware of risks supports the EA to initiate changes preventively in order to avoid letting risks become a crisis.
- **BSI Grundschutz** describes another initiator for change at Lufthansa based on an assessment of applications need for data protection. Its goal is to ensure a high level of confidentiality, integrity and availability. Based on the assessment outcomes, specific data security and data safety measures are taken in order to ensure the necessary protection of application data.

An EA, in his role as crisis manager, needs to understand these practices and initiate necessary measures and actively promote the associated change within the application portfolio and its underlying infrastructure. Note that the type of change dimension is not limited to these four practices. Other practices can be used to get an even more detailed perspective on needed change.

3.2 Ability to control exercised by Lufthansa in crisis situations

At Lufthansa, the Continuity Management process mentioned above helps to define different types of preventive, detective and correcting control measures to ensure IT and business continuity (ITCM/BCM). Depending on the nature of the measures, they detect the situation and act accordingly or monitor it with manual intervention as part of the measure. Concepts from the design of highly available systems, like hot, warm, and cold standby, are relevant here. The Risk Management process includes mitigations (i.e. corrective control). IT Service Management also contains the concept of service monitoring and improving, which can help to continuously detect errors or problems and identify corrective actions. Based on all these processes, the EA gets various inputs for control needs. The EA function translates this into requirements towards monitoring, supervision processes and the application behavior.

3.3 Crisis outcomes at Lufthansa

For the global aviation industry, crises generally have negative impacts resulting in losses (i.e. pandemics like SARS or COVID-19, volcano eruptions disrupting travel routes, or cyberattacks). In terms of Quality Management / IT Governance, the actions and measures defined by the practices described in previous sub-sections (Risk, Continuity, etc.) can be evaluated. Governance management is very helpful to gain some insights about the effect of the identified measures in the event of a crisis. With respect to IT Process Management, we need to have a view of the process landscape and see how the linkage of the different processes will help optimize them. At the very least it helps to identify missing links and/or gaps. The EA can thus be part of the Plan-Do-Check-Act cycle and helps continuously improve crisis outcomes.

3.4 Level of communication exercised within Lufthansa

Lufthansa employs various communication channels. These are in place during nomal business operations and can be used right away in times of crisis by posting current information, depending on release timelines and degree of formality required. For external communication, Lufthansa uses, amongst others, press conferences, social media channels, the company's public blog and press releases. For internal crisis communication, Lufthansa uses e-mail, web casts and the social intranet. In times of crisis the EA is installed as communicator and consolidates information provided by different business units. The EA is closely connected to the responsible parties for internal and external communication to distribute information as effectively as possible.

3.5 Distribution of crisis management in the context of Lufthansa

Lufthansa is a global organization, and it employs decentralized and centralized mechanisms for crisis management. The German headquarters have crisis management procedures in place to be implemented by subsidiaries around the world. However, each subsidiary has its own freedom to adapt these regulations to local standards and needs. This is the most effective approach. The central crisis management team cannot design a "one-size-fits-allßtrategy on crisis management, as each country has its own external and internal factors that need to be incorporated into its specific crisis response. EAs for global business lines incorporate regional differences in the crisis strategy plans to ensure enough freedom for local adaptions. Local EAs have to know local laws, regulations and best practices in order to tailor the strategy to the local needs.

3.6 Criticality of business functions applied to Lufthansa

As explained previously, BCM and ITCM are used to assess the criticality of an organizations business processes, functions and capabilities. These assessments are continuously applied by Lufthansa in order to get a comprehensive and up-to-date overview of its most critical business functions. Based on the outcomes of the BCM and ITCM assessment, initiatives can be started to conceptualize plans on how to ensure continuity of business functions in times of crisis and on how to prevent potential hazards for these business functions. The EA is both the consumer and producer of these assessments. In the role of producer, the EA is delivering valuable input for assessing a business function's criticality. As a consumer, the EA is using the output to start initiatives to ensure business function continuity.

3.7 Lufthansa's architectural approach to crisis management

The main benefit of our model is the ability to connect the dots across all model dimensions through the architectural perspective.

Business and strategy

The business and the strategy layer deliver important input to the EA. As described above, the assessment of the criticality of business processes and the vision of the strategy with regards to disruption, agility and reliability are translated into requirements for the IT landscape. The following example shows the link between these areas. Flight operations are vital to the business continuity of Lufthansa. Ensuring ongoing flights is the key focus of crisis management as flights are the biggest revenue stream. During a crisis, Lufthansa must at least cover the cost of operations in order to ensure long-term continuity. Booking and re-booking flights is considered as an essential business function. Irrespective of the crisis, it must still be possible to book flights in order to ensure ongoing flight operations.

Marketing campaigns are important business functions in order to maintain the brand image and associated value of Lufthansa. However, in times of crisis, marketing campaigns can be postponed—ultimately saving money.

Data

The classification of the data according to the required data protection level is important to identify the effective measures and actions for protecting the data layer. The EA's task is to translate these actions into concrete architectures (reference, domain, and solution) or alternatives, if possible. The assessment of the alternatives regarding the coverage of requirements and the cost efficiency is one of the main tasks of the EA. Measures applied at Lufthansa range from encrypting data at rest and in transit, setting up High Availability in order to optimize metrics such as Recovery-Time-Objective and Recovery-Point-Objective. An example for data classification at Lufthansa shows the following: the integrity of all flight-related data is crucial. Nevertheless, parts of flight data (e.g. arrival and departure times) have only low needs regarding confidentiality. Therefore, different measures of data protection need to be applied. Defining these measures based on the assessment is a main task of the EA.

Application

The application layer is another integral part that Lufthansa's EAs focus on. Their practices are characterized by implementing a lean application portfolio by applying different enterprise application portfolio management techniques. The most important thing for Lufthansa is to ensure high availability of applications (based on their criticality assessment) while getting rid of redundant applications. This ensures business continuity and cost reductions in the long run.

Technology

Together with assessing the crisis resistance of its infrastructure as part of the technology layer, Lufthansa is also evaluating its vendors as part of the overall risk management process. In times of crisis, a lack of support from vendors due to the unexpected circumstances can severely harm the application portfolio of an organization. Therefore, highly reliable, crisis-resistant technology vendors need to be considered during the buying decisions. For example, the infrastructure and applications cloud vendors need to show that their crisis management processes cover the requirements of an organization such as Lufthansa.

3.8 Future directions

Our ongoing research efforts are focused on identifying crisis management requirements for the IT landscape, as EAs have to do through the processes described earlier in the paper. We are following the usual steps as depicted by architecture development frameworks like TOGAF and therefore are developing the needed architecture building blocks serving

to fulfill all of the requirements. Based on the experiences reported in this paper and anticipated future developments, we have created a first set of requirements. These are comprised of: data backup and recovery (data must be recoverable - requirement derived from the BSI Grundschutz), application and server virtualization or containerization (applications must be recoverable by instantiating the image - requirement derived from ITCM and Risk Management), unlimited elasticity and financial flexibility (applications must provide scale up and down capability without boundaries in regards to workload and costs - requirement derived from Risk Management and IT-Strategy), resilience and fault tolerance (applications must able to cope with failures – requirement derived from ITCM and Risk Management), desktop and workplace virtualization (as well as the applications, the workplaces must be recoverable – requirement derived from ITCM and Risk Management), unified communication, collaboration and social media (specific communication channels and collaboration tools - requirement derived from communication needs and types).

4 Conclusions

In this paper, we build a model connecting EAM and crisis management through multiple dimensions that showcase crisis management responsibilities an EA can undertake, and then instantiate the model with insights from a large global aviation company, Lufthansa. Our work showcases a novel interdisciplinary model that connects crisis management concepts (derived primarily from the management and crisis and disaster research fields) with enterprise architecture concepts (derived primarily from IT research and practice). We hope that this model can help other organizations evaluate their enterprise architecture for crisis readiness, and show how enterprise architects can adopt crisis management responsibilities. Future research can apply this model to different organizations in other industries, and test its usefulness for different types of crises.

Bibliography

[Ah12] Ahlemann, F.; Stettiner, E.; Messerschmidt, M.; Legner, C. (Eds): Strategic Enterprise Architecture Management. Heidelberg, Germany: Springer, 2012.

[Au00] Augustine, N. R.: Managing the Crisis You Tried to Prevent. Harvard Business Review on Crisis Management. Harvard Business School Press, 2000.

[Bu] Bundesamt für Sicherheit in der Informationstechnik. Lektion 4: Schutzbedarfsfeststellung.

[Bu17] Bundy, J.; Pfarrer, M.D.; Short, C.E.; Coombs, W.T.: Crises and crisis management: Integration, interpretation, and research development. Journal of Management, 43/6, 2017, pp.1661-1692.

[Hi00] Hiatt, C. J.: A Primer for Disaster Recovery Planning in an IT Environment. Hershey, PA: Idea Group Publishing, 2000.

[Kl19] Klotz, M.: IT-Compliance nach COBIT 2019, 2019.

[Ko20] Kovoor-Misra, S.: Crisis Management Resilience and Change. Thousand Oaks, CA: SAGE Publications, 2020.

[KW17] Kuipers, S.; Welsh, N.H.: Taxonomy of the crisis and disaster literature: Themes and types in 34 years of research. Risk, Hazards & Crisis in Public Policy, 8/4, 2017, pp.272-283.

[La18] Laverdet, C.; Weiss, K.; Bony-Dandrieux, A.; Tixier, J.; Caparos, S.: Digital Training for Authorities: What is the Best Way to Communicate During a Crisis?. In Sauvagnargues, S. (Ed.), Decision-making in Crisis Situations, ISTE Ltd and John Wiley & Sons, Inc., 2018, pp. 149-174.

[PC98] Pearson, C.M.; Clair, J.A.: Reframing crisis management. Academy of Management Review, 23/1, 1998, pp.59-76.

[Su15] Sultanow, E.: Real World Awareness in kollaborativen Unternehmensprozessen. Berlin, Germany: Gito, 2015.

[Th18] The Open Group. The TOGAF Standard Version 9.2., 2018.

[Wa14] Watters, J.: Disaster Recovery, Crisis Response, and Business Continuity. Apress, 2014.

Neue Mobilitätsdienste und ihre Auswirkungen auf Unternehmensarchitekturen: Eine Fallstudie

Mark-Oliver Würtz[1], Kurt Sandkuhl [2]

Abstract: Neue Mobilitätslösungen (NML), wie Carsharing, City Bikes oder E-Scooter, werden in Großstädten immer beliebter und gewinnen zunehmend Einfluss darauf, wie sich Menschen im Stadtgebiet fortbewegen. Mittelfristig ist zu erwarten, dass die Zunahme der NML-Kundenanzahl auch auf öffentliche Verkehrsunternehmen und insbesondere deren IT-Landschaft Auswirkungen haben wird. Das Ziel dieses Papiers ist, dieses Thema aus der Perspektive der Unternehmensarchitektur (UA) zu beleuchten. Obwohl das UA-Management als relevant für öffentliche Verkehrsunternehmen anerkannt ist, fehlt es an Untersuchungen darüber, wie die NML die UA verändern. Der Zweck unserer Forschung ist es, einen Beitrag zu diesem Bereich zu leisten, indem wir die Relevanz des Themas aus der Geschäftsperspektive untersuchen. Der Hauptbeitrag unserer Arbeit ist (a) eine Literaturanalyse der Nutzung von UA im ÖPNV und für die Integration der NML, (b) Anforderungen bei einer NML-Integration aus Sicht von Betreiber, Kunde und Mobilitätsdienstleister mit Relevanz für UA, und (c) erste Anforderungen an die Integration von NML in UA.

Keywords: Digitale Transformation; Unternehmensarchitektur; ÖPNV; neue Mobilitätslösungen

1 Einleitung

Carsharing, City Bikes, e-Scooter, Bike-Sharing und viele weitere Mobilitätslösungen, die in größeren Städten in Pay-as-you-go- oder Abonnement-Modellen angeboten werden, sind immer beliebter geworden und haben auch begonnen, die Art und Weise zu beeinflussen, wie sich Menschen innerhalb kurzer und mittlerer Distanzen in Stadtgebieten bewegen [Ka16]. Diese Lösungen werden oft als neue oder innovative Mobilitätslösungen (NML) kategorisiert und waren Gegenstand der Forschung, z.B. im Hinblick auf neue Architekturen [Pf16], Geschäftsmodelle [Hi15], Plattformen [Ma15], Akzeptanz bei den Kunden [Ka16] oder erforderliche Standards. Wie sich neue Mobilitätslösungen auf ÖPNV-Unternehmen und insbesondere auf die IT-Landschaft in diesen Unternehmen auswirken, wurde bisher jedoch kaum erforscht.

Das Unternehmensarchitektur-Management (UAM) ist eine etablierte Funktion in vielen großen Unternehmen und zielt auf eine koordinierte und langfristige Entwicklung der Geschäfts- und IT-Aspekte eines Unternehmens ab [La17]. Architektur-Denken soll dazu beitragen, strategische Entscheidungen und nachhaltige Lösungen zu unterstützen [Wi14].

[1] Jellyco, Germany, mow@jellyco.de
[2] University of Rostock, Institute of Computer Science, Albert-Einstein-Str. 22, 18057 Rostock, Germany, kurt.sandkuhl@uni-rostock.de, https://orcid.org/0000-0002-7431-8412

doi:10.18420/inf2020_15

Obwohl UAM auch für den Bereich des ÖPNV als relevanter Ansatz anerkannt ist, fehlt es an Forschung darüber, wie sich neue Mobilitätslösungen auf die Unternehmensarchitektur öffentlicher Mobilitätsanbieter auswirken. Bislang gibt es nur wenige Leitlinien oder Vorschläge, wie Dienste und Informationssysteme, die NML erleichtern oder in eine bestehende UA einbinden, am besten integriert werden können.

Das Ziel unserer Forschung ist es, einen Beitrag zu diesem Bereich zu leisten, indem wir - in einem ersten Schritt - den Stand der Forschung sowie die Relevanz des Themas und Merkmale des Problems untersuchen. Die Leitfrage für die hier vorgestellte Arbeit lautet: *Wie wirken sich neue Mobilitätslösungen auf die UA von ÖPNV-Organisationen aus?* Der Hauptbeitrag unserer Arbeit ist (a) eine Literaturanalyse der Nutzung von UA im ÖPNV und für die Integration von NML, (b) Anforderungen bei einer NML-Integration aus Sicht von Betreiber, Kunde und Mobilitätsdienstleister mit Relevanz für UA, und (c) Anforderungen an die Unterstützung der Integration von NML in UA.

Das Papier ist wie folgt strukturiert: Abschnitt 2 fasst den Hintergrund unserer Arbeit zu NML und UAM zusammen. In Abschnitt 3 wird die genutzte Forschungsmethode vorgestellt. In Abschnitt 4 werden verwandte Arbeiten zu UAM im ÖPNV diskutiert und die zentralen Ergebnisse eines Experteninterviews dargestellt, dass in einem vorangegangenen Arbeitsschritt geführt wurde. Abschnitt 5 konzentriert sich auf die Analyse von Auswirkungen der NML-Integration auf UA im ÖPNV. Abschnitt 6 fasst Ergebnisse zusammen und diskutiert zukünftige Arbeiten.

2 Hintergrund und verwandte Forschung

2.1 Unternehmensarchitektur-Management

Das Gebiet der Unternehmensarchitektur (UA) [La17, Nu19] wird seit mehr als einem Jahrzehnt als eine Disziplin sowohl in der Forschung als auch für die praktische Unterstützung entscheidungsunterstützender Funktionen und Modelle in Unternehmen und Organisationen entwickelt [SFS14]. Ziel der Unternehmensarchitektur ist es, wichtige Zusammenhänge und Querbeziehungen zwischen Business und IT zu modellieren, abzustimmen und zu verstehen, um so die Voraussetzung für einen gut angepassten und strategisch ausgerichteten Entscheidungsrahmen sowohl für das digitale Business als auch für digitale Technologien zu schaffen [NP17].

UA Management, wie es heute durch verschiedene Standards wie ArchiMate [Ar17] und TOGAF [TO18] definiert ist, verwendet eine relativ große Anzahl unterschiedlicher Sichten und Perspektiven für die Verwaltung und Dokumentation der Business-IT-Ausrichtung (BITA) [SFS14]. UAM stellt einen Managementansatz dar, der ein kohärentes Set von Richtlinien, Architekturprinzipien und Governance-Regelungen festlegt, pflegt und verwendet, die Orientierung und Unterstützung bei der Gestaltung und Entwicklung einer Architektur bieten, um die Transformationsziele des Unternehmens zu verwirklichen. Ein effektiver

Architekturmanagement-Ansatz für digitale Unternehmen sollte zusätzlich die Digitalisierung von Produkten und Dienstleistungen unterstützen [UA19] und sowohl ganzheitlich als auch leicht anpassbar sein. Dies erfordert oft verteiltes Arbeiten, z.B. in Projektform oder auch mit unternehmens-externen Stakeholdern [HL18].

2.2 Neue Mobilitätslösungen

Neben dem klassischen Personennahverkehr, wie z.B. Bus, Straßenbahn, U-Bahn, Fähren und Taxen, haben sich seit dem Beginn der Digitalisierung sogenannte Neue Mobilitätslösungen Services (NML) zum Teil im Rahmen von sogenannten Mobility as a Service-Ökonomien (MaaS) etabliert. Diese erweiterten den klassischen Personennahverkehr um die Transportkonzepte *Car-Sharing*, *Ridesharing*, *Bike-* und *Scooter-Sharing*. Zukünftig werden hierzu auch noch neue Technologien hinzukommen, wie z.B. autonome Fahr- und Flugdienste. Diese Dienste haben in den letzten 2 Jahren in Deutschland (bzw. weltweit) einen enormen Aufschwung erlebt. Grund für die schnelle Ausbreitung waren Innovationen, die das Produkt durchlaufen hat. Intelligente Schlösser und Entwicklungen in der Geo-Lokalisierungstechnologien haben dazu geführt, dass Fahrräder und Tretroller nicht nur stationär an bestimmten Punkten angeboten werden können, sondern frei schwimmende Flotten in den Städten entstanden sind, die ohne große zusätzliche Infrastruktur voll funktionsfähig betrieben werden können.

Der Markt fängt gerade, an NML-Transportkonzepte grob in die zwei Kategorien zu unterteilen, stationsgebundene und nicht stationsgebundene Angebote. Dies kann man bei allen oben vorgestellten Konzepten anwenden. Dabei basieren diese neuen Dienste auf bestehenden Transportmitteln, die entweder mit Hilfe der Digitalisierung (a) von vielen unterschiedlichen Personen zeitlich entkoppelt voneinander genutzt werden, oder (b) dass individualisierte Angebot kurzfristig einer möglichst großen Zielgruppe transparent gemacht wird und dazu führen, dass sich mehrere Personen ein Beförderungsmittel (Mitfahrdienste) teilen. Dabei ist die Herausforderung, diese Dienste so einfach wie möglich zu gestalten und in eine bestehende Mobilitäts-Informationsinfrastruktur zu integrieren, um möglichst viele potentielle Kunden einer Region zu überzeugen bzw. zu erreichen. Im zweiten Fall sind zukünftig im Besonderen die ÖPNV-Unternehmen gefordert, um diese Anforderung zu realisieren. Erst mit einer weitreichenden Integration wird der Personennahverkehr in der Lage sein, multimodale Mobilität für alle Anbieter und Kunden einfach und komfortabel zu gestalten.

Bei Mobility-as-a-Service (MaaS) handelt es sich um einen integrierten Mobilitätsdienst, der verschiedene Formen von Verkehrsdiensten integriert. Dabei bietet MaaS dem Benutzer einen Mehrwert, indem es für die Nutzung der Mobilität eine einzige Anwendung (App) zur Verfügung stellt, die statt mehrerer unterschiedlicher Ticketing- und Zahlungssysteme (der NML-Anbieter) die Möglichkeit eines einzigen Ticketing- und Zahlungssystem zur Verfügung stellt. Dabei integriert der MaaS-Betreiber für die Kundenzufriedenheit verschiedenste Transportoptionen unter seiner Plattform (z.B. als Mobility Plattform). Ein Erfolgsfaktor

einer solchen Plattform ist die Integration vieler Beförderungskonzepte, wie z.B. öffentlicher Personennahverkehr, Mitfahr-, Auto- oder Fahrradmitbenutzung, Taxi, Autovermietung oder Leasing die entweder autark oder in Kombination zu berücksichtigen sind. Ein weiterer Erfolgsfaktor ist die Schaffung von neuen Geschäftsmodellen und Möglichkeiten für die Unternehmen, Betrieb und Organisation der Transportoptionen optimal zu gestalten. Dabei steht der Zugang zu aktuellen bzw. besseren Informationen (ÖPNV-Daten, Verkehrsdaten, Wetterdaten usw.) für den Kunden im Mittelpunkt der Betrachtung.

Ziel von MaaS ist es einen möglichst hohen Wertbeitrag für seine Nutzer zu leisten, damit sie eine echte Alternative zur Nutzung eines PKWs bietet, die effizienter, nachhaltiger und in vielen Fällen auch günstiger sein kann. Während des laufenden Übergangs zur Mobilität als Dienstleistung (MaaS) sind daher die Integration von möglichst vielen bzw. allen Transportanbietern eines Stadtgebiets zu gewährleisten, um es dem Nutzer zu ermöglichen, multimodale Stadtfahrten zu planen und durchzuführen.

Eine zentrale Herausforderung, die sich MaaS-Anbieter bzw. Unternehmen gemeinsamer Mobilität stellen müssen, ist eine fragmentierte Regulierungslandschaft im kommunalen Sektor. Diese macht es entsprechenden Unternehmen nicht einfach, neue Geschäftsmodelle zu entwickeln und diese erfolgreich zu betreiben. Eine Förderung solcher MaaS-Ansätze auf lokaler Ebene als Best Practices zu fördern wäre daher sinnvoll. Dabei ist zu berücksichtigen, dass gerade im deutschsprachigen Raum die ÖPNV-Unternehmen zu meist Anstalten öffentlichen Rechts - mit zumeist guten Beziehungen in die regionale Politik- sind und in ihrer Region einen optimalen Überblick über den Personennahverkehr haben. Diese Argumente sprechen dafür, dass Betreiber von MaaS-Ansätzen in einer Region die ÖPNV-Unternehmen selbst sind.

3 Forschungsmethodik

Die in diesem Beitrag vorgestellten Arbeiten sind Teil eines Forschungsprojektes, das auf methodologische und instrumentelle Unterstützung für die Integration von NML in die UA von ÖPNV-Organisationen abzielt. Das Projekt folgt dem Paradigma der konstruktionsorientierten Forschung (Desgn Science Research - DSR) [JP14]. Die in diesem Papier vorgestellten Arbeiten behandeln die Analyse der Problemrelevanz und die Ermittlung von Anforderungen an das zu entwickelnde Artefakt: eine UA-methodisch-technologische Unterstützung für die Integration von NML. Der in diesem Beitrag vorgestellte Teil unserer Forschungsarbeit ging von der folgenden Forschungsfrage aus, die auf der in Abschnitt 1 dargestellten Motivation beruht: *FF: Welche geschäftlichen Herausforderungen sind im Zusammenhang mit der Einführung von NML in der betrieblichen Praxis sichtbar und welche Anforderungen bestehen an eine methodisch-technologische Unterstützung zur Integration von NML in UA?*

Die zur Bearbeitung dieser Forschungsfrage angewandte Forschungsmethode ist eine Kombination aus Literaturstudie, Fallstudie und argumentativ-deduktiver Arbeit. Ausgehend

von der Forschungsfrage haben wir zunächst entsprechende Literatur analysiert. Ziel der Analyse war es, bestehende Ansätze oder Theorien zur Unterstützung der NML-Integration und Fallstudien aus der Praxis zu finden, die es uns ermöglichen, die Problemrelevanz im Detail zu untersuchen. Da die Literaturstudie einen Mangel an etablierten Ansätzen zeigte (siehe Abschnitt 4.1), haben wir mittels einer Fallstudie, in deren Rahmen auch Experteninterviews durchgeführt wurden, das Problem und die daraus resultierenden Anforderungen an eine Lösung analysiert (siehe Abschnitt 4.2).

Das Experteninterview basierte auf Richtlinien und Fragen, die auf der Grundlage des obigen FF entwickelt wurden. Der Interviewte war der Leiter des UAM einer ÖPNV-Organisation in einer deutschen Großstadt. Die gewonnenen und dokumentierten Ergebnisse wurden einer qualitativen Inhaltsanalyse nach Mayring [Ma10] unterzogen. Diese Methode dient dem Zweck der empirischen Datenauswertung und enthält Empfehlungen für eine systematische und überprüfbare Textanalyse. Auf der Grundlage der Ergebnisse aus dem Interview argumentieren wir, dass für eine erfolgreiche Integration der NML in die UA von ÖPNV-Unternehmen methodische Unterstützung erforderlich ist. Um die Anforderungen an diese Unterstützung zu ermitteln, haben wir die Sicht verschiedener Stakeholder im Rahmen der Fallstudie zusammen mit dem Experten analysiert (s. Abschnitt 5). Diese Argumentation benötigt weiteres unterstützendes Fallmaterial, um unsere Schlussfolgerungen in der zukünftigen Arbeit zu untermauern.

4 UAM im ÖPNV

Die aktuelle Situation zu UAM im ÖPNV und den Bedarf eventueller Veränderungen haben wir zum einen mittels einer Literaturanalyse untersucht (s. 4.1). Zum anderen haben wir ein Experteninterview in einem bei einem größten deutschen ÖPNV Anbieter durchgeführt, um die Perspektive der Praxis in unsere Arbeiten zu integrieren (s. 4.2).

4.1 Literaturanalyse

Am Anfang der Recherche stützten sich die Autoren auf die Veröffentlichung von Scholz „IT-Systeme für Verkehrsunternehmen“ [Sc16], welches das Branchenmodell (Domain-Modell) ITTC Core Model (dt. ITVU) für den ÖPNV beinhaltet. Dieses Modell beschreibt Geschäftsprozesse und fundamentale Datenstrukturen (Klassen), die ineinander zugeordnet in Paketen (Buildingblocks) dargestellt werden. Die Auseinandersetzung mit der Substitution von ineffizienten Personentransportsystemen durch andere Verkehrsformen ist so alt, wie das Auto selbst. Hervorzuheben ist, dass Heinze und Kill 1994 in „Zukunftsfähige Strategien für den ÖPNV in Berlin-Brandenburg“ [HK94] sowie Beutler und Brackmann in „Neue Mobilitätskonzepte in Deutschland: ökologische, soziale und wirtschaftliche Perspektiven“ [BB99] schon Themen, die uns noch heute beschäftigen, diskutierten.

Neben der Diskussion von NML sind Produkte auf Grundlage neuer Technologien entstanden, die dem Ansatz eines „Mobility as a Service“ (MaaS) folgen. Diese Mobilitätsintegration wurde 2016 in [Ka16] kritisch bewertet. Ergebnis dieses Papiers war, dass MaaS eine „. . . vielversprechende Mobilitätslösung darstellt und voraussichtlich einen bedeutenden Beitrag zur künftigen städtischen Reform darstellen wird.“ [Ka16]. Zudem haben [Hi15] ein nachhaltiges Geschäftsmodell anhand einer Carsharing Dienstleistung skizziert. Dabei war das Ziel des Papiers, ökonomische und ökologische Gesichtspunkte in einem alternativen Geschäftsmodell mit zu vereinbaren. Eine Veröffentlichung mit ähnlicher Intention ist [Al17].

In einem zweiten Schritt der Literaturanalyse wurde nach der Einordnung von NML, MaaS und Mobility Plattform in das Enterprise Architecture Management Verbindung stehenden IT-relevanten Themen gesucht. [Pf16] skizzieren die Architektur einer offenen und modularen Dienstleistungsplattform für mobile Dienste für eine zukünftige Smart City. Schlussfolgerung des Papiers ist, dass durch die Bereitstellung von Daten und Diensten für den Betrieb von Mobility Services eine noch stärkere Innovation bei den Mobilitätsdienstleistern ausgelöst wird. [Ba13] beschäftigt sich mit dem Parken von Autos im urbanen Umfeld und den möglichen Steuerungsmöglichkeiten in einem modernen Mobilitätskonzept. [Ma15] beschreibt eine intelligente Mobilitätsplattform, welche als Kernkomponente eine kartenbasierte Plattform besitzt. Die Veröffentlichung, die an das Thema unserer Arbeit am nächsten heranreicht, hat den Titel „Internet of Things within the Service Architecture of Intelligent Transport Systems“ [LDI17]. Auf Grundlage von IoT-System-Technik wird eine Enterprise Architektur für das St. Petersburger Lagezentrum erörtert. Hierbei geht es primär um Überwachung des Verkehrs und der Verkehrsteilnehmer und es wird zumeist auf IT-Bestandteile und weniger auf Geschäftsprozesse verwiesen.

Alle hier aufgezeigten Artikel und Papiere beinhalten wichtige Teillösungen für das Managen eines integrativen Mobility-Konzepts für Unternehmen des ÖPNV. Eine umfassende Plattform, die alle wichtigen Funktionen umfasst, eine UA vorschlägt, oder beantwortet, wie ein ÖPNV-Unternehmen NML in seine bestehende Unternehmens-Architektur integrieren könnte, wurde im Laufe der Recherche nicht gefunden.

4.2 Experteninterview

Durch die Literaturanalyse sehen wir unsere Vermutung bestätigt, dass das Thema UAM für den ÖPNV und insbesondere die Integration der NML in UAM bisher in der Forschung keine nennenswerte Aufmerksamkeit erfahren hat. Ein Experteninterview mit dem UAM-Verantwortlichen der Berliner Verkehrsbetriebe BVG hatte zudem bestätigt, dass – zumindest aus Sicht der BVG – Unternehmensarchitekturen eine wichtige Rolle bei der Integration spielen können. Konkret wurden im Experteninterview vier Hypothesen untersucht und bestätigt (s. [WS20]):

Hypothese 1: Digitalisierungsvorhaben, wie NML wirken sich auf die UA von ÖPNV-Unternehmen aus.

Die Einführung von NML in einem ÖPNV-Gebiet führen, ob nun selbst durch das ansässige ÖPNV-Unternehmen oder durch einen Fremdanbieter initiiert, immer zu einem Impakt auf das ÖPNV-Unternehmen selbst - und wenn es nur zu Neu- bzw. Umplanung im Angebot des ÖPNV-Unternehmen führt. Dabei ist die Frage, ob dies schon einen signifikanten Einfluss auf die Geschäftsprozesse bzw. IT-Prozesse hat und somit die Enterprise Architektur betrifft. Grundsätzlich konnte festgestellt werden, dass durch die Digitalisierung folgende Bereiche bei der BVG anfangs besonders betroffen waren: Vertrieb (inkl. Angebot), IT und Verwaltung. Zusehends wurde davon auch das Kerngeschäft (Fahrdienstleistungen: U-Bahn, Straßenbahn, Bus) und Personal erfasst.

Hypothese 2: NML bilden einen disruptiven Ansatz gegenüber dem klassischen ÖPNV.

Ob dem Kerngeschäft der BVG mit der Einführung weiterer *NML, wie z.B. UBER, (zukünftig) Flugtaxen oder auch White Sharing-Anbieter wie z.B. Fleetbird mit Emmy,* ein disruptiver Einbruch droht, wollte der Experte nicht bestätigen bzw. bezog dazu keine klare Stellung. Anders sieht der Befragte dies bei B2B-Geschäftsmodellen, wie z.B. Mobilitätsplattformen mit eigenen Preismodellen. Hier sieht der Bereichsleiter sehr wohl Gefahren auf den ÖPNV zukommen, da wenn man sich z.B. die Anbieter „. . . anschaut, zum Beispiel die NOW-Gruppe mit Transdev zusammen, die im B2B-Geschäft ganz neue Preismodelle anbieten, dann ist das, aus . . . " seiner „. . . Einschätzung heraus, eine unmittelbare Bedrohung, des originären BVG Geschäftsmodells." [Me20].

Hypothese 3: UAM hilft bei der Digitalisierungsvorhaben, bzw. bei der Einführung neuer NML, in ÖPNV-Unternehmen.

UAM und im speziellen TOGAF als EAM-Framework stellen nach Aussage des Befragten ein gutes Hilfsmittel dar, um strategische Applikationen bzw. IT-Landschaften zu managen. Es bringt die IT und die Fachbereiche zusammen und ist ein probates Mittel, um zu einer gemeinsamen Sicht (IT und Fachabteilung) zu gelangen. Darüber hinaus erzeugt es die Transparenz zwischen der aktuellen Situation (IST) und dem zukünftig vom Business benötigten (SOLL) in der IT - so dass das IT-Alignment im Unternehmen nach den Vorgaben der Geschäftsstrategie gesteuert und gewährleistet werden kann. Dabei spielt es keine Rolle, ob die Applikationen/IT-Landschaft im Haus oder außer Haus betrieben wird. Auch außen gehostete Applikationen werden durch Betreuung/Prozesse bzw. spätestens auf Prozessebene im EAM „Fußabdrücke" hinterlassen.

5 Herleitung von Anforderungen an UAM zur Integration von NML

Durch die Literaturanalyse (s. 4.1) sehen wir unsere Vermutung bestätigt, dass das Thema UAM für den ÖPNV und insbesondere die Integration der NML in UAM bisher in der Forschung keine nennenswerte Aufmerksamkeit erfahren hat. Das Experteninterview mit

dem CIO der Berliner Verkehrsbetriebe BVG (s. 4.2) hat zudem bestätigt, dass – zumindest aus Sicht der BVG – Unternehmensarchitekturen eine wichtige Rolle bei der Integration spielen können. Wir konzentrieren uns daher in diesem Kapitel darauf, die sich aus der NML-Integration ergebenden Auswirkungen auf UA herauszuarbeiten. Dazu werden im ersten Schritt wichtige Stakeholder bei der NML-Integration (Betreiber, Mobilitätsdienstleister, Kunde) analysiert, um im zweiten Schritt die daraus resultierenden Anforderungen herleiten zu können.

5.1 Anforderungen der Stakeholder Betreiber, Mobilitätsdienstleister, Kunde

In diesem Absatz werden die zentralen funktionalen Anforderungen an eine Mobilitätsplattform geordnet nach den drei wichtigsten Stakeholdern Betreiber der Plattform, Mobilitätsdienstleister (z.B. ÖPNV-Unternehmen) und Kunde (der Mobilitätsplattform) aufgeführt. Aus diesen drei Rollen wird bespielhaft eine ausgewählt und im Kontext des EAM beleuchtet. Nicht-funktionale Anforderungen bleiben in diesem Papier unbeachtet, da diese Punkte, wie Flexibilität, Robustheit usw. beinhalten und in einer späteren gesonderten Betrachtung für beispielsweise die Skalier-Fähigkeit des Modells einer Mobilitätsplattform herangezogen werden.

Tab. 1: Beispiele für Anforderungen aus Sicht des Betreibers

Bereitstellung einer Kundenschnittstelle	Bereitstellung einer APP, die den Kundenanforderungen entspricht (siehe Kundenanforderungen);
Debitoren Verwaltung	Kundenverwaltung
Kreditoren Verwaltung	Mobilitätsdienstleister (Lieferantenverwaltung)
Service Bereitstellung	Unterstützung des Kunden bei allen Transaktionen
Faktura	Fakturierung der vom Kunden gebuchten Verbindungen.
Rechnungsdaten	Abrechnen des Produkts/Tarif, der vom Kunden genutzten Verbindungen mit dem externen Mobilitätsdienstleister. (Vertraglich hinterlegte Preisinformationen)

Tab. 2: Beispiele für Anforderungen aus Sicht des Mobilitätsdienstleister

Tarifdaten	Eigene Tarifdaten für Mobilitätsplattform bereitstellen.
Kundendaten	Abgleich mit der Mobilitätsplattform.
Verbindungen	Bereitgestellte Verbindungen und Einsatz-/Aussetzpunkte
Geleistete Einsatzdaten	Geleistet: Einsatzort, Einsatzzeit und Fahrtziel
Rechnungsdaten	Für die Faktura des Mobilitätsdienstleisters.

Ein klassisches ÖPNV-Unternehmen kann einerseits als reiner Mobilitätsdienstleister bzw. zusätzlich mit der Rolle als Betreiber einer Mobilitätsplattform am Markt auftreten. Es wird schwer darstellbar sein, dass ein Träger öffentlichen Rechts, was ein großer Teil der ÖPNV-Unternehmen ist, nur die Rolle als Mobilitätsplattformbetreiber innehat.

In unserem Fall betrachten wir ein ÖPNV-Unternehmen, welches die Kombination aus Mobilitätsdienstleister und Mobilitätsplattformbetreiber darstellt bzw. darstellen möchte.

Tab. 3: Beispiele für Anforderungen aus Sicht des Kunden/Fahrgäste

Informieren über:	Tarife, Produkte und Verbindungen (von-/nach-; Uhrzeit); Verkehrssituation/-prognose; Parkplatzsituation/-prognose; Ergänzende Mobilitätsdienste; angrenzende Produkte des Fernverkehrs;
Registrieren / Einloggen	Kunde muss sich beim (ersten) Aufruf des Service registrieren/bzw. einloggen.
Buchen	Einer Verbindung bzw. eines Mobilitätsmittel/-angebot.
Zahlen	Zahlung der ausgesuchten Verbindung.
Reisebegleitung	Mitfahrregeln (bei einigen Produkten), Gruppenangebote usw.
Nach-Reise Service	z.B. Fundsachen, Kundenzufriedenheit, usw.

Hierbei ist vorrangig gemeint, dass mit dem Betrieb der Mobilitätsplattform primär der Support und der „fachliche Betrieb“ gemeint sind. Der technische Betrieb kann intern mit eigenen Ressourcen oder extern durch ein Full-Service Provider erfolgen. Auch wenn der Betrieb durch ein Full-Service Provider erfolgt, muss die IT des ÖPNV-Unternehmens wissen, wie die (technische) Architektur aussieht, um eine erfolgreiche Integration in eine bestehende Architektur Landschaft vollziehen zu können.

5.2 Auswirkungen auf die UA im ÖPNV

In der UA des betrachteten Unternehmens ist deutlich erkennbar, dass Daten- und Anwendungsarchitektur vor allem bei speziellen Funktionen der Betriebsführung und -abrechnung einem Transformationsprozess unterzogen werden müssen, der vor allem durch fehlende Schnittstellen und erforderliche Anpassungen in der Datenarchitektur verursacht wird. Am Beispiel des App-Hauptprozess Mobilitätsplattform, aus Sicht der Kunden dargestellt, werden im Folgenden die Auswirkungen auf eine Architektur, die keine Mobilitätsplattform betreibt, aufgezeigt:

1. Registrieren/Einloggen: Standard Vorgehen (wie bei vielen Apps) - Da in der App/Homepage personenbezogene Daten (inkl. Kontodaten) ausgetauscht werden, sollte das Frontend und das Backend besonders gut gesichert sein.
2. Verbindung suchen: Hierfür muss der Mobilitätsplattform aller angeschlossenen Partnern deren Tarif- bzw. Produktinformationen vorliegen. Ein System, welches nicht die gewünschten Verbindungen des Kunden suchen, auflisten und preislich darstellen kann, wird so gut wie keine Akzeptanz beim Kunden erlangen.
3. Verbindung buchen: Entscheidet der Kunde (Fahrgast) sich für eine passende Verbindung, so kann er diese direkt -aus der Verbindungsansicht heraus- buchen.
4. Beförderung (be-) zahlen: Nach der Buchung wird eine Zahlungstransaktion ausgelöst. Hat der Kunde bei der Registrierung seine Zahlungsdaten noch nicht angegeben, so wird er nach der ersten Buchung aufgefordert, dieses nachzuholen.

5. Service bewerten: Nach der Nutzung der Verbindung (Fahrt) besteht die Möglichkeit, dass der Kunde die Verbindung (im Ganzen) bewertet.

6. Fahrschein Kontrolle: Eine (Digitale- oder Sicht-) Kontrolle eines vom Kunden erworbenen digitalen Fahrausweis muss gewährleistet sein.

In der Prozesslandschaft (Level 0) eines ÖPNV-Unternehmens (siehe Abb. 1) stehen traditionell primär entsprechende Änderungen im Bereich der Verkehrsleistungen bzw. in dessen Betriebsführung im Fokus. Sieht man sich nun den (IST-) Geschäftsprozess in Abhängigkeit zu den ÖPNV-Produkten an und ordnet diesen die verwendeten IT-Systeme zu, so kann man erkennen, welchen Impakt die Erweiterung des Verkaufsweg mit Fremdprodukten an welcher Stelle hat. In Abb. 2 ist beispielhaft der Ergebnistyp „Geschäftsprozess/ Produkt zu IT-System“ grafisch dargestellt.

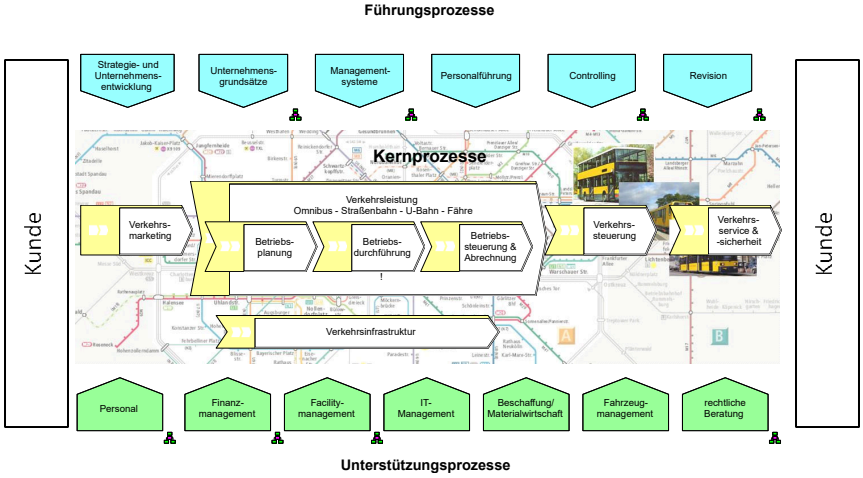


Abb. 1: Prozesslandschaft eines ÖPNV-Unternehmens (Level 0)

An Abbildung 2 lässt sich schnell erkennen, welches ÖPNV-Produkt und dem dazugehörigen Prozess vom IT-System unterstützt wird. Hiermit ist das IT-Management in der Lage schnell zu erkennen, welches IT-System betroffen ist. Das IT-Management ist mit Hilfe der jeweiligen Experten (im Haus/Extern) in der Lage, ob die neuen Anforderungen an das IT-System erfüllt werden kann oder ob: (a) das IT-System erweitert/angepasst bzw. (b) ein neues IT-System beschafft/erstellt werden muss. Dabei ist zusätzlich gewährleistet, dass man sofort erkennt, ob die Änderung weitere Auswirkungen auf andere bzw. weitere, vom selben IT-System unterstützte, Prozesse hat. Enterprise Architecture Management bietet hierdurch die Transparenz, um kurzfristig Entscheidungen hinsichtlich der Gesamtarchitektur treffen zu können.

Im obigen Fall würde das IT-Management die Rolle des Mobilitätsplattformbetreibers und deren Anforderungen der Produkt-Prozess-IT-System Matrix gegenüberstellen. Dabei kann man erkennen, dass z.B. im Quadranten „Self Service Verkauf“ und „Mobiles Ticket“

schon eine App (IT-System) vom analysierten ÖPNV-Unternehmen betrieben wird. Um das Produkt und den Prozess bedienen zu können wird im aktuellen Szenario noch zwei weitere IT-Systeme benötig: SAP PT120-Kontokorrent (Nebenbuchhaltung) und PTcom (Importschnittstelle für Buchhaltungsdaten). Die Anforderungen, die an einen Mobilitätsplattformbetreiber gestellt werden, können nun gezielt auf die Bestandssysteme „gemappt" werden. Hierbei steht außer Frage, dass die bis jetzt verwendete App, die bisher die Verwaltung und Abrechnung eines Mobilitätsdienstleister ermöglichte, nicht die Anforderungen erfüllt, die an einem Betreiber (ggf. auch in Kombination als Mobilitätsdienstleister) gestellt werden. In Tabelle 4 werden beispielhaft die Auswirkungen auf die IT-Systeme skizziert und grob bewertet.

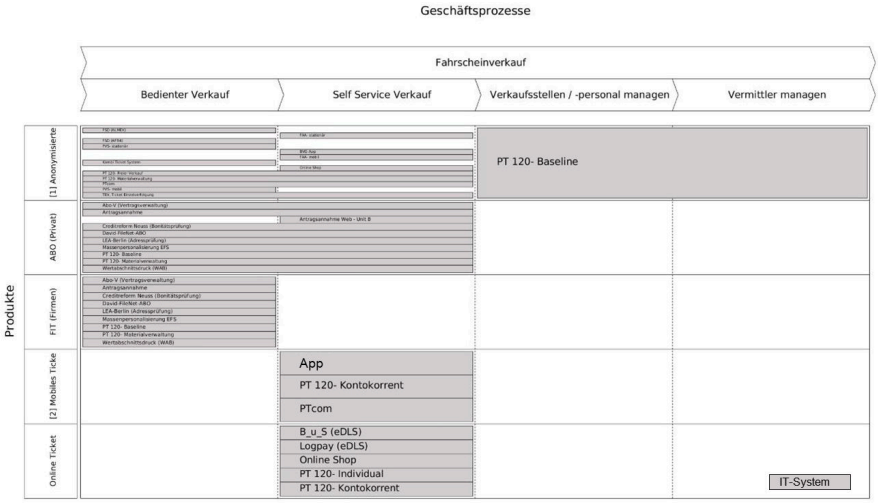


Abb. 2: Bsp. Geschäftsprozess „Fahrscheinverkauf" (Level 2) eines ÖPNV-Unternehmens

Zusammengefasst führen gegenseitige Abhängigkeiten zwischen Geschäfts-, Daten- und Anwendungsarchitektur zu Transformationsbedarf in bestimmten Ausschnitten der UA. So kann, z.B. die Flexibilisierung von Ticketlösungen (Erwerb einer Fahrerlaubnis) dazu führen, dass nicht nur klassische ÖPNV-Fahrberechtigungen, z.B. in Form einer Waben- oder Zonenstruktur verarbeitet werden muss, sondern auch Fremdanbietersysteme, wie z.B. Taxen, Car-Sharing, Bike-Sharing usw., die pro Mieteinheit, bzw. Entfernung oder in einer Kombination aus Beiden abgerechnet werden müssen. Dies führt in diesem Fall nicht nur zu Anforderungen an das IT-System, sondern auch an die Datenstruktur und an die Prozesse. Denn einerseits müssen unterschiedlichere Daten gehalten, verarbeitet und gelöscht werden sowie andererseits müssen Personen diese unterschiedlichen Informationen auch prüfen können, welches zur Folge hat, dass z.B. ein eigener Prozess für die Klärung von Störungen in Fremddaten (externen Anbietern) geschaffen werden muss.

Tab. 4: Beispiele für Anforderungen aus Sicht des Betreibers

IT-System	Auswirkung	Aufwandsabsch.
App	Eine allen Anforderungen gerecht werdende App muss für diesen Fall neu entwickelt bzw. eingeführt werden. **Neu-Entwicklung.**	Hoher Aufwand (> 200 PT)
SAP PT-120 Kontokorrent	PT-120 verfügt über die Funktionalität, Multibetreiber Plattformen und Abrechnungen von Kunden jeglicher Art zu managen sowie über eine Tarifverwaltung, die die benötigten Tarifdaten/Anbieter verwalten könnte. **Konfiguration und kleinere Erweiterungen.**	Mittlerer Aufwand (< 200 PT)
Ptcom	Die Schnittstelle zu SAP-PT120 müsste um einen weiteren Vertriebskanal angepasst werden. Da Feldinformationen variabel anlegbar sind, können auch komplexe Abrechnungskonzepte abgebildet werden. **Konfiguration.**	Geringer Aufwand (< 50 PT)
Zusätzlich benötigte Funktion (Tarif- bzw. Produktinformationen)		
SAP-PT120 Baseline	Das PT-120 System (Module Baseline) verfügt darüber hinaus über eine Tarif-verwaltung aus der man die benötigten Tarifdaten/Anbieter verwalten könnte. **Konfiguration und kleinere Erweiterungen.**	Mittlerer Aufwand (< 200 PT)

6 Schlussfolgerungen

Die dargestellten Ergebnisse führen zum Schluss, dass die erfolgreiche Einführung von neuen Mobilitätsdienstleistungen z.Z. weniger bei den öffentlichen Verkehrsbetrieben stattfindet, sondern von Newcomer (wie z.B. Emmily u.w.) etabliert werden. Dabei sind die Themen, die die etablierten ÖPNV-Unternehmen treibt eher die Ausrichtung am klassischen ÖPNV-Geschäft. Damit ÖPNV-Unternehmen auch langfristig überleben können, müssen sie selber in der Lage sein, Mobilitätsplattform in ihr Unternehmen zu integrieren und zu managen. Hierbei ist es egal, ob sie selbst den Betrieb machen oder dies durch Dienstleister übernehmen lassen. Wichtig hierbei ist, dass die „Owner-Schaft“ für die ÖPNV-Unternehmen gesichert sein muss. Für dieses Ziel sollten die klassischen ÖPNV-Unternehmen wissen, wie eine mögliche Zielarchitektur für Mobilitätsplattformen im Idealfall aussehen, welche Prozesse damit verbunden sind und wie sie die Schnittstellen, Produkte und Dienstleistung so offen gestalten, dass jederzeit weitere Anbieter von Mobilitätssystemen mit einsteigen können.

Das oben dargestellt Bespiel soll zeigen, dass gerade in Zeiten der Digitalisierung EAM Methoden zur Verfügung stellt, die das IT-Management in die Lage versetzt, diese Komplexität effizient zu „managen“. Dies wiederum trägt dazu bei, dass die Unternehmen (hier: ÖPNV) in der Lage sind, Anpassungsbedarf, der durch die Strategie bzw. Governance gefordert wird, schnell zu erkennen und nötige Maßnahmen für eine Umsetzung ggf. zügig einzuleiten. In diesem Fall heißt das, IT-Systeme, über die Eingrenzung von Prozess und Produkt, zu identifizieren, die von neuen Anforderungen betroffen sind und für ein systematisches EA Content Management zu sorgen [Zi19].

Im Kontext der vorgestellten Flexibilisierung von Ticketlösungen wurde dargestellt, dass nicht nur die Darstellung von IT-Systemen und Prozessen den ausschließlichen Vorteil des Enterprise Architecture Management darstellt, sondern auch die Wechselwirkung zu Datenstrukturen, Infrastrukturelementen u.w. Informationen, die es ermöglichen, die Komplexität der Veränderung, die durch die Digitalisierung angestoßen wird, zu beherrschen. Dabei ermöglicht das EAM den Beteiligten die Beantwortung wichtiger Fragen, wie z.B.: Woraus speisen sich die IT-Komplexitätstreiber? Oder Welche Maßnahmen benötigen wir zur Bewältigung der technischen Komplexitätsbeherrschung?

Diese zwei Fragen sollten in einer weiteren Studie untersucht werden. EAM bietet darüber hinaus noch viele weitere Methoden, die das „managen“ der Gesamtarchitektur vereinfacht bzw. erst ermöglicht. Eine entsprechende Referenzarchitektur über den entsprechenden (Funktions-) Bereich würde zusätzlich die Identifikation der Änderungspunkte deutlich erleichtern und dem jeweiligen ÖPNV-Unternehmen eine am „Best Practice“ orientierte Lösungsarchitektur vorschlagen.

Literaturverzeichnis

[Pf16] Pflügler, C., Schreieck, M., Hernandez, G., Wiesche, M., & Krcmar, H.: A concept for the architecture of an open platform for modular mobility services in the smart city. Transportation Research Procedia, 19, 199-206, 2016.

[Hi15] Hildebrandt, B., Hanelt, A., Piccinini, E., Kolbe, L., & Nierobisch, T.: The Value of IS in Business Model Innovation for Sustainable Mobility Services-The Case of Carsharing. In Wirtschaftsinformatik (pp. 1008-1022), 2015.

[Ma15] Marchetta, P., Natale, E., Pescapé, A., Salvi, A., & Santini, S.: A map-based platform for smart mobility services. 2015 Symposium on Computers and Communication (ISCC) (pp. 19-24). IEEE, 2015.

[Ka16] Kamargianni, M., Li, W., Matyas, M., & Schäfer, A.: A critical review of new mobility services for urban transport. Transportation Research Procedia, 14, 3294-3303, 2016.

[La17] Lankhorst, M.: Enterprise Architecture at Work. Modelling, Communication and Analysis. The Enterprise Engineering Series. Springer Berlin, Heidelberg, 2017.

[Wi14] Winter, R.: Architectural thinking. Business & Information Systems Engineering, 6(6), 361-364, 2014.

[JP14] Johannesson, P., & Perjons, E.: An introduction to design science. Springer, 2014.

[Ma10] Mayring, P.: Qualitative Inhaltsanalyse. In Handbuch qualitative Forschung in der Psychologie (pp. 601-613). VS Verlag für Sozialwissenschaften, 2010.

[RE16] Research Service, European Parliament: Uber under the hood, 2016; UITP, 2016.

[AP16] APTA: CityAM, 2016; Orb International, 2017; ADEME, 2016

[Nu19] Nurmi, J., Pulkkinen, M., Seppänen, V., Penttinen, K.: Systems Approaches in the Enterprise Architecture Field of Research: A Systematic Literature Review. EEWC 2018 Proceedings, vol. 334. LNBIP, vol. 334, pp. 18–38. Springer, Cham, Switzerland, 2019.

[SFS14] Simon, D., Fischbach, K., Schoder, D.: Enterprise architecture management and its role in corporate strategic management. Inf Syst E-Bus Manage 12(1), 5–42, 2014.

[NP17] Niemi, E., Pekkola, S.: Using enterprise architecture artefacts in an organisation. Enterprise Information Systems 11(3), 313–338, 2017.

[Ar17] The Open Group: ArchiMate® 3.0.1 Specification, 1st edn. Van Haren Publishing, Zaltbommel, 2017.

[TO18] The Open Group: The TOGAF Standard, version 9.2. TOGAF series. Van Haren Publishing (2018)

[UA19] Urbach, N., Ahlemann, F.: Transformable IT Landscapes: IT Architectures Are Standardized, Modular, Flexible, Ubiquitous, Elastic, Cost-Effective, and Secure. In: Urbach, N., Ahlemann, F. (eds.) IT management in the digital age. vol. 6. Management for Professionals, pp. 93–99. Springer, Cham, 2019.

[Sc16] Scholz, G.: IT systems in public transport: Information technology for transport operators and authorities. Heidelberg: dpunkt.Verlag, 2016.

[HK94] Heinze, G.W. und Kill, H. H.: Zukunftsfähige Strategien für den ÖPNV in Berlin-Brandenburg, ISBN 0722-8287, 1994.

[BB99] Beutler F. und Brackmann, J.: Neue Mobilitätskonzepte in Deutschland: Ökologische, soziale und wirtschaftliche Perspektiven. WZB Berlin Social Science Center, 1999.

[Ka16] Kamargianni, M., Li, W., Matyas M. und Schäfer, A.: A critical review of new mobility services for urban transport. Transportation Research Procedia, Jg. 14, S. 3294–3303, 2016.

[Hi15] Hildebrandt, B., Hanelt, A., Piccinini, E. und Kolbe, L.: The Value of IS in Business Model Innovation for Sustainable Mobility Services-The Case of Carsharing, 2015.

[Al17] Aletà, N. B., Alonso C. M. und Ruiz, R. M. A.: Smart mobility and smart environment in the Spanish cities. Transportation Research Procedia, Jg. 24, S. 163–170, 2017.

[Ba13] Barone, R. E., Giuffrè, T., Siniscalchi, S. M.. Morgano M. A. und Tesoriere, G.: Architecture for parking management in smart cities. IET Intelligent Transport Systems, Jg. 8, Nr. 5, S. 445–452, 2013.

[Ma15] Marchetta, P., Natale, E., Pescapé, A., Salvi, A. und Santini, S.: A map-based platform for smart mobility services. IEEE Symposium on Computers and Communication (ISCC), 2015, S. 19–24, doi: 10.1109/ISCC.2015.7405448.

[LDI17] Levina, A. I., Dubgorn, A. S. und Iliashenko, O. Y.: Internet of Things within the Service Architecture of Intelligent Transport Systems. European Conference on Electrical Engineering and Computer Science (EECS), S. 351–355, 2017.

[Me20] Menge, F.-W.: Digitale Transformation in einem ÖPNV-Unternehmen (BVG). Berliner Verkehrsbetriebe (BVG) AÖR, D-13355 Berlin-Wedding. Zugriff am: 6. Februar 2020.

[WS20] Würtz, M.-O. und Sandkuhl, K.: Impact of New Mobility Services on Enterprise Architectures: Case Study and Problem Definition. 11th BITA Workshop. BIS Workshop Proceedings, Springer LNBIP, 2020.

[HL18] Hacks, S., Lichter, H.: Towards an Enterprise Architecture Model Evolution. INFORMATIK 2018 - Workshops, Workshop on Enterprise Architecture in Research and Practice. LNI P-285, GI e.V., 2018.

[Zi19] Ziemann, J.: Architectural Content Management in Agile Times. INFORMATIK 2019, Workshop on Enterprise Architecture in Research and Practice. LNI P-295, GI e.V., 2019.

Software Architecture Best Practices for Enterprise Artificial Intelligence

Yannick Martel[1], Arne Roßmann[2], Eldar Sultanow[3], Oliver Weiß[4], Matthias Wissel[5], Frank Pelzel[6], Matthias Seßler[7]

Abstract: AI systems are increasingly evolving from laboratory experiments in data analysis to increments of productive software products. A professional AI platform must therefore not only function as a laboratory environment but must be designed and procured as a workbench for the development, productive implementation, operation and maintenance of ML models. Subsequently, it needs to integrate within a global software engineering approach. This way, Enterprise Architecture Management (EAM) must implement efficient governance of the development cycle, to enable organization-wide collaboration, to accelerate the go-live and to standardize operations. In this paper we highlight obstacles and show best practices on how architects can integrate data science and AI in their environment. Additionally, we suggest an integrated approach adapting the best practices from both the data science and DevOps.

Keywords: Software Architecture; Enterprise Architecture; Machine Learning; Artificial Intelligence; MLOps

1 Introduction

Despite gaining more and more popularity (and hype), AI technology is still in an infancy stage and may be intimidating to companies with no AI know-how and experienced personnel. AI use cases involve many challenges and often fail to scale and translate into productive solutions [Wh19]. In addition, specifically data science projects are very deep and challenging in terms of content. Their level of industrialization is not comparable to what we have reached for other business applications. While being at the core, the machine learning part used in the enterprise environment represents only a fraction of the overall system, but machine learning still put some strong constraints on the overall architecture. This can

[1] Capgemini, 147 Quai du Président Roosevelt, 92130 Issy les Moulineaux, France, yannick.martel@capgemini.com
[2] Capgemini, Bahnhofstraße 30, 90402 Nuremberg, Germany, arne.rossmann@capgemini.com
[3] Capgemini, Bahnhofstraße 30, 90402 Nuremberg, Germany, arne.rossmann@capgemini.com
[4] Capgemini, Mainzer Landstraße 180, 60327 Frankfurt, Germany, oliver.weiss@capgemini.com
[5] Capgemini, Olof-Palme-Straße 14, 81829 Munich, Germany, matthias.wissel@capgemini.com
[6] IT Dept. of Germany's Federal Employment Agency, Südwestpark 26, 90449 Nuremberg, Germany, frank.pelzel@arbeitsagentur.de
[7] IT Dept. of Germany's Federal Employment Agency, Tafelhofstraße 4, 90443 Nuremberg, Germany, matthias.sessler@arbeitsagentur.de

doi:10.18420/inf2020_16

result in problems such as technical debts that prevent further development, modification or even operations of the system [Sc15].

Project participants and customers therefore face difficulties that they are not familiar with, for example, from "traditional" enterprise information system projects. These new challenges include:

(1) Model rules and logic exist outside the well-known code base. This makes it difficult to encapsulate and control the behavior of the whole system. In addition to code dependencies, dependency on data is also introduced when using Machine Learning Models.

(2) The model changes its behavior over time. This is referred to as model drift and can lead to a decrease of quality for predictions by a model that indeed is successfully deployed in production but is decaying in its abilities.

(3) A mix of frameworks, different programming concepts and languages is involved. The detection of semantic errors is much more difficult in such an environment, where each component has been selected as optimal for one function but might not fit well in the overall architecture.

Integrating AI into existing systems is a process that is complicated not only due to technical intricacies but also due to the organizational aspects. The data science team primarily aims at accelerating and simplifying experimentation and innovation based on machine learning and data analytics. They commonly work in a fast-paced prototyping and rapidly iterate, test new approaches and explore new areas. The focus of enterprise software development team, on the other hand lies on stable and reproducible software releases. Both disciplines follow different workflows and incorporate different tools.

It seems difficult to incorporate the dynamic nature of ML model development and training into an end-to-end software development cycle in a reproducible fashion. We commonly see a lack of expertise in software engineering in data science teams and experience for corresponding software architectures. On the other hand, IT departments struggle to understand the specificities of machine learning based software systems, and to integrate the exploratory and empirical approaches favored in data science. Some approach is needed to incorporate the two corpuses of knowledge and practices and to apply it to our new AI-based systems.

In our paper we first summarize related work examining the integration of ML components into large enterprise software systems in section 2. Section 3 then presents the basic methodology of comparing case studies which we will use to sharpen the problem domain, followed by a detailed analysis of several case studies in section 4. In section 5 we show a best practices approach for enterprise software development incorporating ML parts by discussing suitable processes, team lineup and eventually presenting our Machine Learning Reference Architecture. Finally, section 6 presents conclusions, limitations and future

research by comparing the discussed approach with other solutions such as commercial platforms.

2 Related Work

The problem of moving machine learning into production has been mostly faced by the large web players of the AI generation, whose business model is built on large-scale automation and massive use of machine learning. Typically, Google, Amazon, LinkedIn, Facebook, Lyft and their colleagues have developed and approached software to face the challenges they encountered. Some have published their learnings, like Google in [Sc15], Facebook in [Ha18] or Uber in [HD17]. Most of the work is on large to very large-scale infrastructure, adapted to the volume of data and size of teams of these players.

MLOps is now one of the *5 Hands-on Skills Every Data Scientist Needs in 2020*[8], and is part of ML modernization movement. Many vendors such as DataRobot, Dataiku, SAS, SAP, ParalellM, and of course the cloud platform vendors such as AWS, Google Cloud Platform and Microsoft Azure have viewpoints, recommendations and reference architectures.

3 Methodology

We chose an exploratory approach and used a case study analysis to identify and categorize architectural shortcomings and problems which occurred in real practice of AI projects. These practical cases include:

- Federal Employment Agency (FEA): An operational analytics tool visualizes and predicts cpu and memory utilizations including peak and idle times [Ch19].
- BSH Hausgeräte GmbH: A novel approach for predicting and visualizing distributed product information queues at BSH Hausgeräte GmbH [Ch20].
- Federal Employment Agency (FEA): The use of Natural Language Processing (NLP) to classify documents. In the case, student certificates are evaluated by our machine learning models and the results are made available to the processor.
- Premium automotive OEM: Anomaly detection in a car fleet.
- Factoring branch of a French bank: Integration of scoring algorithms into productive environment
- A fraud detection startup in need of running machine learning models in a real time streaming environment.

[8] https://www.kdnuggets.com/2020/01/odsc-5-skills-every-data-scientist-needs.html

- Aircraft maintenance and material support analytics: Building a robust and stable data pipeline for many different data sources

To resolve this, we interviewed experts involved in these projects and based on their feedback identified and worked out the problem core in order to cluster problem domains which can be subsequently addressed through our solution approach.

4 Case Study Analysis

4.1 Federal Employment Agency (FEA): Operational Analytics

This case develops a novel tool for data center management that incorporates data visualization and machine learning capabilities for a large government agency in Germany, which hosts three highly available data centers containing more than 10,000 servers [Ch19]. The solution comprises a web-based 3D prototype running on Node.js. The tool provides a significantly better option and enabled visualization of historical data for all server instances at the same time, as well as real-time charts. It also uses the full potential of machine learning for time series forecasting.

There were quite a few challenges faced during implementing the ML-based tool. Setting up the Python environment including all needed dependencies was time intensive. Development of new functions which required new dependencies (which in worst case conflict with existing ones) would block other developers to restart the system after an update. Anaconda was not required at the beginning but later became essentially the only solution to get the system running. At the end containerization simplified a lot, as it allowed easy encapsulation of the dependencies and smooth installation.

4.2 BSH Hausgeräte GmbH: Predicting and visualizing distributed product information queues

The objective of the project was to present processing state and queues in a more understandable and predictable way and to visualize them in order to enable improved decision-making.

When implementing the solution, many challenges came up: there existed many versions of Python scripts, notebooks etc. Defining interfaces was not that easy as we know this in the JEE world. The production environment consisted of a complex infrastructure and many pieces had to be put together before having a running system. Sometimes a small change led to issues where other developers were unable to run the system (as it ran one day before) as they needed to adjust parameters and paths manually etc. The overall build process was not easy and transparent, for details see [Ch20].

4.3 Federal Employment Agency (FEA): Natural Language Processing (NLP) for Confirmation of Studies

There are vast amounts of text documents in public administration, which opens great potentials for the use of ML-algorithms. In our case, certificates, which are uploaded to claim child benefits, are now evaluated by machine learning models. Especially classification and automatic information retrieval foster a simpler, faster and more beneficial way of processing these data.

As this had been the first use case at FEA which had been deployed to production, there were some initial efforts needed. E.g. once we had to recode from Python to Java to hand over the model. Today, we develop and deploy in Python. Other efforts are organizing and implementing a monitoring and feedback loop for the first time, especially if this causes some modifications to operating applications. A further, at first glance trivial aspect, is the organisation of the collaboration of the numerous stakeholders for the first time - i.e. to enable IT-Experts, Business Analysts, Product Owners, Data Scientists etc. to work together with different access, views etc. and within the data science team with script and model management, versioning etc. Due to performance and scalability issues in other use cases, the need for upgrading the laboratory to a ML-factory by an on-premise stack become inevitably. The requirements to take a platform decision are listed in chapter 5.5

4.4 Premium Automotive OEM

Anomaly detection system: All vehicles during the ride are regularly sending data to the endpoint in the cloud. However, often due to technical problems, some data is lost or delayed. The customer wants to be informed as soon as possible about problems with collecting the data.

The solution consists of three models underneath, each one specialized in detecting different kind of anomalies: Linear Learner is an AWS SageMaker built-in model which was used for making prediction for next 24 hours, and each hour a check whether incoming data is within confidence interval of prediction was made. Autoencoder is a neural network, which finds the function would follow the general pattern and would be insensitive to sudden deviations. Then the difference between incoming data and denoised function is assessed using Local Outlier Factor to decide whether observation is anomaly. For retraining purposes, scripts for data preprocessing and training, which must be run by user, were prepared. In the future, there will be a module for automatic retraining.

4.5 Factoring branch of a French Bank

In this case the factoring branch of a French bank required a quick integration of two ML algorithms giving scorings used by a digital platform into production: One for fraud

probability (using graph analysis; for instance if the bank client is linked to many known fraudsters, it may be a fraudster) and another about default of payment probability (if the amount of transaction of the client is very different of the usual amounts seen for this client, the score is higher).

Using a DevOps pipeline, each algorithm has been delivered on production four months after the delivery by the data scientist. The software engineering team decided not to reimplement all the preparation procedure but to reuse the SQL scripts already done and to optimize only the scripts that needed to. This way the work was easier because data scientists and software engineers used the same language. The result of the preparation process was exposed in a NoSQL Database to be used by a microservice with low latency. The prediction was done by a Java microservice merging the data from the NoSQL database with the arguments sent in a SOAP query to call a Python script giving the score as result. This mixed architecture introduced many difficulties due to the closely coupled Python-integration with Java. As key learning we would recommend exposing Python scorings using Python (micro-) services with frameworks like Django[9] for example. We also had difficulties to manage Python packages environments and we started to implement a dependency management automation process to avoid this issue on a follow-up project.

All logs (during the preparation process or during the microservice call) were sent to an ELK cluster and visualized via technical and business dashboards to follow the algorithm usage. A difficulty concerning this kind of monitoring was to train the hosting team to manage all the parts of the project. The team's background was mainly the management of batch loadings for BI and so we chose to hide the technical complexity of Big Data technologies using a framework. In addition, we created trainings for the ops team with reference to BI to give them some examples they already knew.

4.6 Fraud detection at Fintech startup

This case is an example of moving machine learning from the lab to a production environment. This young fintech has developed a machine learning model for detecting payment fraud in real time, based on behavior fingerprints of clients, accounts, banks and payments. This model has been proven in the lab on historical data, with big data tools and a mix of Python and Scala scripts.

Then the model had to be introduced in a real-time process, together with data preparation and feature engineering. Therefore, the fintech has developed a real-time streaming engine supported by a Kafka unified log, with the following steps:

1.) Collection and conversion of transactions and auxiliary data onto a common data model; 2.) Enrichment of transactions with auxiliary data – for instance addition of client profile and bank profile to the transaction; 3.) Calculation of features for the machine learning

[9] https://www.djangoproject.com

model – including short-term (1 minute, 1 hour, 1 day) and longer term (1 year, 1 month...) aggregates based on different keys (account, client, country...) and selection criteria (failed transactions, instant payment...). These behavioral features require the maintenance of pre-features in an in-memory database. This move from a batch lab environment to real-time has required redesign and redevelopment of all the algorithms; 4.) Application of machine learning models, using the same library as during the lab, but with its Scala bindings; 5.) Application of configurable business rules, to assess the risk and make a final decision.

This effort has been done in a separate environment, with independent technologies, and with a pure DevOps / Agile development methodology and tools. The data science team can now develop new models and integrate them in the developed framework.

4.7 Aircraft maintenance and material support analytics

This case is an example for reproducible analytics insights and constantly growing data pipelines to enable end users making data driven decisions in different areas surrounding the operation of military aircrafts. The focus is the aircraft maintenance, where the provided dashboard help support engineers perform root cause analysis of errors and part failures. Another area is material services where the tracking of parts within the repair or maintenance cycle is enabled by dashboards.

Starting with the many source systems on end customer side as well as in house, the data needs to be ingested and further processed on a single platform to leverage the full potential of the collected information. Due to the iterative process and increasing number of included source systems, the challenge of delivering reproducible and consistent results in the required dashboards for the different user groups is a constant challenge. Particularly due to the use of the Hadoop stack, which comes with great processing frameworks like Apache Spark but also the Hadoop Distributed File System (HDFS) which prevents the use of common already established data versioning tools. Other challenges to build a robust and stable data pipeline stem from the different frequencies and volatile quality of the data deliveries from the source systems.

To overcome these issues, a data versioning system was established that connects the version of the code for the processing logic with the used data delivery. Also, a monitoring system for the data quality and consistency of the derived insights was established to be able to spot data issues as soon as possible and provide initial clues for further investigation.

5 Developing Best Practices for a Data Science Architecture

In the following, we distil the best practices from the case studies. This extraction of best practices forms the actual outcome of the performed case study analysis. It is precisely a responsibility of Enterprise Architecture Management to detect these pitfalls and recipes

for success and to anchor them in the organizational memory. We map above mentioned pitfalls into following categories:

- *Organizational*: New professions from the field of big data seem to be taking a firm place in organizations, but their tasks in projects or enterprise programs, and at the intersections to other departments, have yet to be sharpened.
- *Procedural*: New tools and ways to create solutions in the data science environment also bring new problems that hinder agility, high level of data quality and integration into the solution building processes at enterprise level.
- *Architectural*: Well working solutions in lab environments often fail to scale and deliver sufficient performance in real world scenarios or an explosion of technical complexity prevents an integration into manageable products.

5.1 Requirements

First and foremost, governance and leadership are needed which will be considered as critical to move machine learning to production and the introduction of data science into operational processes. This innovation requires the transformation and adaptation of software development processes and thus the leadership and vision to support that. Subsequently, we must integrate data science experiment-oriented activity within the software engineering activities, oriented towards developing industrial-grade software systems. This is a clash of cultures, as well as skills and practices. Finally, we need architectures, technical tools and environments to support an end to end processes, encompassing exploration, prototyping, model development and training, then allowing easy transition into industrial software.

The tooling must allow for automation and repeatability. We also need to support serving and monitoring of machine learning models, both from a technical and business standpoint. The architecture patterns must cover different types of machine learning models in different contexts, varying with the business, the deployment constraints, the temporality (real-time vs slow processing) and the volume of data to process.

In order to cover existing and future requirements of developments in the field of Advanced Analytics and Machine Learning, there is also a need to upgrade and expand the analytics platform. It requires an integrated system of coordinated hard- and software in the local data center. The hardware must be designed to meet the data science specific requirements This results in particular

- due to increasing amounts of data as well as significantly increased demand for computing capacity because of the growing complexity of the methods and algorithms used.

- from the increased demands in the development of analytical models with regard to data protection and security, data ethical principles and legal framework conditions (GDPR, EU Whitepaper on AI).
- from the constantly increasing amount of machine learning increments used in productive operation, requiring larger and more flexible functionalities in terms of data and workflow management, interfaces and monitoring of data and models.

Finally, machine learning requires a high degree of connectivity within the organisation. This places completely new demands, but also opportunities on IT security, see [Be19]. A framework that allows flexibility in tooling and methodology while maintaining the integrity of sensitive data within the organization.

5.2 Defining a MLOps process

In order to overcome the above-mentioned problems and satisfy the requirements, we propose viewing developing ML applications as a software engineering activity, thus in need of a software engineering process. We adapt best practices from the Agile as well as the DevOps movement. Main characteristics are:

- Release early, release often
- Short adaption cycles, small safe increments
- Strong version control
- Testing in every stage and automatically
- Automated integration and deployment (CI / CD)
- Reproducible processes and reliable software releases
- End to end responsibility of teams from Dev to Ops

Fig. 1 shows a typical DevOps process which is basis for continuous integration (CI) and continuous delivery (CD), see also [SWW19].

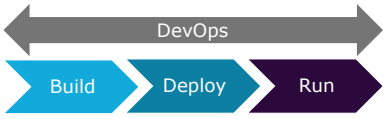


Fig. 1: Typical DevOps process

For ML applications the process for continuous development and deployment gets more complex, because of additional types of artifacts which are subject to change:

(1) Data: A model is chosen, trained and operated upon data, whereas changes in schemes or attribution can occur any time.

(2) Model: Algorithms, parameters and hyperparameters representing the model behavior as outcome from experiments and training over time

(3) Code: Primary artifact in classic software engineering including all kinds of source code, scripts, tests

The idea of Continuous Delivery must adopt these dimensions in order to create a repeatable and reliable software release process including Machine Learning artifacts. We refer to this extension of DevOps as MLOps.

Fig. 2 shows the proposed Software Engineering process which is extended by the steps "Define" and "Train" to deal with above mentioned additional dimensions of data and model. This way we add following feedback loops to the process in order to integrate central ML lifecycle steps:

- Define and build a suitable model and adapt based on demo feedback through experiments
- Include training the model and adapt based on observed model performance
- Retrain an operational model based on new real-life data and performance
- Adapt result of the whole process based on live feedback, preventing model drift

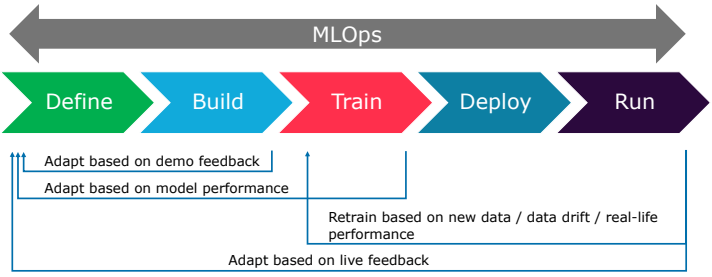


Fig. 2: Extension of the DevOps process to proposed MLOps

The resulting process allows reproducible small increments for reliably releasing machine learning applications in short adaption cycles. This enables managing the additional dependencies on changing model behavior and data (see problems mentioned in chapter 1) allowing continuous model training und monitoring model performance in production. In our opinion, this is crucial for preventing model drifts by iteratively checking for performance degradation with the ability to re-train the productive model or implement a new one based on new key assumptions.

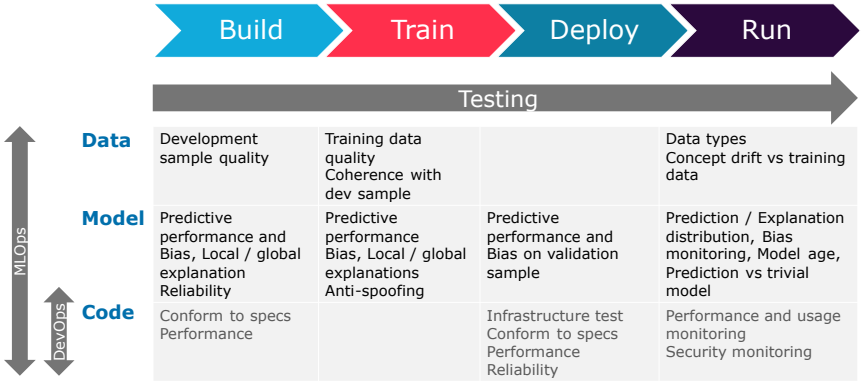


Fig. 3: Tests in every phase of proposed MLOps process

A specific mention concerns testing, as it is not a step somewhere in the process but is to be considered as integral part of the engineering procedure. As so, testing should be carried out in every step on every involved artifact, especially on the model and the data. Our proposed extension of testing in DevOps is shown in Fig. 3: In addition to tests on the code, there are test targets also for data and the model.

5.3 Cross functional teams and mix of competences

From the perspective of a classic ML lifecycle, the role setting of Business Analysts together with Data Scientists and Data Engineers is sufficient for conducting a working ML solution which proves to deliver all required benefits in a simple lab environment. For industrialization purposes in the context of complex enterprise products additional roles must be considered. Again, by applying concepts from the DevOps world these roles may include: Software Developers / ML Engineers, Architects, Project Managers, Designers and DevOps Engineers. These disciplines should form a cross-functional team and act together, else we observe the before mentioned problems of delays and friction, as is common in complex IT projects. This is due to the ownership of different phases of the process (see Fig. 2) to different roles where results are just handed over. We made good experience by creating cross functional teams that has experts from every discipline in order to cover the whole MLOps lifecycle and aim at a common end-result: an application live, serving users and Business.

5.4 Reference Architecture for MLOps

In this chapter we propose a data science reference architecture that works as a blueprint when creating ML solutions using the MLOps approach. This blueprint architecture is part of an overall Data Analytic architecture with three large blocks, see Fig. 4.

- The platform foundations form the common platform elements, providing infrastructure for the other functions; they address the raw storage, raw processing, software engineering DevOps tooling and automation, and security;
- The AI, Analytics & BI Foundations provide tools for performing the processing and analysis of data;

The AI, Analytics & BI Execution block host the custom-made as well as the off-the-shelf algorithms and applications. The AI, Analytics & BI Foundations contains the MLOps foundations, which incorporate common building blocks addressing specific dimensions that form a holistic ML project approach in enterprises. The many commercial or Open Source products and tools address one or more blocks. In this MLOps foundations, we identify the following functional areas:

- An AI Marketplace allows to download and incorporate reusable models and AI libraries, even reusable data sets.
- Analytics Monitoring to supervise models in development and production in order to prevent model drifts or install a trigger on when re-training is needed.
- Analytics Orchestration provides management of data pipelines as well as mechanisms for serving of models and testing them.
- The Data Science Workbench is the development environment for data scientists including management of experiments
- The Model Repository enables the management of the whole lifecycle of machine learning models and stores the various versions together with metadata
- The Data Modelling enables data preparation and features engineering

On the right side of Fig. 4, we located the AI, Analytics & BI Execution which describes platforms solutions that encapsulate and hide many or all the building blocks from the left side in order to deliver ready-to-use AI services for users who don't want to deal with connecting single blocks to the solution themselves. Examples are AI Platform as a Service or API as a Service offerings by Cloud vendors. These products are often referred to as low-code platforms. Following functional areas include:

- Custom exploration providing fixed toolsets for preparation and gaining insights into several kinds of datasets.

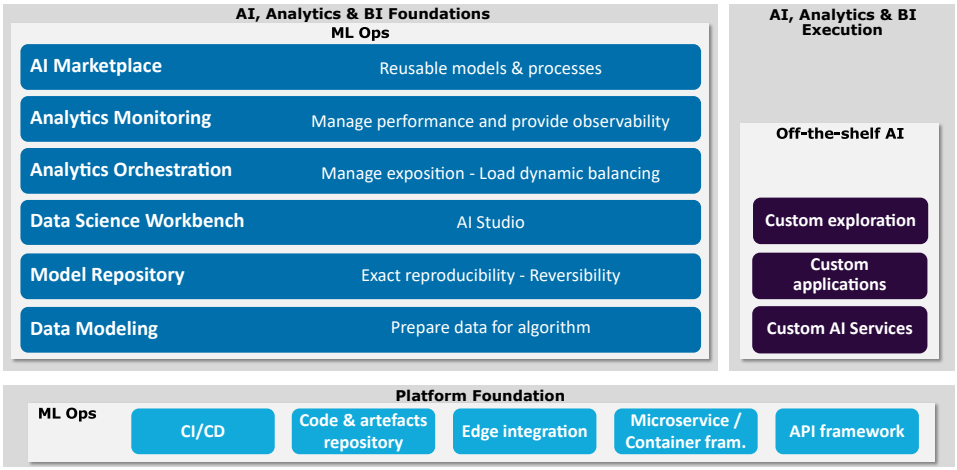


Fig. 4: Reference Architecture containing MLOps-related building blocks

- Custom applications that offering predefined AI solutions for a certain scope of business aspects
- Custom AI Services that can be called to process AI tasks as part of a broader business solution or product.

Platform Foundation forms the bottom part of the proposed architecture defining building blocks that relate to infrastructure and cross-functional aspects:

- CI/CD platform allowing to implement the proposed MLOps process chain in order to supply repeatable and reproducible results
- Code and artefacts repository facilitating version control of model data and code
- Edge integration enabling us to include local computing-, sensor- and other IoT-devices in our AI solution
- Microservice / container framework as a standardized execution environment
- API framework that supports the microservice approach in encapsulating and providing access to AI functionality

5.5 Choosing the right Architecture Approach

We discuss different scenarios on how the before mentioned Architecture can be used to choose an approach for creating AI solutions in a concrete environment, see Fig. 5. A

first question addresses the ability to create IT-based solutions and understand underlying mathematical and technical mechanisms. Organizations that have teams with such a background can pursue the Code-first approach and take over the responsibility to create and deliver suitable business benefits on their own. A Low-code approach on the other hand offers products and platforms suitable for non-developers such as business-minded citizen data scientists but come with higher vendor dependency and limited solution space.

The second dimension describes to what extent the used building blocks can be customized and their associated flexibility and adaptability. Platform as a Service (PaaS) solutions offer a great variety of ready-made solutions but are limited when they must be adapted to different usage scenarios. They also lock you into a proprietary solution with more or less proprietary interfaces. Going direct over an infrastructure, either on the cloud or on premise provides more flexibility and offer a greater accuracy to fit to the needed solution. Unfortunately, it comes with much higher efforts to form the appropriate product, and require stronger expertise in architecture, development, tuning and operations.

There is no "right choice" here and we hope the simple diagram can help make the first step for making a choice in each case – or choosing to combine approaches for an optimal combination.

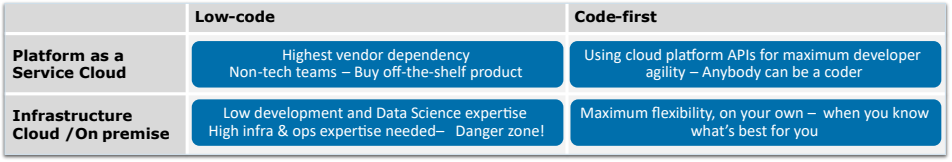


Fig. 5: Scenarios when adopting AI solutions

5.6 Summary of best practices

The discussed best practices from this chapter can be summarized by again using the before mentioned problem categories: 'Organizational', 'procedural' and 'architectural'. Organizational best practices start with governance and leadership: It is necessary to align a cross-functional project team and establish a common vision, but also to solve differences with respect to skills, practices, tools and environments that are used. In addition, data protection and security, ethical principles or legal factors such as GDPR play an important role that is often underestimated.

Second, as the process for continuous development and deployment is getting more complex for ML applications, we advise to combine elements from both, the Agile as well as the DevOps movement. Therefore, we proposed a process that allows reproducible small increments for reliably releasing ML applications in short adaption cycles.

Last, a sound reference architecture as well as architectural approach complete the picture. The building blocks for this consist of the "platform foundations" that provide the infras-

tructure for other functions, the "AI, Analytics & BI Foundations" that provide the tools for performing the processing and analysis of data, and the "AI, Analytics & BI Execution" block that hosts the custom-made as well as off-the-shelf algorithms and applications.

6 Conclusions and future research

The analysis of cases presented in section 4 and the extraction of best practices by chapter 5 provides a consolidated set of reusable templates for approaching architectural challenges in large AI projects as well as a general set of pitfalls that need to be considered during realizing such projects.

Moving machine learning to production is necessary to avoid the Death Valley of technologies which were the latest hype one day but failed to fulfill their promises. ML has high potential and can transform an organization into a leader in its field, as it has done with Google, Amazon, Netflix, but also Experian and Amex. These organizations did not become leaders because they were the first or because they lacked competition – they did because at the core, they developed the expertise of applying machine learning each to its field. As organizations are struggling to succeed on their markets and overcome competition, machine learning has the potential to improve their chances for success.

As we have shown through the examples, many organizations are experimenting and doing true innovation – that is doing new things and bringing them to serve – with machine learning in production. Tools and processes are not yet mature, but they are improving fast. Software engineering is providing solid foundations and making the Agile organization a reality. The case studies of this report and the experiences of the authors show that combining practice of software engineering and machine learning can realize enterprise grade solutions in a industrialized fashion like we are used to through DevOps.

Many of the above-mentioned case studies where in dedicated environments. You can go a long way with Open Source, but cloud platforms and commercial products are easier paths, especially if you are in a hurry and/or do not have the full technical competencies. Both ways have (dis-) advantages. However, the ability to integrate new and updated Open Source tools is a key requirement in view of the rapid development ongoing in this field. Cloud adoption will probably hugely facilitate the move of machine learning to production by providing easy access to supporting tools with low administration or installation effort. Even then, the dominant factor in the maturation of machine learning in production will be the confirmation and the democratization of best practices, dedicated tools and architecture patterns, based on the first experiences.

In future research, our model could be extended by considering commercial cloud vendors for AI or mixing with Open Source and other cloud vendor tools. First insights suggest that considering wide use of Open Source may increase the control of the overall architecture (and therefore solution fitting accuracy), while the collaborative support of large or distributed

teams might be easier on commercial platforms. Especially when it comes to platform foundations and AI execution, commercial platforms seem to add significant value for ML projects compared to a "traditional" Open Source setup, often seen in pure DevOps solutions. Experience has yet to be gained here and should be topic for further research activity.

References

[Be19] Ben Zid, I.; Parekh, M.; Waedt, K.; Lou, X.: The application of Articial Intelligence for Cyber Security in Industry 4.0. In (Draude, C.; Lange, M.; Sick, B., eds.): INFORMATIK 2019: 50 Jahre Gesellschaft für Informatik – Informatik für Gesellschaft (Workshop-Beiträge). Gesellschaft für Informatik e.V., Bonn, pp. 255–260, 2019.

[Ch19] Chircu, A.; Sultanow, E.; Baum, D.; Koch, C.; Seßler, M.: Visualization and Machine Learning for Data Center Management. In (Draude, C.; Lange, M.; Sick, B., eds.): INFORMATIK 2019: 50 Jahre Gesellschaft für Informatik – Informatik für Gesellschaft (Workshop-Beiträge). Gesellschaft für Informatik e.V., Bonn, pp. 23–35, 2019.

[Ch20] Chircu, A.; Sultanow, E.; Hain, T.; Merscheid, T.; Özcan, O.: Real-Time 3D Visualization of Queues with Embedded ML-Based Prediction of Item Processing for a Product Information Management System. In (Zimmermann, A.; Howlett, R. J.; Jain, L. C., eds.): Human Centred Intelligent Systems. Springer Singapore, Singapore, pp. 347–358, 2020.

[Ha18] Hazelwood, K.; Bird, S.; Brooks, D.; Chintala, S.; Diril, U.; Dzhulgakov, D.; Fawzy, M.; Jia, B.; Jia, Y.; Kalro, A.; Law, J.; Lee, K.; Lu, J.; Noordhuis, P.; Smelyanskiy, M.; Xiong, L.; Wang, X.: Applied Machine Learning at Facebook: A Datacenter Infrastructure Perspective. In: 2018 IEEE International Symposium on High Performance Computer Architecture (HPCA). Pp. 620–629, 2018.

[HD17] Hermann, J.; Del Balso, M.: Meet Michelangelo: Uber's Machine learning platform, Sept. 5, 2017, URL: https://eng.uber.com/michelangelo-machine-learning-platform/.

[Sc15] Sculley, D.; Holt, G.; Golovin, D.; Davydov, E.; Phillips, T.; Ebner, D.; Chaudhary, V.; Young, M.; Crespo, J.-F.; Dennison, D.: Hidden Technical Debt in Machine Learning Systems. In (Cortes, C.; Lawrence, N. D.; Lee, D. D.; Sugiyama, M.; Garnett, R., eds.): Advances in Neural Information Processing Systems 28. Curran Associates, Inc., pp. 2503–2511, 2015, URL: http://papers.nips.cc/paper/5656-hidden-technical-debt-in-machine-learning-systems.pdf.

[SWW19] Sato, D.; Wider, A.; Windheuser, C.: Continuous Delivery for Machine Learning, martinFowler.com, Sept. 19, 2019, URL: https://martinfowler.com/articles/cd4ml.html.

[Wh19] White, A.: Our Top Data and Analytics Predicts for 2019, Gartner Blog Network, Jan. 3, 2019, URL: https://blogs.gartner.com/andrew_white/2019/01/03/our-top-data-and-analytics-predicts-for-2019.

Positioning IT4IT in the face of classic Enterprise Architecture Frameworks

A critical review

Andreas Hartmann [1], Gunnar Auth [2]

Abstract: IT4IT was introduced by the industry consortium The Open Group (TOG) in 2015 as a new reference architecture for the business view of IT management. Since TOG declared IT4IT to be a new standard and it apparently has an architecture focus, its potential use in enterprise architecture management has become a topic under discussion. In this study IT4IT is reviewed and compared with the classic enterprise architecture frameworks TOGAF and ARIS using evaluation criteria collected from literature. The results show that, although IT4IT has structural and topical similarities to classic EAFs, in particular the architecture focus and different views, its purpose and context of use is clearly different. Conforming to the concept of a service-oriented architecture (SOA), IT4IT is value oriented, service-centric, data driven, and automation focused; its position can be described best as part of an IT-related extension of a comprehensive EAF. Thus, IT4IT calls for integration with a classic EAF – without erroneously replacing or overwriting existing standards.

Keywords: IT4IT; TOGAF; ARIS; enterprise architecture; enterprise architecture framework; management of technology framework

1 Introduction

Firstly introduced by the international consortium The Open Group (TOG) in 2015 [TF19], IT4IT is a recent reference model for analyzing, designing, and improving the value creation of an IT organization aiming at efficiency and agility. With its latest version 2.1 dating from 2017 [Th17b], it addresses both corporate IT departments providing IT products/services to internal customers and IT vendors selling to a buyer market. While IT4IT is based on the process-oriented value chain model by Michael Porter, it describes and explains the "business of IT" [Th17b] from an architectural point of view, covering the dimensions information (data), function, integration and IT service. Consequently, TOG refers to the IT4IT model as a "reference architecture".

According to Winter and Fischer [WF06], Enterprise Architecture (EA) is understood as "the fundamental organization of a government agency or a corporation, either as a whole,

[1] HfT Leipzig, Gustav-Freytag-Str. 43, 04277 Leipzig, hartmann@hft-leipzig.de, https://orcid.org/0000-0003-1340-5325
[2] Meissen University of Applied Sciences, Herbert-Böhme-Straße 11, 01662 Meißen, Gunnar.Auth@hsf.sachsen.de, https://orcid.org/0000-0002-3013-2739

doi:10.18420/inf2020_17

or together with partners, suppliers and / or customers (“extended enterprise”), or in part (e.g. a division, a department, etc.) [. . .]” and should include both IT and business related artefacts. Following this definition, and in accordance with [Wa16], [Pr16] et al., we review IT4IT as an EAF. In addition to IT departments and IT vendors, IT4IT is also relevant for enterprises with their value chain depending on IT, especially software. Accordingly, TOG promotes IT4IT as “a modern framework for managing a digital enterprise” [Th20b].

On the other hand, IT4IT may not be yet another EAF. According to TOG’s strong value proposition, the model is designed to become “a game-changing foundation for IT” [Th20a]. The core concept of this new foundation is an IT value chain formed by end-to-end value streams and supported by a single integrated system of record containing all relevant IT information. This way, IT4IT aspires to close a “growing capability gap” [Ak16] in today’s IT organizations which results from a combination of rising demand and increasing technology trends. Typical problems IT4IT wants to overcome include “fragmented teams, processes, and tools”, leading to “disorganized handovers from business to the IT function and from development to operations” [Ak16].

As a new EAF, IT4IT is competing with a rich choice of existing EAFs. Sultanow et al. [Su16], for instance, identify and evaluate 55 EAFs. The problem of evaluating and selecting the best fitting EAF for an individual enterprise has been object to research in EAM for over two decades [AKL99]. Over this time, several methods and approaches for evaluating EAFs have been proposed (e. g., [AG06], [BNS03], [Fr09], [Su16], [UM06]). Against this background, we reviewed the new IT4IT framework to understand its characteristics in relation to existing EAFs as well as the appropriateness of common criteria for evaluation. As underlying research questions, we formulated:

1. How can IT4IT be characterized, regarding commonalities/differences and advantages/disadvantages, if comparing it to classic EAFs?
2. Can or shall it be applied alongside, without or even replacing a classic EAF?

For conducting the review, a structured process was applied: first, we reviewed the (sparse) existing literature on IT4IT, including both scientific and professional sources. Then, we collected and analyzed existing evaluation criteria for EAFs from literature. From the result, we created a consolidated set of evaluation criteria, suiting our evaluation objectives. Then the actual review was performed. We decided to include the two classic EAFs TOGAF and ARIS in the review and compared them to IT4IT. TOGAF was selected because of its high profile and because it is also developed by TOG. ARIS, on the other hand, was selected because of similarities to IT4IT in its focus on the IT domain and its basic dimensions (views).

2 EAF Evaluation Criteria

The evaluation of different EAFs is usually targeted at selecting the best-fit EAF for a given purpose. Our slightly different evaluation objective is to characterize the new IT4IT model through comparing it to existing EAFs. Since in both cases the object under evaluation is the same (EAF), we expect evaluation criteria developed for EAF selection to be appropriate for our evaluation objective too. Based on this assumption, we conducted a structured literature review to collect existing criteria for evaluating EAFs, following the recommendations of [vo15]. Starting with the search words 'enterprise architecture framework' + {evaluation | selection | comparison}, we used Google Scholar, Springer Link, IEEE Xplorer, and the ACM Digital Library. From the results, relevant articles were selected via title/abstract and further analysed, including forward and backward search. As expected, we found several criteria which are suitable for our evaluation objective and are supported by multiple authors. Tab. 1 presents the consolidated criteria along with a short description and the corresponding references.

Tab. 1: Evaluation criteria for EAFs

Criterion	Description	Reference(s)
Scope	Area of application as stated by the EAF	[Sc06], [Su16], [Wi10]
Goals	Future state of enterprise achieved through applying the EAF as intended	[AG06], [Sc06], [Wi10]
Design principles	Basic normative statements on how to achieve a to-be EA	[AFW11], [Bu17], [Fr09], [Sc06], [Wi10]
Views	Representations of certain aspects of EA from a specific point of view, often in the form of graphical models, e. g., business, process, infrastructure	[AFW11], [Bu17], [Fr09], [UM06], [WF06], [Wi10]
Metamodel	A model that defines the building blocks, their relationships, and notation for modeling the EA views	[AFW11], [Fr09], [LZ06], [WF06]
Method	Method(s) for EA design and evolution	[Fr09], [Bu17], [LZ06], [WF06]
Terminology	Terms and definitions	[Fr09], [WF06]
Reference model	Normative model(s) for re-use as template(s)	[Fr09], [WF06]
Adaptivity	Explicit support of adapting the EAF to technological and economical change, in particular digital transformation	[BY15], [Ga18], [HMS14], [KH17], [Ma17], [MV19], [Zi16]
Tool support	Availability of dedicated software tools	[AG06], [Wi10]

3 Introducing the Frameworks

In the following section we give a brief introduction to the IT4IT model as well as to TOGAF [Th18] and ARIS [Sc92]. The latter have been extensively discussed in the literature and should be familiar to many readers (e.g., [BS12], [DH11], [LZ06]).

3.1 TOGAF

The Open Group Architecture Framework (TOGAF) has been developed on the base of the US Department of Defense Technical Architecture Framework (TAFIM) [De96] as a tool for "assisting in the acceptance, production, use, and maintenance of enterprise architectures" [Th18]. Key element is the TOGAF Architecture Development Method (ADM). The ADM forms a requirement-centric, nine phase cycle, starting from architecture vision to architecture change management, and claims to be framework agnostic [Th18]. However, like in other EAF the definition of different views – to reduce complexity – is included in TOGAF as well. The so-called domains are business, data, application, and technology. With respect to reference architectures, TOGAF includes the Technical Reference Model (TRM) and the Integrated Information Infrastructure Reference Model (III-RM), both being capable to model information systems at a conceptual level as well as systems architectures. TOGAF does not include any own modelling notation, but supports UML, BPMN and TOG's Archimate [TO20].

3.2 ARIS

The ARchitecture of integrated Information Systems (ARIS) can be considered as a process-driven EAF. Starting from a holistic view of business processes ARIS is based on an integration concept and has the capability to create highly complex models. However, this complexity can be broken down using different views and abstraction levels. The views are organization, data, process, function, and output [Sc99]. Relationships between views and levels are essential and thus, a proper EAM tool is necessary to handle complex architecture models. Complexity reduction is further achieved through a lifecycle concept and the various description methods for information systems that are classified based on their proximity to IT. Thus, ARIS calls itself a framework for developing and optimizing integrated information systems, with an emphasis on the business related description level ([Sc92], [Sc99]).

Further, ARIS supports several notation/modelling standards, e.g. Unified Modelling Language (UML) [Ob20], Object Modelling Technique (OMT) [Ru91], and Archimate [Th17a].

3.3 IT4IT

With IT4IT the Open Group claims to provide a vendor neutral, technology agnostic and industry agnostic reference architecture for the business view of IT management. There are several reasons being named, e.g. that other common standards are proprietary and slow to adapt. It is capable to evolve and to enable continuous improvement. IT4IT uses a layered architecture, a meta model and the Archimate notation language version 3.0 [Th17a] as well as UML [Ob20]. Compared to e.g. ITIL the IT4IT reference architecture is

almost process agnostic and focusses on capabilities to run the IT operation model instead. Thus, the standard describes what needs to be done to enable an IT value chain – not how. However, one single process view is given: the IT value chain itself. The simple process model consists of four connected value streams forming a continuous value chain. Figure 1 depicts the four value streams "strategy to portfolio", "requirement to deliver", "request to fulfil" and "detect to correct". These value streams have been extended from the high-level "plan, build, deliver, run" process model and supplemented with supporting activities, e.g. finance, governance, and risk management [Th17b].

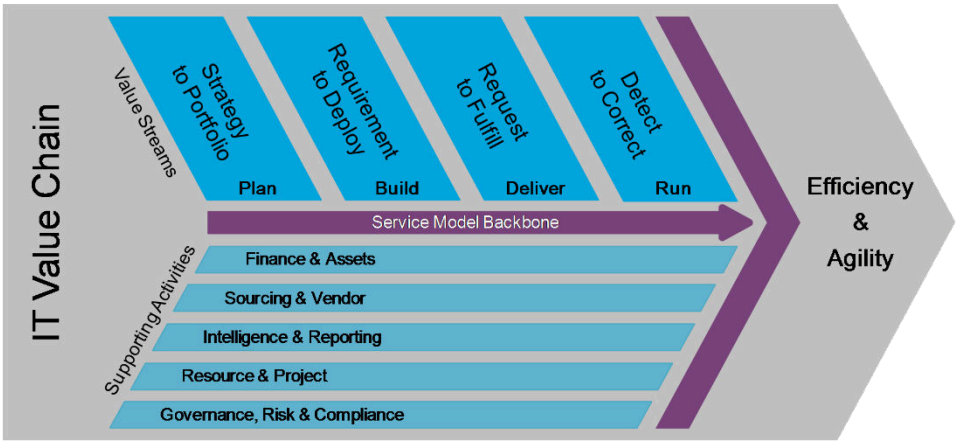


Fig. 1: IT Value chain and Service model ("backbone") of the IT4IT Reference Architecture [Th17b]

Starting from this high-level process representation IT4IT rapidly focusses on a combined functional, informational, service- and capability-based view. Key aspects of the value streams are described as follows: the service model, the essential data objects (information model), goals and capabilities, and functional components (functional model) that support the value stream. Functional components are integrated via well-defined interfaces (integration model) and thus, tie together the whole value chain.

The IT4IT reference architecture defines several levels of abstraction. This approach is similar to those used by eTOM [TM17] or ARIS [ID10]. Level 1 shows a high-level holistic view – the end-to-end overview of the IT value chain containing the four *plan-build-deliver-run* value streams (process model). The information, service and functional model have been defined at high-level as well, including service objects, data objects, functional components, and high-level relationships. Level 2 provides value stream documentation, including objectives, KPIs and capabilities. The focus shifts from high-level towards a more detailed integration model, e.g. relationships are updated with cardinality attributes. The concept of data flow between functional components is introduced at this level, too. Still vendor-independent, level 3 shifts further from general informational reference architecture to a more solution-based architecture. Data object definitions are being updated with more details and essential attributes. The user (solution architect) will be provided with the

capability of scenarios and essential services, the scenario being a "narrative that describes foreseeable interactions of user roles (or 'actors') and a system (or functional component)" [Th17b]. Example given; the architect may describe the implementation of system of records integrations using essential services to maintain the relationship between data objects [Th17b]. Levels 4 and 5 are vendor-specific and thus, controlled by suppliers of IT management products and services (e. g., Atlassian, ServiceNow, HP). The IT4IT model limits itself to some recommendations and example solution architectures.

4 IT4IT compared to classic EAFs

In this section, the results of applying the criteria shown in Tab. 1 to all three frameworks/standards are summarized. The valuation is based on the primary literature on the frameworks. Where interpretations were necessary, they were made based on the experiences of the authors. Comparing the different valuation allows for concluding on the relative position of IT4IT to the two classic EAF at the end of this article.

4.1 Valuation

Tab. 2: Valuation of the three standards

Criterion	**TOGAF**	**ARIS**	**IT4IT**
Scope	TOGAF is targeted at the "*'enterprise'* to be any collection of organizations that have common goals [. . .] encompassing all of its business activities and capabilities, information, and technology" [Th18]	Integrated information system of an enterprise from a business process perspective comprising function, data, organization and process view. The business related description level has priority over implementation issues [Sc99].	Addresses the IT operation model, thus its scope is limited to the IT domain of an enterprise, including any aspect of the business of IT. Since IT is essential to most business domains, IT4IT can be applied in those enterprises – no matter what business process is supported with IT. Enterprises with IT as core business may use IT4IT as basis of an (to be completed) EAF. IT4IT names but does not define enabling capabilities (e.g., financial and HR mgmt.).

Tab. 2: Valuation of the three standards

Criterion	**TOGAF**	**ARIS**	**IT4IT**
Goals	Optimizing business processes towards an integrated organization that is responsive to change and supportive of the business strategy. "Providing a strategic context for the evolution and reach of digital capability in response to the constantly changing needs of the business environment." [Th18]	Designing, analyzing, and optimizing business processes with an enterprise scope. Designing and implementing integrated enterprise systems to automate business processes.	IT organizations shall be enabled to "identify the activities that contribute to business competitiveness" [Th17b]. Introducing the IT value chain, IT4IT aims to align the IT value streams towards "efficiency and agility".
Design principles	Enterprise and architecture principles are defined in TOGAF standard part IV, concept of the enterprise continuum is defined in part V.	No explicit design principles are formulated in ARIS but can be reconstructed from the underlying business process theory: 1) process orientation with its focus on activities for creating customer value and 2) integration as the guiding principle for the design of process oriented enterprise systems. ARIS also incorporates explicit "generally accepted modeling principles" which aim to ensure the quality of individual models [Sc99].	Value-oriented: IT activities are aligned to value streams. Service-centric: IT value is delivered and managed as service/product following the service paradigm. Data driven: data entities are central elements in the architecture. Automation-focused: determination, automation, and integration of the IT toolchain for delivering IT services.
Views	Four *architecture domains*: - Business architecture - Data architecture - Application architecture - Technology architecture	ARIS views: - Function view - Organization view - Data view - Output view - Control / process view	IT4IT views: - Service model - Information model - Functional model - Integration model
Metamodel	Architecture Content Framework (content metamodel), defining e.g. entities and relationships.	Comprehensive metamodel in UML [Sc99].	Metamodel in Archimate & UML [Th17b]

Tab. 2: Valuation of the three standards

Criterion	**TOGAF**	**ARIS**	**IT4IT**
Method	ADM iterative process, starting with preliminary phase, and eight phases (A-H) surrounding the centric requirements management. Top-down evolution from generic architectures to organization specific solutions.	General procedural model for modeling business process oriented information systems as well as specific procedural models for implementing standard software / workflow systems and developing (object oriented) systems [Sc99].	No explicit method defined in the IT4IT model, may be used within the TOGAF ADM [Es18]
Terminology	TOGAF standard part I includes a full list of terms and definitions.	Related terms are defined in the publications by Scheer.	A full glossary with terms and definitions is provided. Some terms are explained in additional white papers, e.g. [BJ16].
Reference model	TOGAF's highly generic ADM can be complemented with architecture patterns [Bu09]. However, reference architectures and patterns are not included.	A full set of reference models is available for industrial enterprises [Sc94]. Although the latest revision dates from 1994 (1997 for German ed.) it still can provide support for today's enterprises.	IT4IT itself is designed as reference model to manage the business of IT.
Adaptivity	Framework or parts of it can be applied to any enterprise and situation, including digital transformation. Full enterprise scope leads to extensive, interdependent model sets which are difficult to oversee and maintain.	Framework or parts of it can be applied to any enterprise and situation, including digital transformation. Full enterprise scope leads to extensive, interdependent model sets which are difficult to oversee and maintain.	Reference architecture is idealized [TF17] and bound to the context IT operation model, specialized for digital transformation. Limited scope and reduced detail level facilitate changes.
Tool support	Many modern EAM tools provide TOGAF support, including ARIS Platform, Abacus, and LeanIX.	Originally introduced as ARIS Toolset in 1993. Later renamed to ARIS Platform and further developed until today.	No explicit support but may be implemented with EAM tools if modelling reference architectures is supported (e.g., ARIS Platform).

4.2 Positioning IT4IT

Through comparing IT4IT with classic EAFs, we intend to extend the understanding of the nature and characteristics of the new approach and its relations to existing concepts. Although IT4IT has structural and topical similarities to the two classic EAFs TOGAF and ARIS, its purpose and context of use are different. Other than classic EAFs, IT4IT includes and makes heavily use of the service perspective. Typical EAFs come to modelling business processes at some point, followed by determining the process' supporting IT artifacts. In contrast, IT4IT defines service entities with associated data entities and groups functional components around them. The Service Model Backbone integrates the service entities and their relationships on a conceptual data level "to ensure end-to-end traceability of a service from concept to instantiation and consumption" [Th17b]. This approach conforms to the concept of a service-oriented architecture (SOA) and follows the principles of cohesion and coupling. In fact, IT4IT's functional components encompass one service or data entity, no more than two. This is best practise object-oriented design: high cohesion and loose coupling. The integration model at level 1 follows the service perspective as well, leading to a component-based service architecture.

This way, IT4IT supports the implementation of a micro services architecture and further paves the way to automating the service lifecycle, especially for software intensive environments. When service delivery heavily depends on software or is completely implemented through software, more and more data is produced. Utilizing this growing amount of heterogeneous and fragmented data for monitoring and automation requires a means for understanding what data exists and how it is related to service delivery. IT4IT addresses this demand with its integrated view on services, functional components, and data entities. Other than common EAFs it also includes KPIs for the evaluation of the architecture management.

In their comparison of EAFs, it was stated by Urbaczewski & Mrdalj that „most if not all frameworks were weak in addressing the maintenance of an information system" [UM06]. Regarding IT4IT, we can assert the opposite: maintenance and operations are one of its strengths. However, this strength may also be considered a risk. The roots of IT4IT are in the IT operation model. Well-orchestrated, agile, and scalable IT services are essential to modern enterprises, and IT4IT gives them a reliable architecture. But setting the focus on IT4IT only and disregarding its connections to classic EAFs, e.g. in the context of financial management, an organization may erroneously replace or even reinvent existing standards. The strength seems only a strength, when IT4IT is combined to a common (and maybe already existing) EAF – providing the reference architecture for the business of IT, while the common EAF delivers a method and the overall picture. It has been shown, that IT4IT and TOGAF can work together. But how about other EAFs?

5 Conclusion

The main characteristic IT4IT has in common with TOGAF, ARIS and also other classic EAFs is its focus on architecture. Furthermore, several structural elements of classic EAFs also do exist in IT4IT (e.g., a metamodel, different views and a well-defined terminology), leading to a structural compatibility.

The main differences are the scope and the intended character of a reference model. While the enterprise-wide scope is part of the nature of an EAF, IT4IT's scope is limited to the IT organization of an enterprise. Only, when the business model of a company is based on providing IT services or IT products to external customers, IT4IT could serve as an enterprise-wide framework. Nevertheless, for non-IT companies, because of the structural compatibility, IT4IT can be integrated with an EAF already in use to further align IT to business with a focus on efficient and agile service delivery. Because of their close relationship, an integration of IT4IT into TOGAF does not encounter methodological problems. An example is described in [Es18]. For an integration in EAFs other than TOGAF things become more difficult. Since, for instance, ARIS' expressiveness is significant higher compared to IT4IT, a model mapping is required. Further research is required for creating and validating such IT4IT mappings. Proof of concept is needed to validate that IT4IT can be used alongside other EAFs than TOGAF.

The reference architecture character results from the normative description of how to manage IT as a business for improving efficiency and agility. Similar to other best practice frameworks originating from industry (e.g., ITIL or COBIT), IT4IT was developed from the professional experience of its creators. The model incorporates expert knowledge on a specific way of managing IT. Applying IT4IT means to follow this direction. In contrast, classic EAFs may include reference models but are not intended to be reference architectures themselves. Rather, they provide a method for designing an individual EA tailored to an individual enterprise. This implies more freedom of design on the one hand but also more effort of design on the other hand.

IT4IT includes no method and its scope is the business of IT. Thus, this reference architecture cannot replace common EAFs. Its position may be described as IT-related extension of a comprehensive EAF, one of its main advantages being the possibility to define a tooling (automation) strategy and creating IT-city planning models as well as roadmaps to evolve towards best practise IT toolchains.

Bibliography

[AFW11] Aier, S.; Fischer, C.; Winter, R.: Construction and Evaluation of a Meta-Model for Enterprise Architecture Design Principles. In (Bernstein, A.; Schwabe, G. Eds.): Proc. of the 10th Int. Conf. on Wirtschaftsinformatik, Zurich, 2011; pp. 637–644.

[AG06] Abdallah, S.; Galal-Edeen, G. H.: Towards a framework for enterprise architecture frameworks comparison and selection: Proc. of the 4th Int. Conf. on Informatics and Systems (INFOS2006), 2006.

[Ak16] Akershoek, R.: IT4IT™ for Managing the Business of IT – A Management Guide. The Open Group, 2016.

[AKL99] Armour, F. J.; Kaisler, S. H.; Liu, S. Y.: A big-picture look at enterprise architectures. In IT Professional, 1999, 1(1); pp. 35–42.

[BJ16] Betz, C.; Jahn, K.: Defining "IT Service" for the IT4IT™ Reference Architecture, San Francisco, 2016.

[BNS03] Bernus, P.; Nemes, L.; Schmidt, G.: Handbook on Enterprise Architecture. Springer, Berlin, Heidelberg, 2003.

[BS12] Buckl, S.; Schweda, C. M.: On the State-of-the-Art in Enterprise Architecture Management Literature, 2012.

[Bu09] Buckl, S. et al.: Using Enterprise Architecture Management Patterns to Complement TOGAF: 2009 IEEE International Enterprise Distributed Object Computing Conference. IEEE, 2009; pp. 34–41.

[Bu17] Bui, Q. N.: Evaluate Enterprise Architecture Frameworks Using Essential Elements. In Communications of the Association for Information Systems, 2017, 41; pp. 121–149.

[BY15] Babar, Z.; Yu, E.: Enterprise Architecture in the Age of Digital Transformation. In (Persson, A.; Stirna, J. Eds.): Advanced Information Systems Engineering Workshops. Springer, Cham, 2015; pp. 438–443.

[De96] Defense Information Systems Agency Center for Standards: Department of Defense Technical Architecture Framework for Information Management. https://apps.dtic.mil/sti/citations/ADA321171, accessed 30 Jun 2020.

[DH11] Dietz, J. L. G.; Hoogervorst, J. A. P.: A critical investigation of TOGAF - based on the enterprise engineering theory and practice. In (Albani, A.; Dietz, J. L. G.; Verelst, J. Eds.): Advances in Enterprise Engineering V. First Enterprise Engineering Working Conference (EEWC2011). Springer, Berlin, Heidelberg, 2011; pp. 76–90.

[Es18] Estrem, W. A. et al.: How to Use the TOGAF® and IT4IT™ Standards Together. White Paper, 2018.

[Fr09] Franke, U. et al.: EAF2 - A Framework for Categorizing Enterprise Architecture Frameworks: 10th ACIS Int. Conf. on Software Engineering, Artificial Intelligences, Networking and Parallel/Distributed Computing. IEEE, 2009; pp. 327–332.

[Ga18] Gampfer, F.: Managing Complexity of Digital Transformation with Enterprise Architecture. In (Pucihar, A. et al. Eds.): 31st Bled eConference. University of Maribor Press, Maribor, 2018; pp. 635–641.

[HMS14] Henfridsson, O.; Mathiassen, L.; Svahn, F.: Managing Technological Change in the Digital Age: The Role of Architectural Frames. In Journal of Information Technology, 2014, 29(1); pp. 27–43.

[ID10] IDS Scheer AG: ARIS-Method Manual. ARIS IT Architect 7.1. Technical Document, 2010.

[KH17] Korhonen, J. J.; Halen, M.: Enterprise Architecture for Digital Transformation: 2017 IEEE 19th Conf. on Business Informatics (CBI). IEEE, 2017; pp. 349–358.

[LZ06] Leist, S.; Zellner, G.: Evaluation of current architecture frameworks. In (Haddad, H. M. Ed.): Proc. of the 2006 ACM symposium on Applied computing. ACM, New York, NY, 2006; pp. 1546–1553.

[Ma17] Masuda, Y. et al.: An Adaptive Enterprise Architecture Framework and Implementation. In International Journal of Enterprise Information Systems, 2017, 13(3); pp. 1–22.

[MV19] Masuda, Y.; Viswanathan, M.: Direction of Digital IT and Enterprise Architecture. In (Masuda, Y.; Viswanathan, M. Eds.): Enterprise Architecture for Global Companies in a Digital IT Era. Springer, Singapore, 2019; pp. 17–59.

[Ob20] Object Management Group (OMG): Unified Modeling Language™ (UML®) Specification. www.uml.org, accessed 28 Aug 2020.

[Pr16] Price, T.: IT4IT™: the new enterprise architecture framework. https://de.slideshare.net/TonyPrice11/it4it-bcs, accessed 28 Aug 2020.

[Ru91] Rumbaugh, J.: Object-oriented modeling and design. Prentice Hall, Englewood Cliffs, NJ, 1991.

[Sc06] Schekkerman, J.: How to survive in the jungle of enterprise architecture frameworks. Creating or choosing an enterprise architecture framework. Trafford, Victoria, 2006.

[Sc92] Scheer, A.-W.: Architecture of integrated information systems. Foundations of enterprise modelling. Springer, Berlin, 1992.

[Sc94] Scheer, A.-W.: Business Process Engineering. Reference Models for Industrial Enterprises. Springer, Berlin, Heidelberg, 1994.

[Sc99] Scheer, A.-W.: ARIS - Business Process Frameworks. Springer, Berlin, 1999.

[Su16] Sultanow, E. et al.: A multidimensional Classification of 55 Enterprise Architecture Frameworks: Surfing the IT innovation wave. 22nd Americas Conf. on Information Systems (AMCIS2016). Curran Associates, Red Hook, NY, 2016; pp. 1520–1527.

[TF17] Tambo, T.; Filtenborg, J.: IT4IT™ as a management of technology framework: perspectives, implications and contributions: Proc. of the 26th Conf. of Int. Association for Management of Technology (IAMOT2017), Vienna, 2017; pp. 1–14.

[TF19] Tambo, T.; Filtenborg, J.: Digital services governance: IT4IT™ for management of technology. In Journal of Manufacturing Technology Management, 2019, 30(8); pp. 1230–1249.

[Th17a] The Open Group: ArchiMate® 3.0.1 Specification, 2017.

[Th17b] The Open Group: The Open Group IT4IT™ Reference Architecture. Version 2.1, 2017.

[Th18] The Open Group: The TOGAF® Standard. Version 9.2, 2018.

[Th20a] The Open Group: IT4IT™ FAQ | The Open Group. https://www.opengroup.org/membership/forums/it4it-forum/it4it-faq, accessed 14 Jun 2020.

[Th20b] The Open Group: The IT4IT™ Reference Architecture. https://www.opengroup.org/it4it.

[TM17] TM Forum: GB921 Business Process Framework (eTOM) R17.0.1. https://www.tmforum.org/resources/suite/gb921-business-process-framework-etom-r17-0-1/, accessed 30 Jun 2020.

[TO20] TOGAF-Modeling: Modeling Enterprise Architecture with TOGAF. https://www.togaf-modeling.org, accessed 30 Jun 2020.

[UM06] Urbaczewski, L.; Mrdalj, S.: A Comparison of Enterprise Architecture Frameworks. In Issues in Information Systems, 2006, 7(2); pp. 18–23.

[vo15] vom Brocke, J. et al.: Standing on the Shoulders of Giants: Challenges and Recommendations of Literature Search in Information Systems Research. In Communications of the Association for Information Systems, 2015, 37, Article 9.

[Wa16] Warfield, D.: Why the IT4IT™ Standard is Good News for Architects. In Journal of Enterprise Architecture, 2016, 12(2); pp. 25–29.

[WF06] Winter, R.; Fischer, R.: Essential Layers, Artifacts, and Dependencies of Enterprise Architecture: 2006 10th IEEE International Enterprise Distributed Object Computing Conference Workshops (EDOCW'06). IEEE, 2006.

[Wi10] Winter, K. et al.: Investigating the State-of-the-Art in Enterprise Architecture Management Methods in Literature and Practice: Proc. of the 5th Medit. Conf. on Information Systems (MCIS201). AIS Electronic Library (AISeL), 2010.

[Zi16] Zimmermann, A. et al.: Adaptive Enterprise Architecture for Digital Transformation. In (Celesti, A.; Leitner, P. Eds.): Advances in Service-Oriented and Cloud Computing. Springer, Cham, 2016; pp. 308–319.

Die IT-Finanzarchitektur im Cloudumfeld

Darstellungsweisen und Empfehlungen aus der Beratungspraxis

Dr. Carsten Brockmann[1], Christian Schneider[2], Mario Schmitz[3], Thomas Klingspor [4]

Abstract: Im Bereich des Architekturmanagements existiert eine Vielzahl von Architekturmodellen und branchenspezifischen Ausprägungen. In diesem Beitrag stellen wir einen Ansatz vor, mit dem die IT-Architektur der Finanzfunktion in einem Cloudumfeld dargestellt werden kann.

Keywords: Business; Finance; IT; Architecture

1 Einleitung

Referenzarchitekturen verfolgen üblicherweise verschiedene Ziele. Zum einen soll der Wissenstransfer in einem bestimmten Bereich durch Beschreibung der relevanten Spezifika ermöglicht werden. Zum anderen soll die strukturelle Darstellung als Ausgangsbasis für die inhaltliche Diskussion anhand der Klassifikationskriterien dienen.

2 Die Finanzfunktion als Gestalter von Werteflüssen

Die Finanzfunktion umfasst ein breites Spektrum an Aufgaben, Prozessen und Funktionen, beginnend von der buchhalterischen Erfassung der Zahlungsvorgänge bis zur Gestaltung von Werteflüssen. Durch die kontinuierliche Aggregation von Zahlungsströmen zu Informationen und Wissen über die Unternehmenslage verfügt die Finanzfunktion über ein einzigartiges Know-How vor Ort, dass es erlaubt, das Handeln der anderen Bereiche zielgerichtet zu analysieren um eine effiziente, kosten- und leistungsoptimale, Ressourcenallokation zu ermöglichen. Dies könnte durch die Rolle als Business Partner [SK14] erfolgen.

Die durchgängige Transparenz von Finanzinformationen und die Integrität dieser Informationen entlang der Wertschöpfungskette ist ein Imperativ in der heutigen, vernetzten Organisation. Idealerweise sind die rein betragsmäßigen Finanzinformationen auch um entsprechende Mengen- und weitere Detailinformationen angereichert. Es sind insb. diese

[1] Deloitte Consulting GmbH, Kurfürstendamm 23, 10719 Berlin, cbrockmann@deloitte.de
[2] Deloitte Consulting GmbH, Rosenheimer Platz 4, 81669 München, chrischneider@deloitte.de
[3] Deloitte Consulting GmbH, Schwannstr. 6, 40476 Düssledorf, marschmitz@deloitte.de
[4] Deloitte Consulting GmbH, Rosenheimer Platz 4, 81669 München, tklingspor@deloitte.de

doi:10.18420/inf2020_18

erweiterten Informationen, die in Zukunft gerade im Anwendungsfeld von Big Data und Analytics der Finanzfunktion die Möglichkeit bieten, das Handeln der anderen Organisationseinheiten zu orchestrieren.

3 Die Modellierung der Finanzfunktion

Bei der Darstellung der Finanzfunktion sind der Forschungsstand aus Mehrschichtenarchitekturen, TOGAF sowie Erfahrung aus diversen Beratungsprojekten eingeflossen. Ziel der Modellierung ist es, den Sachverhalten einen Ordnungsrahmen zu geben und Auswirkungen auf Veränderungen darzustellen [Br14, Su13]. Die in Abbildung 1 dargestellten Elemente werden im Folgenden näher beschrieben.

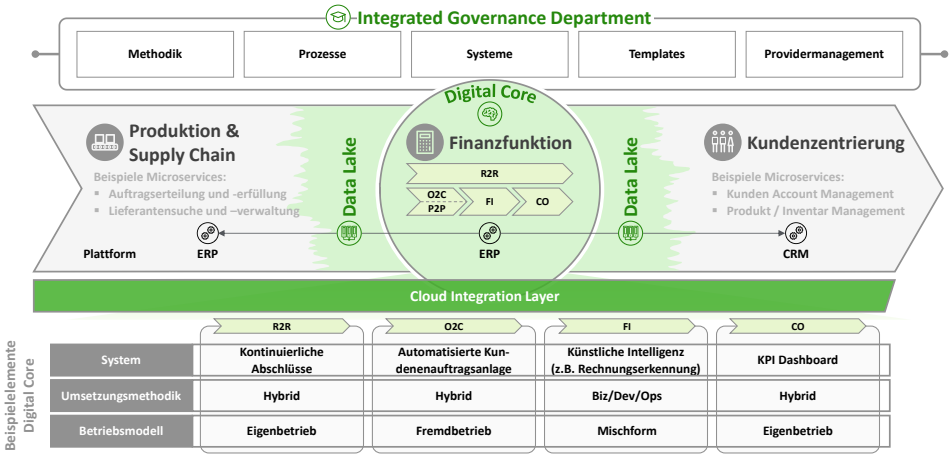


Abb. 1: IT Finanzarchitektur

3.1 Governance

Im Rahmen der Governance-Funktion werden übergreifende Standards definiert. Bei der Methodik werden die Vorgehensmodelle zur Projektumsetzung vorgegeben und deren Ausgestaltung definiert, beispielsweise mittels agiler Ansätze [Ep19]. Daneben bekommen Aspekte der Prozess Governance zunehmend Bedeutung. Diese umfassen zum einen Vorgaben hinsichtlich der Prozessmodellierung (insb. Methodik), aber auch hinsichtlich Prozessstandardisierung (Etablierung und kontinuierliche Überwachung des Prozessharmonisierungsgrads) und ggfs. Automatisierung. Hierbei finden sogenannte Template-Prozesse Verwendung, die neben der Vereinheitlichung von Prozessen auch die Beschleunigung von Rollouts zum Ziel haben. Im Hinblick auf die Systeme werden Rahmenparameter bezüglich des Softwareökosystems und der Anbindung von Microservices/Drittlösungen gesetzt. Das Providermanagement setzt den Handlungsrahmen für externe Dienstleister.

3.2 Der Finanzfunktion im Kern der Prozessdarstellung

Traditionelle Wertschöpfungsketten beginnen mit der Fertigung von Produkten oder Erstellung von Dienstleistungen. Damit verknüpft sind die Finanzprozesse, die Wertströme darstellen, dokumentieren und Handlungsoptionen aufzeigen. Die letzte Komponente stellt die Kundenzentrierung dar, in der sich die Interaktionspunkte mit dem Kunden sowie zugrundeliegenden Abwicklungssysteme befinden. Hier kann die Finanzfunktion über ihre Rolle als Orchestrator kundenfokussierte Deckungsbeiträgsanalysen ermöglichen um beispielsweise Produkt-, Kunden- oder regionale Unterschiede aufzudecken.

Für die Erstellung von Produkten/Dienstleistungen sowie deren Abbildung in den Finanzprozessen werden üblicherweise die meisten Prozessschritte in einer ERP-Plattform ausgeführt, bei der Kundenzentrierung hingegen in der CRM-Plattform. Die Datenanbindung und Datenvorhaltung findet entweder über die spezifischen Datenquellen oder über einen Data-Lake statt.

In der Vergangenheit war die Finanzfunktion oft stark nach einzelnen Einheiten aufgebaut. In Zukunft ist die Kundenzentrierung sowie die damit verbundenen Werteflüsse ausschlaggebend für die Ausgestaltung der Finanzfunktion. Kunden der Finanzfunktion sind zum einen andere organisatorische Einheiten wie beispielsweise der Einkauf, Vertrieb, Produktion. und zum anderen externe Adressaten.

Waren Controlling, i.S.v. interne Leistungssteuerung, Managementberichtswesen und Planung, und Rechnungswesen, i.S.v. Leistungserfassung, Abrechnung und externe Berichtslegung, bisher klassisch getrennt, so wachsen diese Teildisziplinen der Finanzfunktion zunehmend zusammen, um so den internen und externen Kunden ein umfassendes und holistisches Bild der aktuellen Unternehmenssituation zu geben, aber auch valide Entscheidungsgrundlagen für operative, taktische und strategische Fragestellungen zu leisten.

Die digitalen Kernprozesse der Finanzfunktion beginnen mit den Order to Cash (O2C) und Purchase to Pay (P2P) Prozessen. Bei O2C werden die Schritte von der Kundenbestellung bis zur Zahlung durch den Kunden betrachtet. Bei P2P werden die Schritte von der Bestellung einer Organisation bis zur Zahlung des Lieferanten betrachtet. Während diese Prozesse klassisch in den operativen Funktionen (hier: Vertrieb und Einkauf) beginnen und enden, so werden durch den zugrundeliegenden Belegfluss entsprechende Finanzinformationen generiert. Im O2C sind dies beispielsweise (verkürzt) das Angebot, die Kundenbestellung, die Kundenrechnung, sowie der Lieferschein. In P2P werden entsprechend die Bestellanforderung, Bestellung, Lieferschein und Lieferantenrechnung generiert. Die innerhalb dieses Belegflusses erzeugten Finanzinformationen bilden zum einen die Basis für die durchgängige Betrachtung und Steuerung durch Finanzen, aber auch nachgelagert die Grundlage für den sog. Record to Report (R2R) Prozess. Im Zuge des R2R Prozess werden aus den o.g. Informationen des Belegflusses die jeweiligen Informationen für monatliche bzw. quartalsweise Berichte erzeugt, sowie der vorgeschriebene Jahresabschluss erstellt.

Durch die steigende Industrialisierung von Prozessen der Finanzfunktion wurde der Effizienzgrad kontinuierlich erhöht - so werden beispielsweise heute schon Berichte durch sogenannte Reportingfactories [SLD16] in sehr standardisierter Form erstellt und damit Skalen- und Qualitätsvorteile erzielt, sowie ein Leverage von Lohnarbitragen gerade bei parallelen Standortverlagerungen möglich.

3.3 Cloud Integration Layer

Die prozessübergreifende Integration von Applikation ist schon immer eine Herausforderung in vernetzten, arbeitsteilig agierenden Unternehmen. Die dabei zu erzielende Informationstransparenz und –durchgängigkeit stellt insb. für die Finanzfunktion einen wesentlichen Erfolgstreiber dar.

Integrationsarchitekturen unterlagen in den letzten Jahren starken technologischen Innovationen – beispielhaft sei an dieser Stelle nur an das Aufkommen service-orientierter Architekturen genannt [FT]. Aktuell werden hierfür durch Unternehmen zunehmen sogenannte Cloud Architekturen diskutiert und sukzessive etabliert. In diesem Zusammenhang wird in der Praxis auch vom „Cloud Integration Layer“ gesprochen. Dieser soll mittels Verwendung von Cloud Methoden und Paradigmen die kontinuierliche, elastische und insb. infrastrukturunabhängige Integration von Prozessen, Daten und Anwendern ermöglichen.

Prozesse können dabei in Legacy Applikationen ausgeführt werden, als Funktionalitäten neuer Applikation bereitgestellt werden oder auch im Sinne von externen der intern Microservices konsumiert werden. Microservices können dabei neben der reinen Prozessausführung auch zunehmend durch Prozessautomatisierung (Robotic Process Automation) und künstliche Intelligenz (KI) zu einer integrierten digitalen Prozesswertschöpfung kombiniert werden.

Mit der Nutzung des Cloud Integration Layers gehen verschiedene Vorteile einher:

- **Verkürzte Time-to-Market:** Durch die Anbindung spezifischer Cloud-Systeme kann ausgewählte Funktionalität deutlich schneller bereitgestellt werden, als dies bei der Umsetzung auf der ERP-Plattform der Fall wäre
- **Skalierbarkeit:** Bei steigendem Transaktions- und Datenvolumen gibt es keine Änderung an der Ausführungsgeschwindigkeit, da der Cloudanbieter die notwendigen Kapazitäten auf seiner Seite entsprechend bereitstellt
- **Agiles Arbeiten und höherer Innovationsgrad:** Die Anpassung von Funktionalität in Cloud-Systemen erfolgt sehr häufig im agilen Kontext, so dass der Fachbereich, die IT und der Cloudanbieter sehr zügig neue und innovative Funktionalität entwickeln und Live setzen

3.4 System

Die Verarbeitung von Daten kann durch unterschiedliche Anwendungen erfolgen. ERP-Plattformen stellen üblicherweise die Basisfinanzdaten bereit. Bei analytischen Anwendungen wie beispielsweise KI werden Daten analysiert und Schlussfolgerungen daraus gezogen, die die Grundlage für die Verarbeitung durch den Menschen, einen Bot oder eine klassische Transaktion darstellen. Über Microservices werden verschiedene Drittanwendungen angebunden, die üblicherweise einen klar umrissenen und kleinen Funktionsumfang haben. Individualentwicklungen können gezielte Funktionalität beisteuern, wie z.B. Anweisungen zum steuerlichen Umgang mit Mitarbeiterkleidung.

3.5 Umsetzungsmethodik

IT Organisationen adaptieren zunehmen agile Methoden [Ep19, Br19]. Im Zuge dessen wird das klassische Modell der Erstellung und des Betriebs von Applikationen auch für den Finanzbereich auf DevOps umgestellt. DevOps beschreibt dabei das Zusammenlegen von bisher getrennten Aktivitäten zur Weiterentwicklung und zum Betrieb von IT Anwendungen, mit dem Ziel hierdurch sowohl die Agilität dieser Anwendungen zu erhöhen, dabei aber auch die Total Cost of Ownership (TCO) zu reduzieren. Die dadurch neu etablierten interdisziplinären (IT)-Teams werden in der Regel um Business Ansprechpartner erweitert, bspw. aus der Finanzfunktion, um so frühzeitig die Fachanforderungen in Entwicklung und Betrieb von Applikationen zu berücksichtigen (in diesem Zusammenhang wird auch von BizDevOps gesprochen).

Neben dem Einsatz von agilen Methoden (bspw. Scrum) finden sich in den Unternehmen aber weiterhin klassische Wasserfallmethoden im Bereich Applikationsentwicklung – mit der zugehörigen Organisation.

Ebenso lässt sich auch der Einsatz hybrider Modelle beobachten, die Elemente der Agilität mit traditionellen Elementen verknüpfen (bspw. im Bereich ERP Entwicklung: Neuentwicklung unter Verwendung agiler Methoden vs. Rollout der ERP Lösung in Länder / Märkte nach Wasserfallmethodik).

3.6 Betriebsmodell

Beim Eigenbetriebs werden alle IT Aktivitäten durch die Organisation erbracht. Im Rahmen des Fremdbetriebs werden insb. Hardware, Infrastruktur, aber auch Serverbetrieb und –wartung an externe Dienstleister ausgelagert. Application Management Services (AMS) erweitern oder ergänzen dies um den Betrieb und die Wartung einzelner Anwendungen (bspw. ERP System). In letzterem Szenario verbleiben im Unternehmen vornehmlich Key User mit tiefem Applikationsverständnis, die jedoch nicht mehr in Wartungs- und

Entwicklungsaufgaben für die jeweilige Applikation eingebunden sind. Diese werden von der IT Organisation in sog. On-demand Szenarien abgebildet.

Möglich sind gemischte Modelle, bei der zwar das eigene Personal den Betrieb übernimmt, die Infrastruktur allerdings von einem externen Dienstleister bereitgestellt wird.

4 Fazit

In dem hier vorgestellten Modell werden die Spezifika der Finanzfunktion durch verschiedenen Ebenen hinweg mitsamt der Anbindung an die Cloud berücksichtigt. Der prozessorientierte Aufbau ermöglicht die Darstellung von Brüchen in der Nutzung der verschiedenen Anwendungssysteme und kommt gleichzeitig der Forderung nach, Geschäftsprozesse und Technologie enger zu verzahnen [BSC19].

Die intensivere Nutzung von Microservices und Drittsystemen sowie der zunehmende Funktionsumfang und Integrationsgrad der ERP Plattformen führt zu einer steigenden architekturellen Komplexität die ex ante bei heutigen Entscheidungen zur Renovierung oder Ausbau bestehender Finanzapplikationslandschaften zu berücksichtigen sind. Dabei ist die organisationale Komplexität heutiger, vernetzter Unternehmen mit zu berücksichtigen: Prozessvarianten in Tochter- oder Landesgesellschaften werden zunehmend zum Hemmnis für Digitalisierungsbestrebungen, so dass diese zunehmend abgebaut bzw. vorwärtsgerichtet vermieden werden sollten. Um diese Anforderungen zu erfüllen, kristallisiert sich zunehmend eine zentrale Governancestelle als Lösungsform heraus. Das Aufgabengebiet der Governancestelle besteht aus dem Setzen und überwachen von Standards bezüglich der Prozesse und Systeme sowie den damit in Verbindung stehenden Templates und Providermanagement. Als weiteres Element stellt die Governancefunktion die Einhaltung der Grundsätze ordnungsgemäßer Buchführung und Datenhaltung sicher.

5 Acknowledgements

Wir möchten Herrn Florian Kuchler aus dem Digital Finance Chapters bei Deloitte Consulting für seine Mitwirkung herzliche danken. Ebenso möchten wir Alexander Jahnke von Deloitte Consulting für seine Impulse danken.

Literaturverzeichnis

[SK14] Schmitz, M., Koelzer, C.: Was bedeutet Business Partnering im Controlling? Controlling & Management Review 2, 33 (2014)

[Br14] Brockmann, C.: An approach to design the business model of an ERP vendor. GITO, Berlin (2014)

[Su13] Sultanow, E., Brockmann, C., Pratt, R., Andresen, K.: Integrate Enterprise Systems to our Hyperconnected World: Anything, Anywhere, Anytime through architectural design. 19th Americas Conference on Information Systems (AMCIS), Chicago, USA (2013)

[Ep19] Epstein, R., Klingmann, P., Kroker, M., Brockmann, C.: Die Finanzfunktion als Accelerator einer agilen Organisation. CFO Insights November, 1 - 7 (2019)

[SLD16] Schmitz, M., Lawrenz, A., Drerup, B.: Reporting Factory in Controllerbereichen. In: Becker, W., Ulrich, P. (eds.) Handbuch Controlling, pp. 427 - 458. Springer, Wiesbaden (2016)

[FT] https://www.ft.com/content/e47f68d6-6eae-11d9-94a8-00000e2511c8

[Br19] Brockmann, C., Nagel, C., Kahl, S., Biermann, A.: Stepping stones to an agile enterprise. Deloitte Review 36 - 47 (2019)

[BSC19] Brockmann, C., Sultanow, E., Czarnecki, C.: Is Enterprise Architecture still relevant in the Digital Age? In: Draude, C., Lange, M., Sick, B. (eds.) INFORMATIK 2019: 50 Jahre Gesellschaft für Informatik – Informatik für Gesellschaft (Workshop-Beiträge), pp. 21. Gesellschaft für Informatik e.V., Bonn (2019)

"HySLAC" – A Conceptual Model for Service Level Agreement Compliance in Hybrid Cloud Architectures

Michael Seifert[1] Stephan Kuehnel [2]

Abstract: Cloud computing provides IT infrastructures and services via networks, and it enables economic potentials for end users as well as a focus on core competencies. In addition to its extensive potentials, cloud computing in general, and hybrid cloud computing in particular, pose new challenges in the negotiation and formulation of Service Level Agreements, as well as in the monitoring of and compliance with contractual requirements. An understanding of cloud service and deployment models, perspectives, roles, and contractual terms is essential for a successful and compliant adoption of hybrid clouds. Consequently, this paper proposes the novel HySLAC model, focusing on service level agreement compliance in hybrid cloud architectures. Based on eight model requirements and a systematic literature review, the HySLAC model was conceptualized with UML 2.0. It comprises eight UML classes and five associated enumerations, and it is instantiated by means of a case study. The model offers scientific and practical application capabilities for the analysis of service components as well as hybrid cloud service compositions, and it opens up potentials for decision support.

Keywords: Cloud Computing; Service Level Agreement; Quality of Service; Compliance

1 Introduction

Enterprise Architectures (EA) are constantly faced with new issues in the digital age [BSC19]. A central challenge in this context results from the increasing spread of cloud computing, whose various forms directly impact organizational EA [KGN09]. Cloud computing is characterized as ubiquitous on-demand access to available computing resources via networks, and it has become very attractive for companies [PRS09] [MG11]. Cloud service providers today are offering highly available storage, complex services, development platforms, and massive parallel computing resources at relatively low cost and without any need for implementation on the customer side [Li15]. The possibility of paying for necessary services and resources depending on use, i.e., a so-called “pay-per-use"basis, offers companies substantial economic advantages [Ya16] [PRS09]. Therefore, it is not surprising to see a growing trend in practice to move formerly in-house service systems into the cloud [PRS09]. Such a shift also enables companies to concentrate on their core competencies while avoiding unnecessary back-office activities [PRS09]. However, in addition to the potentials that cloud computing opens up for companies, there are also new challenges for Enterprise Architects in general and Enterprise Cloud Architects in

[1] GISA GmbH, Leipziger Chaussee 191A, 06112 Halle (Saale), michael.seifert@gisa.de
[2] Martin Luther University Halle-Wittenberg, Chair for Information Management, Universitaetsring 3, 06108 Halle (Saale), stephan.kuehnel@wiwi.uni-halle.de, https://orcid.org/0000-0002-6959-9555

doi:10.18420/inf2020_19

particular [BB18], such as the negotiation and formulation of contracts between cloud providers and cloud customers, so-called Service Level Agreements (SLAs), and the monitoring and verification of compliance with service quality requirements, so-called Qualities of Service (QoS) [JAB12]. These challenges become even more complex when cloud providers adopt the dual role of provider and customer, e.g., when their own cloud service offerings build on the services of external cloud providers. The resulting "hybrid cloud" (HC) architectures establish SLA hierarchies [Co09] and hamper both monitoring and compliance with contractual requirements and QoS [JAB12].
The understanding of different cloud deployment and cloud service models, different perspectives, dependencies, SLAs, QoS, and roles associated with the introduction and user acceptance of cloud computing in general and HC computing in particular is of fundamental importance for Enterprise Cloud Architects, as it provides the foundation for a successful and compliant (hybrid) cloud deployment. The current state of research already provides conceptual models addressing individual relevant aspects of SLA compliance in HC architectures, such as cloud service models [YBD08], cloud deployment models [LMB16], SLAs, and QoS [KPP16] [Gu11] [Co09] [Th10]. However, to the best of our knowledge and belief, a conceptual model holistically addressing all cloud service models, cloud delivery models, contractual aspects, associated roles, and existing dependencies in the context of SLA compliance of HC architectures is still missing. Consequently, the research objective (RO) of our study is as follows:

RO: *The goal of the study is to conceptualize a model for service level agreement compliance in hybrid cloud architectures, considering cloud service models, cloud deployment models, involved perspectives and roles, as well as related contractual aspects.*

In order to address this research goal, we first derive relevant model requirements (MRs) from the theory in Section 2. Using the rigorous method of vom Brocke et al. [vo09] described in Section 3, we subsequently analyze relevant theoretical domain knowledge with respect to the identified MRs (Section 4). Based on the insight that no existing conceptual model maps all aspects of SLA compliance in HC architectures, we present a novel conceptual model, called "HySLAC," in Section 5. The HySLAC model maps cloud deployment models, cloud service models, business and IT perspectives, SLAs, SLA templates, QoS, as well as relevant roles and dependencies in the context of SLA compliance of HC architectures. Our model is instantiated by means of a case study in Section 6. The paper concludes with a summary and a discussion of limitations in Section 7.

2 Theoretical Background and Model Requirements

In the following section, we address the theoretical background underlying this study as well as necessary concepts relevant to the problem context. The theoretical background serves to derive MRs with relevance for Enterprise Cloud Architects and provides us with a theoretical basis for a conceptual model of SLA compliance in HC architectures.

2.1 Cloud Theory

Our investigation is based on Cloud Theory (CT). Following the well-known treatise of the National Institute of Standards and Technology (NIST), three service models and four deployment models can be distinguished in the context of CT [MG11]. The service models are differentiated into Software as a Service (SaaS), Platform as a Service (PaaS), and Infrastructure as a Service (IaaS) [MG11]. IaaS addresses the operation of IT infrastructures by a provider who makes them available to end users "as a service" and provides the foundation for both PaaS and SaaS, since it is impossible to deliver platforms and/or software without infrastructure in terms of a service [MG11]. SaaS addresses the provision of software and related IT infrastructures for end users [MG11], and PaaS provides platforms and related IT infrastructures for application developers [Ya16] [MG11]. These service models give rise to MR1:

MR1: *Consideration of Software, Platform, and Infrastructure as a Service.*

In CT, the deployment models are divided into Private Clouds (PiC), Community Clouds (CC), Public Clouds (PuC), and Hybrid Clouds (HC) [MG11]. The PiC model is characterized by the fact that the cloud infrastructure is provided for exclusive use by a single organization [Go14]. However, PiC are often used by several customers, such as those of different business units [MG11] [Go14]. The PuC model includes cloud infrastructures for open use by the general public, and these can be owned by a business, academic, or governmental organization [Go14] [MG11]. The CC model combines the properties of the PuC and PiC models in line with requirements of one target user group [Go14]. The CC model is quite similar to the PiC model, but the infrastructures and computing resources are not exclusively available to one organization, but to two or more organizations [MG11] [Go14]. Thereby, CC aims to address common concerns of multiple organizations. The architecture of the HC model is far more complex than the others as it is composed of two or more different deployment models (PuC, CC, and/or PiC) that remain independent entities but are technologically combined [Go14] [MG11]. The second requirement for the conceptual model can be derived from the discussion of the deployment models as follows:

MR2: *Consideration of Private, Community, Public, and Hybrid Clouds.*

2.2 Involved Perspectives and Roles

The adoption of cloud services is associated with economic risks [PJW10], e.g., arising from downtime in business activities. These risks affect IT services (i.e., the IT perspective), business processes (i.e., the business perspective), and their intersections. Thus, the model-based connection between IT services and business processes represents MR3:

MR3: *Connection between IT services and business processes.*

The separation of involved roles into service providers and service customers is an essential feature for mapping cloud scenarios [MSY16]. In contrast to traditional, internal IT provisioning, the distinction between the interests of the roles involved must be taken into consideration [PM15]. However, in the context of HC architectures, the separation of roles into service providers and service customers reaches its limits. These limits result from the fact that the provider of a HC service is itself a user of a non-hybrid (ordinary) cloud service [PJW10]. In the given problem context, a total of four roles have to be distinguished: 1) the non-hybrid (ordinary) cloud service customer, 2) the non-hybrid (ordinary) cloud service provider, 3) the HC service customer, and 4) the HC service provider. Thus, we derive the fourth model requirement, which addresses the distinction of the four roles involved:

MR4: *Consideration of hybrid (4a) and non-hybrid (4b) cloud service customers, as well as hybrid (4c) and non-hybrid (4d) cloud service providers.*

2.3 Operational/Service Level Agreements and Qualities of Service

A SLA is a contract for an agreed IT service between a provider and a customer [PRS09]. The definition and formulation of SLAs is not trivial for either (hybrid) cloud service customers or (hybrid) cloud service providers. The reason for the non-triviality is that SLAs have to be negotiated individually in each case as uniform standards are largely lacking [Al15]. For example, the definitions and formulations of contractual penalties for non-compliance with promised performance levels of cloud services are generally diverse, which hampers a uniform risk assessment for service providers and customers [Al15]. A promising approach to address this challenge is the definition of SLA templates allowing for both providers and customers to use a consistent set of specifications [Br09]. On this basis, we define the consideration of SLA/SLA templates as the fifth model requirement:

MR5: *Consideration of Service Level Agreements/Service Level Agreement Templates.*

Based on the role model presented, we distinguish between SLAs and Operational Level Agreements (OLAs). An OLA is defined as an agreement between different service providers to ensure a SLA [KHB02]. Furthermore, an OLA is understood to be an agreement between different units within one service provider company [NK07]. The main difference between SLAs and OLAs is that OLAs are usually easier to negotiate and do not constitute a contract. We need to represent OLAs as a specialized form of SLA whenever two service components of the same service provider are used as part of a service composition. Thus, the sixth model requirement is as follows:

MR6: *Consideration of Operational Level Agreements.*

To measure and evaluate agreed performance levels of cloud services, QoS are commonly used [Su12]. Complex (hybrid) cloud architectures require an aggregation of QoS of individual service components that depend on the underlying service model and service architecture based on horizontal and vertical integration [BN13]. Common QoS aggregations are based, e.g., on multiplication and addition, the determination of local and global minima/maxima, and mean value calculations [Qi12]. Thus, the seventh model requirement addresses the representation of QoS as an offspring of the SLA and the consideration of horizontal and vertical aggregations of QoS across involved service components:

MR7: *Consideration of Qualities of Service and their horizontal/vertical aggregations.*

SLAs of PuC, especially for services that follow the SaaS model, are often non-negotiable due to the "one-to-many" relationship between a service provider and multiple service customers. According to [Co09], PuCs are usually defined as one-sided SLAs, i.e., as non-negotiable. Considering the different cloud deployment models which we already specified, different levels of negotiability can be derived: 1) full negotiability, 2) non-negotiability, and 3) partial negotiability. Partial negotiability occurs whenever at least one component of a service composition is non-negotiable. Consequently, the representation of the negotiability of an SLA and its levels represents our eighth and last model requirement:

MR8: *Consideration of the negotiability of Operational/Service Level Agreements and their levels.*

3 Research Method

The conceptual model for SLA compliance in HC architectures is to be constructed based on domain knowledge that we extract from the knowledge base of information systems research. Following the recommendations of Levy and Ellis [LJ06], we investigate this knowledge base by means of a systematic literature review (SLR). Moreover, to ensure scientific rigor in the procedure for carrying out our research, we rely on the well-known method used for SLRs proposed by vom Brocke et al. [vo09]. This method comprises a total of five steps: 1) definition of a review scope, 2) conceptualization of the topic, 3) literature search, 4) literature analysis, and 5) a summary and discussion of the research output.

3.1 Review Scope

For defining our review scope (step 1 according to [vo09]), we use Cooper's taxonomy of literature reviews [Co88]. As shown in Figure 1, the scope of our SLR is specified in terms of six characteristics.

Characteristics	Categories			
(1) Focus	**Research outcomes**	Research methods	Theories	Applications
(2) Goal	**Integration**	Criticism		**Central issues**
(3) Organization	Historical	**Conceptual**		Methodological
(4) Perspective	**Neutral representation**		Espousal of position	
(5) Audience	**Specialised scholars**	General scholars	**Practitioners**	General public
(6) Coverage	Exhaustive	**Exhaustive but selective**	Representative	Central or pivotal

Legend: **Selected category** | Unselected category

Fig. 1: Review scope of the systematic literature analysis

We investigate the literature base according to conceptual models, domain models, meta-models, and ontologies, hence our research focuses on the identification of corresponding research outcomes (1). In this context, the aim of our investigation is both to identify domain-specific concepts and to integrate related relevant issues (2). The organization of our review results is conceptual (3), as we aim to integrate the core concepts of the subject area as part of a comprehensive conceptual model. The research papers identified by the SLR and the conceptual model are presented and discussed neutrally (4). The research results will be particularly relevant for specialized scholars who deal with current problems of SLA compliance in the context of HC architectures, such as Enterprise Cloud Architects. Moreover, the conceptual model can be used as a starting point for practical reflections on dependencies, opportunities, and risks of cloud adoptions (5). Finally, the coverage of the SLR is comprehensive but selective (6). On the one hand, we selected well-known databases and specific search terms for our literature search. On the other hand, we conducted a comprehensive analysis of all search results based on this selective specification.

3.2 Conceptualization of Topic and Search for Topic-Related Literature

The conceptualization of the topic (step 2 according to [vo09]) has already been discussed in Section 2, in which we analyzed the theoretical background of our study, including necessary concepts, and derived relevant MRs for our later concept-centered analysis.
The five well-known databases Springer Link, EBSCOhost, Science Direct, Association for Computing Machinery Digital Library (ACM DL), and Association for Information Systems electronic Library (AISeL) were used for the literature search (step 3 according to [vo09]; see Figure 2). Following our discussion of the theoretical background, the literature search was initially begun with the search terms «service level agreement*» and «hybrid cloud*». However, the search with these terms using truncation (*) and an OR operator in all search fields led to an exorbitant amount of results (>10,000). Consequently, the search string was adjusted. Following the definition of the service models of NIST [MG11], we limited the search scope by adding the subject-relevant search term «*as a service». An initial look at the search results showed that QoS play a major role in complying with SLAs. Therefore, the term «qualities of service» was integrated by an OR operator. Finally, the search string

for searching the five databases was «(service level agreement* OR (quality OR qualities) of service) AND hybrid AND *as a service».

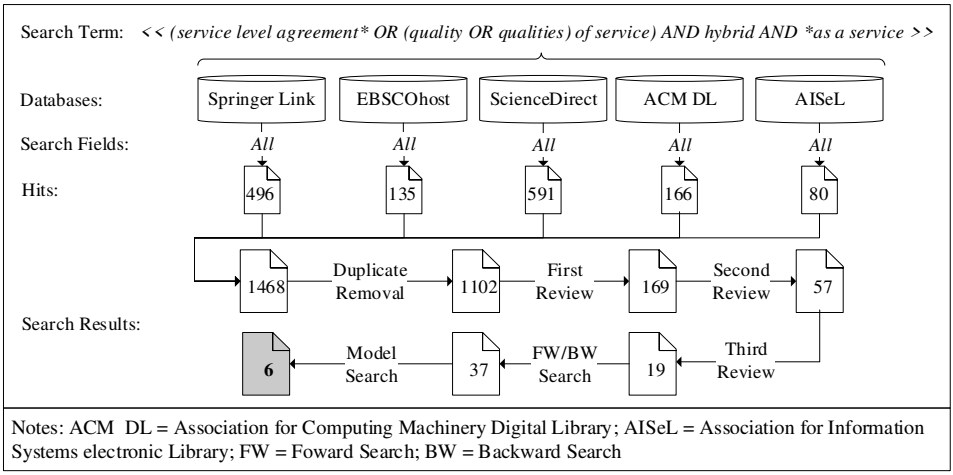


Fig. 2: Literature search process and search results

The literature management was done with Citavi 6. After implementing all search results in Citavi, we identified 366 duplicates. The remaining 1,102 hits were subjected to the first title-based selection, with 933 papers clearly sorted out due to irrelevance to the research problem. In the second evaluation phase, the abstracts of the remaining 169 publications were analyzed. The analysis of the abstracts was done by two independent researchers. Papers that were classified inconsistently were considered for further review. In the third review phase, the remaining 57 papers were subjected to a full text evaluation. As a result, 19 relevant papers were identified. Following the recommendations of Webster and Watson [WW02], we carried out an additional forward and backward search. This procedure enabled us to identify 18 additional relevant papers. The final 37 relevant papers were examined for conceptual models, domain models, meta-models, ontologies, frameworks, and further kinds of model-based conceptualizations which at least tackle the problem space of our study. The final search step led to a total of six relevant search results, which were subjected to a detailed model analysis. Brief summaries of the core contributions of the papers can be found in Appendix 1 (available at *https://bit.ly/3eKzp5h*).

4 Systematic Review of relevant Literature

The six relevant studies were analyzed by two independent researchers with regard to the fulfilment of the eight MRs (step 4 according to [vo09]). Due to consistent evaluation results, there was no need to discuss deviating evaluations. The results of the literature analysis can

be found in Table 1, where a complete and partial representation of an MR by a source is represented by half (◐) and completely filled (●) circles.

	Model requirements (MR)										
Source	MR1	MR2	MR3	MR4a	MR4b	MR4c	MR4d	MR5	MR6	MR7	MR8
[Co09]	◐		●		●	◐	◐	●			◐
[KPP16]					●		●	●		●	●
[Gu11]										●	
[LMB16]	●	●			●	◐	●	●	◐	●	
[Th10]	◐		●	●	●	●	●	●			
[YBD08]	●										

Notes: ● = model requirement is completely represented; ◐ = model requirement is partially represented; empty cell = model requirement is not represented.

Tab. 1: Evaluation of search results

The SLA management framework of Comuzzi et al. [Co09] and the architecture for multi-level SLAs by Theilmann et al. [Th10] partially address MR1, since they distinguish software and infrastructure layers but disregard the PaaS model. The unified cloud computing ontology of Youseff et al. [YBD08] completely addresses MR1 by describing the cloud service models as cloud layers. The ontology "CSLAOnto" of Labidi et al. [LMB16] contains specifications of deployment and service models completely addressing MR1 and MR2. The model of [Co09] completely addresses MR3 by linking software bundle SLAs and business SLAs. The model of [Th10] completely depicts MR3 as the role of the business manager agreeing with the SLA and the service provider being responsible to the customer. The model of [Co09] completely represents the non-hybrid service customer (MR4b), but cloud and HC service providers are only partially mapped (MR4c/d), as these roles are not clearly delineated. [KPP16] present a fine-grained depiction of SLA facets and related non-hybrid roles, thus completely addressing (MR4b/d). [LMB16] completely address MR4b/d by "SupportingParty" and "SignatoryParty" classes as well as corresponding cloud customers and cloud providers. Non-HC service providers (MR4c) are addressed, but they are not sufficiently differentiated from HC service providers. [Th10] completely distinguish ordinary and HC service providers (MR4c/d), i.e., so-called "Service Aggregators," as well as customers of hybrid and non-HC services (MR4a/b). The consideration of SLAs (MR5) is completely represented by the SLA management framework of [Co09], the SLA facets of [KPP16], the SLA classes of [LMB16], and the SLA definitions of [Th10]. In [LMB16], OLAs (MR6) are only partially addressed (via "SupportingParty"), as the possibility of drawing conclusions on corresponding QoS is lacking. The capability to model QoS (MR7) is completely addressed by the QoS facet of [KPP16], the QoS level models of [Gu11], as well as the CSLAOnto of [LMB16]. The model of [Co09] partially depicts negotiability (MR8), but it lacks negotiation levels and implications. In contrast, [KPP16] completely represent negotiability through attributes that can be modeled at the finest QoS levels.

5 Conceptualization of the "HySLAC" Model

After analyzing the literature and determining that none of the existing models fully covers the identified MRs, we propose a novel conceptual model for SLA compliance in HC architectures called "HySLAC" (step 5 according to [vo09]). The model construction was carried out considering the methodology of March and Smith [MS95]. Hence, conceptual models can be built on the basis of domain knowledge and serve to represent new theories and/or phenomena by elements and their associations [MS95]. According to [MS95], such models are not primarily about truth but about utility.
The HySLAC model builds on the knowledge gained from our previous literature analysis, takes into account the eight MRs discussed in Section 2, and was created using the Unified Modeling Language 2.0 (UML). Since the UML allows the modeling of class diagrams on different levels of abstraction, it can also be used to create conceptual models that focus on domain concepts instead of software entities. Figure 3 shows HySLAC as a UML class diagram mapping core elements for SLA compliance in HC architectures as classes and their relationships as associations.

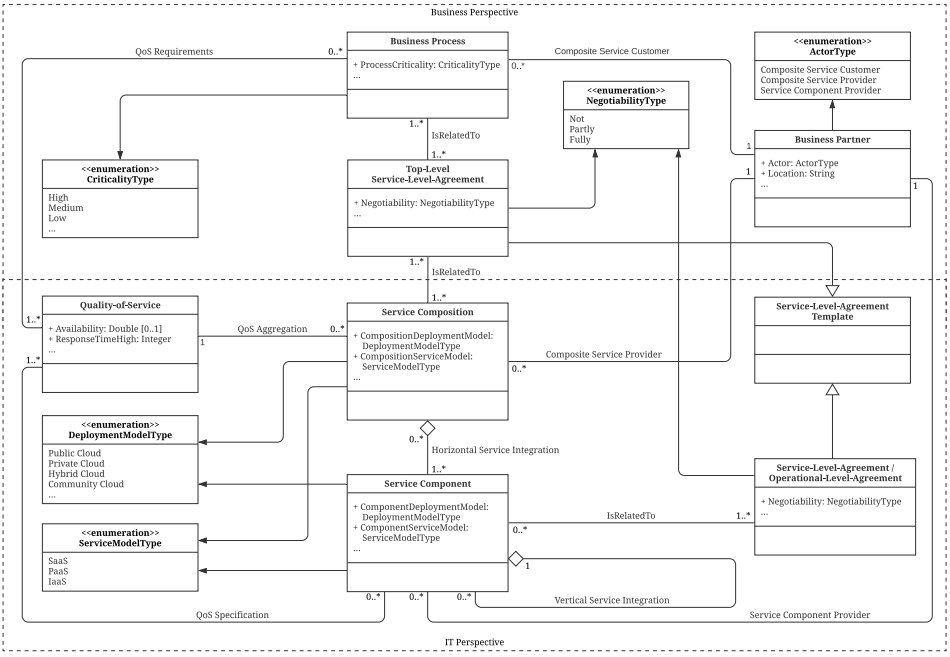


Fig. 3: The HySLAC model, represented as a UML 2.0 class diagram

The HySLAC model is divided into two thematic clusters. The first cluster represents the *Business Perspective*, which focuses both on *Business Partners* and *Top-Level-SLAs* of *Service Compositions*. This involves integrating the *Business Process* into the problem

context. In contrast, the second cluster focuses on the *IT Perspective*. This cluster reflects the responsibilities of the IT department in the context of IT service management. It aims at ensuring the contractually agreed SLAs, taking into account *Qualities of Service* and Service Compositions. The central class of the conceptual model is the Top-Level SLA, which is an integral part of the negotiations between *Composite Service Customers* and *Composite Service Providers*. This class is used to compare the different *Requirements* from a business perspective with the technical possibilities, conditions, and prices of the Service Compositions.

In the Business Perspective, business criticality is evaluated based on the Business Process to be supported, and it is represented as an enumeration called *CriticalityType* with the attributes high, medium, or low. For example, core Business Processes can be specified with high criticality where the failure of an IT service would result in economic consequences (risks). The business-relevant evaluation of criticality is related to the *Negotiability* of (Top-Level) SLAs, which we modelled as an enumeration in line with the definition of negotiation levels as discussed in the context of MR8. By modeling the Business Perspective, it can be quickly determined whether changes in Service Compositions, such as the addition of non-negotiable PuC services, would jeopardize the compliant execution of business-critical processes. Another relevant aspect is the assignment of the Service Composition Provider to a Business Partner, which allows the business management to identify existing partnerships based on the enumeration *ActorType*, as well as resulting legal risks depending on the *Location* of the provider.

In the IT Perspective, the Service Composition is the key element resulting from the composition of Service Components. It always consists of at least one Service Component and represents the relationship to the Top-Level SLA. The Service Composition can consist of horizontally and/or vertically integrated components. *Vertical Service Integration* means that one service component serves as the technical basis for another Service Component, while *Horizontal Service Integration* means that services are combined and are not technically built upon one another. Service Components and Service Compositions are specified by the enumerations *ServiceModelType* and *DeploymentModelType*, whose characteristics are currently considered according to the NIST definitions [MG11], but which can easily be extended. If a Service Component or a Service Composition is provided by the same service provider, *Operational Level Agreements* are used instead of SLAs. However, if the service providers are different, SLAs are negotiated. QoS are modeled as measurable and predictable metrics for agreed OLAs/SLAs for the respective Service Components. The Service Composition aggregates (several) QoS of Service Components to a QoS of the Top-Level SLA. SLAs can also inherit from an *SLA Template* to achieve a consistent formalization of relevant QoS in tiered categories (e.g., gold, silver, or bronze). This allows service providers to specify established sets of characteristics of common QoS for service customers in order to simplify negotiations. SLA Templates have a high practical relevance, especially for PuC services, which are static due to their non-negotiability. Using public cloud SLA Templates, even Service Composition Providers can prepare service levels for HC services.

6 Instantiation of the HySLAC Model

We instantiated the HySLAC model based on a case study with a medium-sized company (see Figure 4). To conduct the case study, we 1) presented our HySLAC model and research project to the company, 2) discussed challenges of SLA compliance in the context of HC architectures, and 3) worked through a common application context.

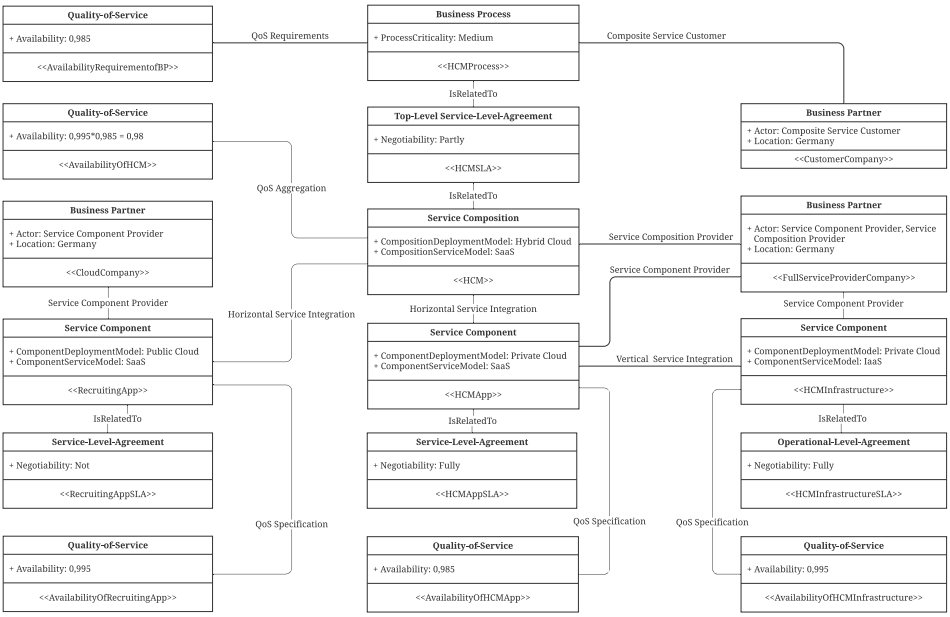


Fig. 4: Instantiation of the HySLAC model, represented as a UML 2.0 class diagram

The starting point of phase 3) was an ongoing Human Capital Management (HCM) service of the company under review which was previously provided as a PiC service for one of its customers. This service was to be extended by functionalities of a new PuC service (RecruitingApp) for the business process of recruiting. The new application was intended to replace the existing recruiting functionality within the company's IT system, while all other functionalities required for the HCM process would continue to be provided from the PiC service. Some data between the new application and the existing HCM system (e.g., applicant data in case of hiring) had to be synchronized systematically. This required the integration of the two service components into a HC service and the aggregation of the respective QoS. The original PiC service was provided through vertical integration of an underlying IaaS HCM infrastructure with a fully negotiable OLA. To illustrate our use case, service availability is shown as a QoS example in Figure 4. The instantiation shows a SLA adjustment after non-negotiable availability constraints are added to the aggregated QoS.

7 Conclusion

Cloud computing provides IT infrastructures and services via computer networks in order to achieve economic potentials for end customers and to enable companies to concentrate on core competencies. However, besides its extensive potentials, cloud computing in general, and HC computing in particular, pose challenges in the negotiation and formulation of SLAs, as well as in monitoring and complying with contractual requirements. An understanding of cloud deployment models, cloud service models, perspectives, roles, and contractual conditions (i.e., SLA and QoS) is of fundamental importance for Enterprise Cloud Architects as it provides the foundation for a successful and compliant (hybrid) cloud deployment. Therefore, this paper aimed to provide a conceptual model focusing on SLA compliance in HC architectures – the so-called HySLAC model.

Based on the derivation of eight MRs and a systematic analysis of domain knowledge according to the rigorous method of [vo09], we conceptualized the novel HySLAC model in UML 2.0, which comprises eight UML classes and five associated enumerations. In general, the model opens up practical application potentials for the analysis of service components and service compositions with regard to connected roles, as well as dependencies on SLAs, SLA templates, and QoS. The model specifically opens up practical application potentials for Enterprise Cloud Architects to analyze HC service compositions with regard to SLA hierarchies and QoS aggregations. Thus, the HySLAC model can be used as a tool for analyzing dependencies and restrictions before HC adoptions, and it opens up potentials for decision support. Scientists can specify the model for different cloud scenarios and new application contexts. Furthermore, the model provides a starting point for deriving context-specific SLA hierarchies and, based on this, opens up the potential for deriving new QoS aggregation algorithms across different hierarchical levels.

In order to adequately assess the explanatory power and scope of the model, the limitations of our investigation have to be considered. The derivation of MRs for the HySLAC model is founded purely in theory. Although the construction of a conceptual model based on domain knowledge is considered legitimate according to [MS95], we cannot fully ensure that all relevant model requirements have been considered. An additional empirical analysis, e.g., by means of a survey with experts, would address this limitation and represents a research desideratum. Moreover, our systematic literature review cannot guarantee that all relevant literature has been identified. However, the rigorous documentation of our literature review according to the method of [vo09] ensures traceability and reproducibility. Although our model was instantiated on the basis of a use case, a well-founded empirical evaluation is still pending. As a consequent next step of research, it is planned to extensively test the HySLAC model with partners from practice for further development.

References

[Al15] Aljoumah, E.; Al-Mousawi, F.; Ahmad, I.; Al-Shammri, M.; Al-Jady, Z.: SLA in Cloud Computing Architectures: A Comprehensive Study. Int. Journal of Grid and Distributed Computing 8/5, pp. 7–32, 2015.

[BB18] Bensberg, F.; Buscher, G.: Die Kunst der Systeme. Kompetenzen und Berufsbilder des Enterprise Architecture Management, INFORMATIK, pp. 43-58, 2018.

[BN13] Breiter, G.; Naik, V. K.: A Framework for Controlling and Managing Hybrid Cloud Service Integration, Int. Conf. on Cloud Engineering, pp. 217-224, 2013.

[Br09] Brandic, I.; Music, D.; Leitner, P.; Dustdar, S.: VieSLAF Framework: Enabling Adaptive and Versatile SLA-Management, GECON, pp. 60-73, 2009.

[BSC19] Brockmann, C.; Sultanow, E.; Czarnecki, C.: Is Enterprise Architecture still relevant in the Digital Age?, INFORMATIK, p. 21, 2019.

[Co09] Comuzzi, M.; Kotsokalis, C.; Rathfelder, C.; Theilmann, W.; Winkler, U.; Zacco, G.: A Framework for Multi-level SLA Management, ICSOC Service Wave, pp. 187-196, 2009.

[Co88] Cooper, H. M.: Organizing Knowledge Syntheses: A Taxonomy of Literature Reviews. Knowledge in Society 1/104, pp. 104–126, 1988.

[Go14] Goyal, S.: Public vs Private vs Hybrid vs Community - Cloud Computing. Int. Journal of Computer Network and Information Security 6/3, pp. 20–29, 2014.

[Gu11] Guo, G.; Yu, F.; Chen, Z.; Xie, D.: A Method for Semantic Web Service Selection based on QoS Ontology. J.Comput. 6/2, pp. 377–386, 2011.

[JAB12] Javadi, B.; Abawajy, J.; Buyya, R.: Failure-aware Resource Provisioning for Hybrid Cloud Infrastructure. Journal of Parallel and Distributed Computing 72/10, pp. 1318–1331, 2012.

[KGN09] Khan, K. M.; Gangavarapu; N. M.: Addressing Cloud Computing in Enterprise Architecture: Issues and Challenges. Cutter IT Journal 22/11, pp. 27–33, 2009.

[KHB02] Kneer, H.; Haeuschen, H.; Bauknecht, K.: Tradable Service Level Agreements to Manage Network Resources for Streaming Internet Services, European Conf. on Information Systems, pp. 611-624, 2002.

[KPP16] Kritikos, K.; Plexousakis, D.; Plebani, P.: Semantic SLAs for Services with Q-SLA. Procedia Computer Science 97/1, pp. 24–33, 2016.

[Li15] Li, J.; Li, Y. K.; Chen, X.; Lee, P. P.; Lou, W.: A Hybrid Cloud Approach for Secure Authorized Deduplication. IEEE Transactions on Parallel and Distributed Systems 26/5, pp. 1206–1216, 2015.

[LJ06] Levy, Y.; J. Ellis, T.: A Systems Approach to Conduct an Effective Literature Review in Support of IS Research. Informing Science 9/1, pp. 181–212, 2006.

[LMB16] Labidi, T.; Mtibaa, A.; Brabra, H.: CSLAOnto: A Comprehensive Ontological SLA Model in Cloud Computing. Journal on Data Semantics 5/3, pp. 179–193, 2016.

[MG11] Mell, P.; Grance, T.: The NIST Definition of Cloud Computing. National Institute of Standards and Technology, Gaithersburg, 2011.

[MS95] March, S. T.; Smith, G. F.: Design and Natural Science Research on Information Technology. Decision Support Systems 15/4, pp. 251–266, 1995.

[MSY16] Masuda, Y.; Shirasaka, S.; Yamamoto, S.: Integrating mobile IT/Cloud into Enterprise Architecture, Pacific Asia Conf. on Information Systems, 2016.

[NK07] Nurmela, T.; Kutvonen, L.: Service Level Agreement Management in Federated Virtual Organizations, Lecture Notes in Computer Science, pp. 62-75, 2007.

[PJW10] Paquette, S.; Jaeger, P. T.; Wilson, S. C.: Identifying the Security Risks Associated with Governmental Use of Cloud Computing. Government Information Quarterly 27/3, pp. 245–253, 2010.

[PM15] Pan, W.; Mitchell, G.: Software as a Service (SaaS) Quality Management and Service Level Agreement, INFuture2015, pp. 225-234, 2015.

[PRS09] Patel, P.; Ranabahu, A. H.; Sheth, A. P.: Service Level Agreement in Cloud Computing. Kno.e.sis Publications 1/78, pp. 1–10, 2009.

[Qi12] Qi, L.; Dou, W.; Zhang, X.; Chen, J.: A QoS-aware Composition Method Supporting Cross-platform Service Invocation in Cloud Environment. Journal of Computer and System Sciences 78/5, pp. 1316–1329, 2012.

[Su12] Suakanto, S.; Supangkat, S. H.; Suhardi; Saragih, R.: Performance Measurement of Cloud Computing Services. Int. Journal on Cloud Computing 2/2, pp. 9–20, 2012.

[Th10] Theilmann, W.; Happe, J.; Kotsokalis, C.; Edmonds, A.; Kearney, K.; Lambea, J.: A Reference Architecture for Multi-level SLA Management. Journal of Internet Engineering 2/2, 2010.

[vo09] vom Brocke, J.; Simons, A.; Niehaves, B.; Riemer, K.; Plattfaut, R.; Cleven, A.: Reconstructing the Giant: On the Importance of Rigour in Documenting the Literature Search Process, European Conf. on Information Systems, pp. 2206-2217, 2009.

[WW02] Webster, J.; Watson, R. T.: Analyzing the Past to Prepare for the Future: Writing a Literature Review. Management Information Systems Quarterly 26/2, pp. xiii–xxiii, 2002.

[Ya16] Yangui, S.; Ravindran, P.; Bibani, O.; Glitho, R. H.; Ben Hadj-Alouane, N.; Morrow, M. J.; Polakos, P. A.: A Platform as-a-Service for Hybrid Cloud/Fog Environments, 22nd IEEE LANMAN, pp. 1-7, 2016.

[YBD08] Youseff, L.; Butrico, M.; Da Silva, D.: Toward a Unified Ontology of Cloud Computing, Grid Computing Environments Workshop, pp. 1-10, 2008.

8. Workshop Umweltinformatik zwischen Nachhaltigkeit und Wandel

8. Workshop Umweltinformatik zwischen Nachhaltigkeit und Wandel (UINW 2020)

Stefan Naumann[1] Kristina Voigt[2] Eva Kern[3] Volker Wohlgemuth[4] Grit Behrens[5]

Abstract: Der Workshop „Umwelinformatik zwischen Nachhaltigkeit und Wandel“ schlägt auch in seinem bereits achten Durchlauf die Brücke zwischen Informatik, Umwelt- und Nachhaltigkeitswissenschaften und präsentiert interdisziplinäre wissenschaftliche Lösungen aus der Schnittstelle zwischen Digitalisierung und Nachhaltiger Entwicklung.

Keywords: Umweltinformatik; Nachhaltigkeit; Informationstechnik

1 Ziele und Motivation des Workshops

Wie selten zuvor wird der Themenkomplex nachhaltige Entwicklung, Klimaschutz, Digitalisierung und Künstliche Intelligenz auch in der Öffentlichkeit zusammenhängend intensiv diskutiert. Nicht zuletzt Schlagzeilen wie „Streamen ist das neue Fliegen“ verdeutlichen, dass hier Lösungen und Herausforderungen durch Informations- und Kommunikationstechnik zwei Seiten derselben Medaille sind. Digitalisierung und Nachhaltige Entwicklung sind somit Schüsselherausforderungen des 21. Jahrhunderts. Schnittpunkt der hieraus resultierenden Forschungsbedarfe ist die Umweltinformatik als ein wesentliches Forschungsfeld der Angewandten Informatik. Die Umweltinformatik besteht und arbeitet seit ca. 35 Jahren kontinuierlich an der Verknüpfung von Informatik mit aktuellen Umwelt-themen. Im Rahmen des Workshops 2020 möchten wir unter diesen Prämissen die Frage nach den Perspektiven der Umweltinformatik stellen: Wie werden diese Themen in der Forschung behandelt? Wie beeinflussen neue Techniken und Phänomene wie Internet of Things, CO2-Load, Blockchain, Citizen Science oder auch Künstliche Intelligenz und Deep Learning die Methoden und Anwendungen der Umweltinformatik? Sind neue Modelle, Vorgehensweisen, Architekturen erforderlich?

[1] Hochschule Trier, Umwelt-Campus Birkenfeld, Institut für Softwaresysteme, Postfach 1380, 55761 Birkenfeld, s.naumann@umwelt-campus.de
[2] 97, Route de Luxembourg, L–6562 Echternach, kvoigtvoigt@web.de
[3] Leuphana Universität Lüneburg, Universitätsallee. 1, 21335 Lüneburg, mail@nachhaltige-medien.de
[4] Hochschule für Technik und Wirtschaft Berlin, Wilhelminenhofstraße 75a, 12459 Berlin, volker.wohlgemuth@htw-berlin.de
[5] Fachhochschule Bielefeld, Campus Minden, Artilleriestraße 9, 32427 Minden, grit.behrens@fh-bielefeld.de

doi:10.18420/inf2020_20

2 Thematische Schwerpunkte

- Ressourcenverbräuche durch IKT
  - Herstellung, Entsorgung
  - Nutzung
  - Hardware und Software
- Nachhaltige Informatik als Lehr- und Lerngebiet
  - Hochschulen
  - Citizen Science
  - NGO, öffentliche Hand und KMU
- Anwendungen zum Nutzen und zum Schutze von Umwelt und Klima
  - Betriebliche Umweltinformationssysteme (BUIS)
  - Geographische Informationssysteme (GIS)
  - Große Datenmengen in Umweltanwendungen
  - Klimawandel und Adaptation
  - Smart Meter und Smart Grid
  - Umwelt- und Energieinformationssysteme
  - Umwelt- und Gesundheitsthemen in der Informatik
- Methoden der nachhaltigen Informatik
  - Modellierung und Simulation von großen Datensätzen
  - Umweltinformatik und Design
  - Künstliche Intelligenz und Machine Learning

3 Eingereichte Fachbeiträge

Insgesamt wurden für den Workshop 5 Fachbeiträge eingereicht, die aufgrund ihrer Qualität auch sämtlich angenommen werden konnten. Die thematische Bandbreite reicht dabei von der Modellierung von Energieinfrastrukturen über die energetische Optimierung von Raumlüftung mittels Machinellem Lernen sowie der Frage der Energieeffizienz von Streaming bis hin zur Betrieblichen Umweltinformatik aus ägyptischer Perspektive und der übergreifenden Fragestellung, ob die Umweltinformatik letztlich Geschmackssache ist.

4 Programmkommitee

- Prof. Dr. Hans-Knud Arndt, Otto-von-Guericke-Universität Magdeburg
- Prof. Dr.-Ing. Grit Behrens, Fachhochschule Bielefeld, Campus Minden
- Dr. Kristina Voigt, Luxemburg
- Dr. Eva Kern, Leuphana Universität Lüneburg
- Prof. Dr. Stefan Naumann, Hochschule Trier, Standort Umwelt-Campus Birkenfeld
- Prof. Dr. Wolf-Fritz Riekert, Hochschule der Medien Stuttgart
- Prof. Dr. Jochen Wittmann, Hochschule für Technk und Wirtschaft Berlin
- Prof. Dr. Volker Wohlgemuth, Hochschule für Technk und Wirtschaft Berlin

R. Reussner, A. Koziolek, R. Heinrich (Hrsg.): INFORMATIK 2020,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2021 225

Towards an Infrastructure for Energy Model Computation and Linkage

David Georg Reichelt[1] Stefan Kühne[2] Fabian Scheller[3] Daniel Abitz[4] Simon Johanning[5]

Abstract: Decision makers strive for optimal ways of production and usage of energy. To adjust their behavior to the future situation of markets and technology, the execution of different models predicting e.g. energy consumption, energy production, prices and consumer behavior is necessary. This execution is itself time-consuming and requires input data management. Furthermore, since different models cover different aspects of the energy domain, they need to be linked. To speed up the linkage and reduce manual errors, these linkage needs to be automated. We present *IRPsim*, an infrastructure for computation of different models and their linkage. The *IRPsim*-infrastructure enables management of model data in a structured database, parallelized model execution and automatic model linkage. Thereby, *IRPsim* allows researchers and practitioners to use energy system models for strategic business model analysis.

1 Introduction

Prediction of future developments is a key value for acting successfully in markets. This prediction is often done by models, which mirror the reality using simplified assumptions about the real world. This is particularly relevant in the energy domain, since the transition of power production from conventional power plants to renewable power plants changes market behaviors [Fa16]. Therefore, municipal energy utilities need to rearrange their portfolio. These portfolios need to be adjusted based on the adoption behavior of customers. To support decision makers of municipal energy utilities, *IRPopt* [Sc18] models the economic effects of changed portfolios and *IRPact* allows insights in the adoption behavior of the customers.

The manual execution of these models suffers from three problems: (1) Their input data, including parameters of different commercial actors such as customers or producers, engineering components such as markets or loads and component relations such as energy flow edges, are complex to handle. (2) The execution of models is too resource-intensive to be executed on a desktop PC on a regular basis. (3) While single models cover their

[1] Universität Leipzig, Universitätsrechenzentrum, Abteilung Forschung und Entwicklung, dg.reichelt@uni-leipzig.de
[2] Universität Leipzig, Universitätsrechenzentrum, Abteilung Forschung und Entwicklung, kuehne@uni-leipzig.de
[3] Technical University of Denmark (DTU), Department of Technology, Management and Economics, Energy Systems Analysis, fjosc@dtu.dk
[4] Universität Leipzig, Institut für Schwarmintelligenz und Komplexe Systeme, abitz@informatik.uni-leipzig.de
[5] Universität Leipzig, Institut für Infrastruktur & Ressourcenmanagement, johanning@wifa.uni-leipzig.de

doi:10.18420/inf2020_21

respective perspective on the domain, their linkage enables a broader view while preserving the advantage of each model [We96]. Manual organizing the linkage of models is time-consuming and error-prone.

To overcome these problems, we present *IRPsim*, a software framework capable of (1) managing input and output data for energy system models, (2) executing different energy system models and (3) linking different models. The implementation of *IRPsim* aims for usability by domain experts without modeling knowledge. In this paper, we describe the architecture of *IRPsim*. The basic *IRPsim*-infrastructure and the *IRPopt*-model, which can be plugged into *IRPsim*, are developed and in use [Kü19; Sc18; Sc20]. Currently, the *IRPact*-model and the model linkage are in an advanced development state. A screenshot of the graphical user interface is depicted in Figure 1.

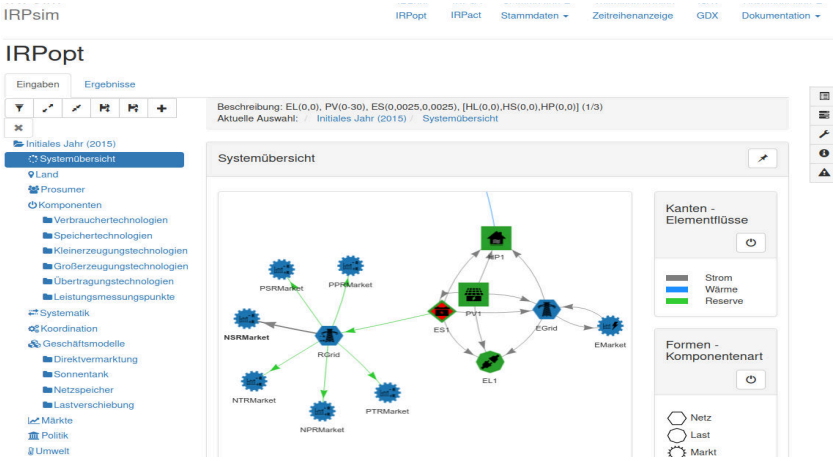


Fig. 1: Graphical User Interface of IRPsim

The remainder of this paper is organized as follows: At first, we describe models suitable for execution in *IRPsim* (Section 2). Based on the models, we describe the architecture of *IRPsim* itself (Section 3). Afterwards, an overview over related work is given (Section 4). Finally, we summarize our work and give an outlook to future work (Section 5).

2 Techno-socio-economic Perspective on Municipal Energy Systems

Decision makers in the energy domain are confronted with making informed decisions within the scope of continuously evolving systems. With the help of techno-economic optimization models, e.g. deeco [Br97], DER-CAM [St14], EnergyHub [Ge07], XEONA [MWB05] or *IRPopt* [Sc18], decision makers of municipal utilities can investigate the performance of an energy system under different circumstances from different market actors perspectives. In subsection 2.1, we give an overview about the optimization model *IRPopt*. While techno-economic modeling can capture technological interactions, it cannot

cover commercial processes that arise between multiple market participants. These can be modeled by an agent models called *IRPact*, which is described in subsection 2.2. To evaluate new business models by providing insights into the operational performance of the energy supply system and the interactions between the adoption decision of market actors, we propose a combined analysis of the techno-economic and socio-economic dynamics. This combination is described in subsection 2.3.

2.1 The Techno-economic Optimization Framework *IRPopt*

The techno-economic optimization framework *IRPopt* (Integrated Resource Planning and Optimization) [Sc18] supports decision makers of municipal energy utilities regarding future portfolio management. The mathematical optimization model allows for a policy-oriented, technology-based and actor-related assessment of varying energy system conditions in general, and innovative business models in particular. The integrated multi-modal approach is based on a novel six-layer modeling framework built on existing high-resolution modeling building blocks.

The optimization model, which has been implemented in GAMS/CPLEX (General Algebraic Modeling System[6]), allows for solving mixed-integer problems in a (quarter-)hourly resolution for perennial periods. The major objective is to maximize profits from different actor perspectives. Thereby, *IRPopt* provides a novel actor-oriented multi-level optimization framework. This is achieved by explicitly modeling municipal market actors on one layer and state-of-the-art technology processes on another layer. Resource flow interrelations and service agreements mechanism are modeled on and between the different layers. Individual participating market actors and the spatially distributed load, storage and generation technologies are modeled separately. Furthermore, multi-party cooperation is incorporated. Individual actors hold bilateral contracts with each other that handle the business transactions.

Due to the chosen approach, decision making of different modeled market actors is unbounded rational [WB09]. In addition to models covering local utilities and large independent energy producers as fully rational actors [WB09], *IRPopt* permits to determine the optimal operation dispatch and thus the optimal profitability index from different market actor perspectives. Thereby, the effects of decentralized business models, such as self consumption, regional self marketing and neighborhood energy storage systems can be modeled [Sc17].

2.2 The Socio-economic Agent-based Model *IRPact*

While techno-economic modeling can capture technological interactions, it cannot endogenize the commercial processes that arise between multiple market participants. Structural

[6] https://www.gams.com/

decisions of different market actors are often related to bounded rationality and thus are not fully rational. The adoption of technology innovation does not just depend on the qualities of the innovation. Instead, it takes place within a complex social system, in which the diffusion of the respective innovations depends on many factors and mechanisms [Sc07]. Innovations need to encompass the dynamics of the market setting by including the mental decision structures, such as personal characteristics and behavioral attitudes, as well as conscious and subconscious purchase decisions of stakeholders in general and of customers in particular.

For the representation of such socio-economic processes the approach of empirically grounded agent-based modeling turned out to be one of the most promising approaches as it allows for considering various influences on the adoption process on a micro-level [ZV19]. Additionally, a large share of available applied research already deals with environmental and energy-related innovations [SJB19].

Socio-economic modeling does not only account for the heterogeneity of bounded-rational mental behavior patterns, which are not only based on economic thinking, but also considers the social structures of market actors [Bo02]. This approach makes it possible to simulate acceptance and diffusion of innovations by various customer types and utilities considering different decision-making and network models, as well as the temporal and regional differences in the diffusion process.

The agent framework *IRPact* (Integrated Resource Planning and Interaction) allows the simulation of the mentioned diffusion processes. *IRPact* is implemented in Java using Jadex [BP12][7]. The Jadex framework allows the implementation of agents, based on the Belief-Desire-Intention model of Bratman [Br87], which is grounded on folk-psychology and permits the simulation of human reasoning and irrational decisions making. Therefore, this model is used in a wide range of social simulations [AG16] and diffusion processes in social networks [BCP18].

2.3 Integration of *IRPopt* and *IRPact* in *IRPsim*

Model	Domain	Input	Output	Time Scale	Language
IRPopt	Energy Dispatch Modeling	Techno-economic Parameters	Profitability Indices	$\geq$ 15 mins	GAMS
IRPact	Techn. Diffusion Modeling	Socio-economic Parameters	Adoption Rates	Monthly	Java

Tab. 1: Summary of Model Properties

The properties of *IRPopt* and *IRPact* are summarized in Table 1. The adoption rate of individual market actors regarding energy-related business models directly affects energy supply networks and process technologies. In contrast, the optimized operation dispatch (profitability index) of individual actors in terms of a given supply network can be considered

[7] Official website: `https://www.activecomponents.org/`

a single influencing aspect of the decision behavior of other market participants. In general, socio-economic behavior patterns of market actors have system impacts on the techno-economic business performance of the energy supply system and vice versa. Such feedback effects between decisions of market actors and the performance of a certain energy supply network can be simulated by a combination of a bounded rationality model with an unbounded rationality model as initially described by [WB07].

The multi-model *IRPsim* (Integrated Resource Planning and Simulation) [SJB18] represents such a combined approach by integrating the bounded and unbounded rationality modeling approaches *IRPact* and *IRPopt*. While the model *IRPact* (Integrated Resource Planning and Interaction) calculates the adoption rates of individual market actors, the model *IRPopt* (Integrated Resource Planning and Optimization) optimizes their profitability indices. The mutual dependencies of the coupled models result in an interactive and dynamic energy model application for multi-year business portfolio assessment.

The integration of both modeling approaches is realized by a common data basis and by linking input and output parameters. Both models consider the same market participants, but from different perspectives. While in *IRPopt*, for example, contractual relationships and optimization authorities of actors are parameterized, the underlying mental models and social relationships are relevant for *IRPact*. Where optimization results of *IRPopt* of a certain system infrastructure provide costs and revenues for each of the participating actors in terms of operational management, the simulation results of a certain social system in *IRPact* shows the adoption rate for each of the participating actor. This, in turn, affects the system infrastructure and changes the parametrization of *IRPopt*. At the same time, the reevaluated profitability of the adoption decision influences the decision making of the participating actors and thus the adoption process of *IRPact*.

3 Technical Architecture of the Simulation Platform *IRPsim*

The application of the *IRPsim* models *IRPopt* and *IRPact* implies a number of technical and non-functional requirements. The practical application within industrial usage scenarios requires systematic management of input and output data, handling of execution resources and organization of model linkage. These aspects require support by a software infrastructure. The workflow is depicted in Figure 2.

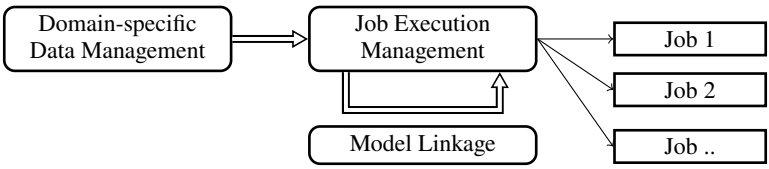


Fig. 2: Requirements Overview

In [Kü19] we describe how the essential aspects usability, adaptivity and flexibility are realized in the *IRPsim*-infrastructure based on master data management, scenario-based

configuration and a model-driven development. In the following we concentrate on aspects aiming at the integration of *IRPopt* and *IRPact*.

IRPopt and *IRPact* process hundreds of input and output parameters – scalars, arrays, sets and time series. The management and assignment of input and output data is described in subsection 3.1. Based on the input data and model specifications, the execution is managed, which is described in subsection 3.2. To ease the use of different models and their interaction, model linkage is handled, which is described in subsection 3.3.

3.1 Data Management

Input and output data collection of both models is done by researchers. They gather concrete input data from various sources, e.g. stock exchange publications, weather forecast databases and scientific literature. The future of the energy world may be shaped by different scenarios, as e.g. business may be continued as usual or the political incentives for the energy transition might be increased heavily. During the temporal scope of the model, these scenarios change, therefore the forecast scenarios of 2015 are not the same as the scenarios of 2020. Furthermore, some input parameters are functions of other parameters, e.g. the tariff for 1 kWh of a private household in year n might be the tariff of year $n-1$ plus 5%.

The *IRPsim*-infrastructure supports researchers by storage, checks for completeness and checks for correctness of data. This is done using the data model depicted in Figure 3. All input and output data are `DataSets`, i.e. instances of data defined by year and scenario. Most of them consist of `StaticData`, i.e. a concrete value or a concrete time series with values in different resolutions, e.g. quarter-hourly, hourly or weekly.

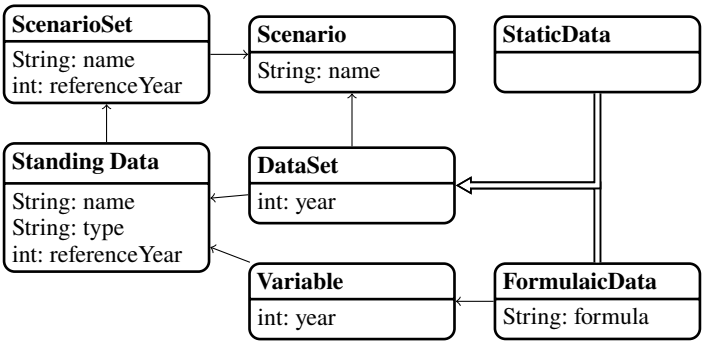


Fig. 3: Input Data Model

To assign input `DataSets` to input parameter of models, the model input parameters are enhanced by annotations which specify unit and name of each parameter. For the concrete assignment of `DataSet`-instances, they contain references to a `StandingData`-instance which contains the *IRPsim*-DSL name as `type` (e.g. *par_F_E_EGrid_energy*, an energy tariff), a human readable name (e.g. *power tariff spot market*) and a reference year, e.g. 2020. For

each `StandingData` instance, for every scenario and every concrete year, a `DataSet` instance may be present. Besides being `StaticData`, a `DataSet` might be `FormulaicData`, which contains formulas referencing `Variables`, which also are `StandingData`, e.g. the tariff might be 5% more then the data contained in the `StandingData`-instance of the last year. This data model is generic for every energy model currently used, since they all require input data in the form of time series or scalar values with the same metadata.

3.2 Execution of *IRPsim* models

The model execution consists of the parametrization with values from the common data basis and the subsequent model call. Both will be described in this subsection.

Parametrization To execute the *IRPsim* models, the input parameters need to be transformed to a format readable by the underlying execution environment. The transformation process includes the formatting of the content, e.g. to roll out a time series to a given resolution, and the technical formatting, e.g. creating a GDX database for the GAMS environment or creating a JSON file for an agent-based model in Java.

This parametrization process uses a model experiment specification as input, which is created by the user using a web frontend. The model experiment specification contains mappings from parameters and its dependencies to DataSets or manually configured data. The *IRPsim* infrastructure contains generic code which reads model experiment specifications from the frontend, queries the database for the concrete data and performs the roll out of time series. Afterwards, for every model type, the parametrization needs to be implemented separately.

Model Call Model calls are initiated as independent Java sub-processes. In the case of *IRPopt*, which is represented as a GAMS model, the associated API of the GAMS environment is used. *IRPact* is executed as a JAR, which manages its own call.

Since model call jobs are long-running and input data or models itself may contain bugs, the view of intermediary results might speed up the modeling and bug fixing efforts. Therefore, models may write intermediary results to CSV files which are continuously read by the system and imported into the database. Thereby, the user may view intermediary results and spot anomalies or unexpected behavior during the call.

Currently, the parallelization of model call jobs are limited to single-server systems. To keep the called environment stable and efficient, a maximum amount of parallel running jobs must be specified. As a rule of thumb, we usually allow one job for two CPUs and 2 GB of RAM. Since model calls are time-consuming and resource-intensive, and the capacities of single-server machines are limited, we plan to expand the possibilities for parallelization of the *IRPsim* infrastructure to cluster environments. The distribution of model calls will be realized by the job scheduling system Slurm.

3.3 Model Linkage

IRPopt and *IRPact* represent the techno-economic and socio-economic perspective on future developments of municipals energy supply systems. They are parameterized based on the same data basis. The feedback loops between both perspectives, i.e. the impacts of the techno-economic model to the socio-economic model and vice versa are handled as functional dependencies between the two models.

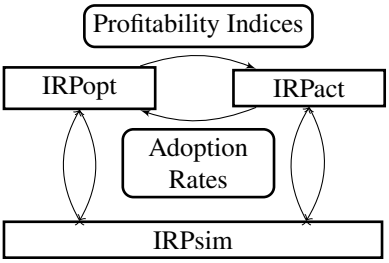


Fig. 4: Interaction of Different Models

In Figure 4 this approach is visualized based on the example of profitability indices and adoption rates as discussed in section 2: *IRPopt* predicts business-specific values, like profitability indices, and needs data of technology adoption rates. *IRPact* requires business-specific values, like profitability indices, and produces technology adoption rates. To support these inter-model dependencies, combined model execution provision was implemented. In these combined executions, the user selects a combined execution mode specifying which models to combine. Afterwards, they define which years are executed. The remaining years are interpolated. When the definition of parameters is finished, the model can be started.

Parameter Exchange Output parameters are by convention prefixed by `par_out`, e.g. `par_out_PowerMeasurement`. To exchange parameters between models, the input-parameters of one model are parsed and matched with the output parameters of another model. If their names match, e.g. if *IRPact* has an input parameter `par_PowerMeasurement` and *IRPopt* has an output parameter `par_out_PowerMeasurement`, the output values of the preceding year are transformed to an input parameter of the current year.

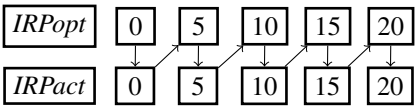


Fig. 5: Execution Dependencies

Execution Dependencies Since parameters are exchanged, the years can only be executed sequentially, i.e. since year 0 of *IRPact* relies on the profitability indices of year 0 of *IRPopt*, it needs to wait until *IRPopt* is finished, and since year 5 of *IRPopt* relies on the adoption rates of year 0 of *IRPact*, it needs to wait until *IRPact* is finished. Therefore, a parallelization of one computation is not possible. However, different scenarios may be computed in parallel.

4 Related Work

A combination of techno-economic and socio-economic modeling perspectives is necessary to provide comprehensive support to decision makers of municipal energy systems. The approach to look at different levels of abstraction, views or sectors of energy systems by

means of specialized models and to combine them to answer complex questions is discussed in existing literature. Advantages include easier application, greater flexibility and better maintainability of the specialized models [KKS20; MS00].

The combination of different models can be achieved by different approaches. Wene [We96] distinguishes between soft-linking (informal linking) and hard-linking (formal linking). With soft-linking, the processing and transmission of the information transferred between models is carried out externally, for example by the model user. The output of one model is used as input for the other. With hard linking, the degree of integration between the models is higher. The transfer of information is an essential part of the modeling itself. Böhringer and Rutherford [BR08] distinguish three categories of integration. Linking of independently developed models, coupling of models by choosing one of the models as the main one and complementing it with a representation of the other in a reduced form, and completely integrated models based on developments of solution algorithms for mixed complementarity problems. Soft-linking is typical for combining energy-sector specific and other models [KKS20].

In *IRPsim* we follow the soft-linking approach. The exchange between *IRPopt* and *IRPact* is realized by mutual parameter transfer, which happens at defined synchronization points (annual slices). By using the common simulation platform IPRsim further synergy effects are created: we use a common database for the parameterization of the models, scenario-based configuration, parallelization of execution processes and merge simulation results synchronously during execution and afterwards.

5 Summary and Outlook

We presented *IRPsim*, an infrastructure for energy model computation. *IRPsim* enables the input and output data management, the configuration of model executions, the linkage of models and the creation of output graphs. This is done using an input database, relying on MariaDB, a parametrization component, rolling out the data and producing GDX and JSON, an execution component automating the call of GAMS and Java models and an output database, managing CSV and GDX result data. Our approach is summarized in Figure 6. Thereby, *IRPsim* supports researchers and practitioners in predicting the effects of the energy transition and enables practitioners to react accordingly in their particular markets.

In the future, *IRPsim* might be extended in the following ways: (1) *Model Support Extension*: Currently, *IRPsim* supports *IRPopt* and *IRPact* and their data exchange. We plan to extend *IRPsim* to be usable for more models, e.g. a balancing energy model, a spot market price model or a political economy model. (2) *Analysis Capability Extension*: Currently, *IRPsim* allows for the creation of graphs to create insights into the model behavior. We plan to further extend the amount of available graphs. (3) *Duration Prediction*: The same model has different performance with different input data, e.g. *IRPopt* increases its computation time when more customer groups are defined. To enable reasonable model execution job

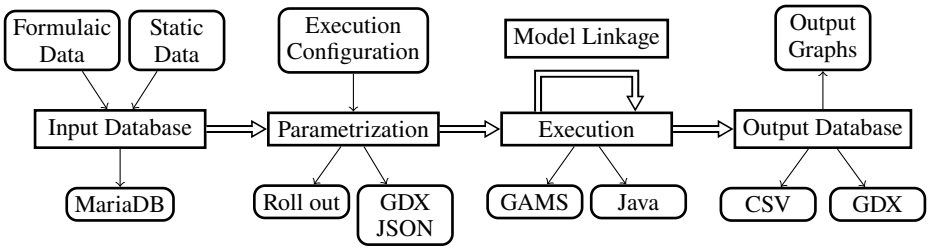


Fig. 6: Architecture Overview

prioritization, we plan to implement a duration prediction based on input data. Thereby, *IRPsim* will be even more efficient enabling research of effects of the energy transition.

Acknowledgement David Georg Reichelt, Fabian Scheller, Daniel Abitz, and Simon Johanning receive funding from the project SUSIC (Smart Utilities and Sustainable Infrastructure Change) with the project number 1722 0710. The project is financed by the Saxon State government out of the State budget approved by the Saxon State Parliament. Fabian Scheller also kindly acknowledges the funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement no. 713683 (COFUNDfellowsDTU).

References

[AG16] Adam, C.; Gaudou, B.: BDI agents in social simulations: a survey. The Knowledge Engineering Review 31/3, pp. 207–238, 2016.

[BCP18] Bulumulla, C.; Chan, J.; Padgham, L.: Enhancing diffusion models by embedding cognitive reasoning. In: 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM). IEEE, pp. 744–749, 2018.

[Bo02] Bonabeau, E.: Agent-based modeling: Methods and techniques for simulating human systems. Proceedings of the National Academy of Sciences of the United States of America 99/Suppl 3, pp. 7280–7287, 2002.

[BP12] Braubach, L.; Pokahr, A.: Developing distributed systems with active components and Jadex. Scalable Computing: Practice and Experience/, pp. 100–120, 2012.

[BR08] Böhringer, C.; Rutherford, T. F.: Combining bottom-up and top-down. Energy Economics 30/2, pp. 574–596, 2008.

[Br87] Bratman, M.: Intention, plans, and practical reason. Harvard University Press Cambridge, MA, 1987.

[Br97] Bruckner, T. J. C.: Dynamische Energie- und Emissionsoptimierung regionaler Energiesysteme, PhD thesis, Uni Würzburg, 1997.

[Fa16] Falthauser, M.: Der Deutsche Strommarkt und seine Entwicklung. Politische Studien 67/, pp. 52–61, 2016.

[Ge07] Geidl, M.: Integrated Modeling and Optimization of Multicarrier Energy Systems, PhD thesis, Eidgenössische Technische Hochschule Zürich, 2007.

[KKS20] Kiss, B.; Kácsor, E.; Szalay, Z.: Environmental assessment of future electricity mix – Linking an hourly economic model with LCA. Journal of Cleaner Production 264/, p. 121536, 2020.

[Kü19] Kühne, S.; Scheller, F.; Kondziella, H.; Reichelt, D. G.; Bruckner, T.: Decision support system for municipal energy utilities: approach, architecture, and implementation. Chemical Engineering & Technology 42/9, pp. 1914–1922, 2019.

[MS00] Messner, S.; Schrattenholzer, L.: MESSAGE–MACRO: linking an energy supply model with a macroeconomic module and solving it iteratively. Energy 25/3, pp. 267–282, 2000.

[MWB05] Morrison, R.; Wittmann, T.; Bruckner, T.: Policy-oriented energy system modeling with xeona./, 2005.

[Sc07] Schwarz, N.: Umweltinnovationen und Lebensstile: Eine raumbezogene, empirisch fundierte Multi-Agenten-Simulation, PhD thesis, 2007.

[Sc17] Scheller, F.; Reichelt, D. G.; Dienst, S.; Johanning, S.; Reichardt, S.; Bruckner, T.: Effects of implementing decentralized business models at a neighborhood energy system level: A model based cross-sectoral analysis. In: 2017 14th International Conference on the European Energy Market (EEM). Pp. 1–6, 2017.

[Sc18] Scheller, F.; Burgenmeister, B.; Kondziella, H.; Kühne, S.; Reichelt, D. G.; Bruckner, T.: Towards integrated multi-modal municipal energy systems: An actor-oriented optimization approach. Applied Energy 228/, pp. 2009–2023, 2018.

[Sc20] Scheller, F.; Burkhardt, R.; Schwarzeit, R.; McKenna, R.; Bruckner, T.: Competition between simultaneous demand-side flexibility options: the case of community electricity storage systems. Applied Energy 269/, p. 114969, 2020.

[SJB18] Scheller, F.; Johanning, S.; Bruckner, T.: IRPsim: A techno-socio-economic energy system model vision for business strategy assessment at municipal level. Research Report No.02 (2018), Leipzig University, Institute for Infrastructure and Resources Management (IIRM)/, 2018.

[SJB19] Scheller, F.; Johanning, S.; Bruckner, T.: A review of designing empirically grounded agent-based models of innovation diffusion: Development process, conceptual foundation and research agenda, tech. rep., Beiträge des Instituts für Infrastruktur und Ressourcenmanagement, 2019.

[St14] Stadler, M.; Groissböck, M.; Cardoso, G.; Marnay, C.: Optimizing Distributed Energy Resources and building retrofits with the strategic DER-CAModel. Applied Energy 132/, pp. 557–567, 2014.

[WB07] Wittmann, T.; Bruckner, T.: Agentenbasierte Modellierung urbaner Energiesysteme. Wirtschaftsinformatik 49/5, pp. 352–360, 2007.

[WB09] Wittmann, T.; Bruckner, T.: Agent-based modeling of urban energy supply systems facing climate protection constraints. In: Fifth Urban Research Symposium. 2009.

[We96] Wene, C.-O.: Energy-economy analysis: Linking the macroeconomic and systems engineering approaches. Energy 21/9, pp. 809–824, 1996.

[ZV19] Zhang, H.; Vorobeychik, Y.: Empirically grounded agent-based models of innovation diffusion: a critical review. Artificial Intelligence Review/, pp. 1–35, 2019.

Energie-Effizienz von Streaming-Plattformen und Möglichkeiten zur Verbesserung

Fabian Gaukler[1]

Abstract: Dieses Paper beschäftigt sich mit der Energie-Effizienz von Video-Streaming-Diensten. Zunächst wird die grundlegende Funktionsweise beschrieben, was die dahinterliegenden Technologien mit einbezieht. Neben der Erwähnung der bekanntesten Video-Streaming-Plattformen wird insbesondere auf Netflix als Fallbeispiel eingegangen. Zunächst werden die aktuellen Daten ermittelt und verschiedene Studien zur Verbesserung der Energie-Effizienz betrachtet. Außerdem wird analysiert, inwiefern in bestimmten Studien aus falschen Annahmen schlechte Prognosen ermittelt werden. Letztendlich werden Vorschläge präsentiert, wie die Energie-Effizienz nachhaltig und stetig weiterentwickelt werden kann. Dabei wird speziell eine Unterteilung in Unternehmen und Endverbraucher herangezogen, um die Möglichkeiten detaillierter zu spezifizieren.

Keywords: Energie-Effizienz; Netflix; Datencenter; Video-Streaming; Ökologischer Fußabdruck

1 Einleitung

Die Nutzung von Streaming-Plattformen für Filme ist heutzutage nicht mehr aus dem Alltag vieler Menschen wegzudenken. Das klassische Fernsehen verliert mit jeder neuen Generation zunehmend an Bedeutung und wird bei weitem nicht mehr so häufig wie früher konsumiert. Dabei nehmen sowohl der Datenverbrauch als auch der Stromverbrauch stetig zu, da immer höher auflösende Videos möglich sind und dementsprechend für den Endverbraucher produziert werden. So war es vor ein paar Jahren noch undenkbar, Filme in 4K-Auflösung zu streamen, weil die Bandbreite einfach nicht ausreichend war. Doch diese wächst parallel zur Datenmenge mit, weil beide in Wechselwirkung miteinander stehen und eine Steigerung des einen Teils eine Steigerung des anderen Teils erfordert. Auf diese Art und Weise beschleunigt sich die Entwicklung der Technologie selbst und wächst stetig schneller.

Doch in den letzten Jahren spielt die Nachhaltigkeit eine immer größere Rolle, was sich ebenfalls im Bereich des Video-Streamings zeigt. Die Plattformen müssen geschickte Methoden anwenden, um eine ausreichend hohe Qualität zu bieten und gleichzeitig die Waage zur Reduzierung von Datenmengen und Stromverbrauch zu halten. Dabei kommt es etwa darauf an, welches Endgerät der Nutzer verwendet, welche Codecs verwendet werden, ob automatisch kleine Regulierungen laufen und ob Algorithmen flexibel auf alle Anforderungen zum optimalen und nachhaltigen Streaming-Erlebnis Acht geben.

[1] Hochschule Trier, Umwelt-Campus Birkenfeld, Campusallee, 55768 Neubrücke, Deutschland, s15cb3@umwelt-campus.de

doi:10.18420/inf2020_22

2 Verwandte Arbeiten

Im Juni 2019 veröffentlichte die französische gemeinnützige Organisation „The Shift Project“ einen Bericht über den nicht-nachhaltigen und wachsenden Einfluss von Online- Videos. Dort ist zu lesen, dass Online-Streaming mehr als 300 Mio. Tonnen CO2 jährlich erzeugt. Davon sind 34% des Verbrauches VoD-Diensten (Video on Demand) zuzuschreiben, zu denen Anbieter wie Netflix, Hulu oder Amazon Prime zählen. Davon abzugrenzen sind andere Videoplattformen wie YouTube, Dailymotion oder Vimeo. Der Bericht basiert zu einem großen Teil auf dem Paper „On Global Electricity Usage of Communication Technology: Trends to 2030“ aus 2015 [AE15].

3 Funktionsweise von Video-Streaming

Wie viele andere Webseiten benutzt Netflix ein benutzerdefiniertes Content Delivery Network (CDN), um eine zuverlässige Bereitstellung ihres Video-Services zu gewährleisten. Ein CDN sorgt dafür, dass der Datendurchsatz auch bei großer Auslastung noch funktioniert, indem genug Speicher- und Auslieferungskapazitäten zur Verfügung gestellt werden. CDNs bieten den Streaming-Service Knotenpunkte in einem verteilten Array von Datencentern, welche die Videos lokal für nahegelegene Nutzer cachen und streamen. Die Server im CDN sind miteinander vernetzt, um Daten zu transportieren und auf verschiedenen Systemen bereitzustellen. Dazu erstellen CDNs einen Pfad, der auf die Dauer der Übertragung reduziert ist und für schnellere Übertragung sorgt. Zusammengefasst sorgen CDNs für eine gute Möglichkeit zur Skalierung, für Zuverlässigkeit bei großen Auslastung von vielen Nutzern, für gute Uptime, für Sicherheit vor Angriffen wie „Distributed Denial of Service“ (DDoS) und für Rahmenbedingungen wie SSL-Zertifikate, Geschwindigkeit und generelle Qualität der Verbindung.

Genauer gesagt, benutzt Netflix Amazon Web Services (AWS) als Server-Hosting und Cloud-Service, um die großen Mengen an Video-Daten zu cachen und die Datenmenge, die von Cloud Hosting Services gebraucht wird, je nach Auslastung zu beschränken. Dadurch sind bessere Konnektivität und hochwertigere Streams möglich.

Eine weitere wichtige Komponente von Video-Streaming ist der Over-The-Top Content (OTT), der in diesem Fall die Übermittlung von Videoinhalten über Internetzugänge beschreibt, ohne dass ein Internet-Service-Provider (ISP) mit eingebunden ist. Weitere Informationen zu OTT und den Regulierungen können unter [Ta15] nachgelesen werden. Das Verfahren läuft laut [As19] in den vier Schritten „Ingestion“, „Transcoding“, „Management“ und „Delivery und Playback“ ab. In der Ingestion-Phase nutzt das Video- Streaming entweder aufgezeichneten Content oder Live-Content, der direkt vom gefilmten Standort aus eingespeist wird. Mit Transcoding bezeichnet man das Konvertieren von einem Dateiformat in ein anderes. Sowohl zur Ermöglichung von Cross-Browser- und Cross-Plattform-Einsatz als auch zur optimierten Darstellung von Videos ist diese Phase ein wichtiger Schritt. In der Management-Phase findet die Content Production statt, gefolgt von der Verwaltung des

Digital Rights Managements (DRM) und der dynamischen Werbe-Einspeisung. Im letzten Schritt, Delivery und Playback, wird der Stream durch Transcodierung zu einem RAW-Video decodiert wird, dann geändert und wieder codiert. Das hat den Zweck, dass somit basierend auf Streams mit einer Auflösung von 1080p alternative Streams mit einer Auflösung von 480p oder 720p erstellt werden können. Konsumenten können so auf unterschiedlichen Endgeräten erreicht werden.

Eine Besonderheit von Netflix besteht in der Nutzung von Chaos Engineering [As19]. Damit wird ein Ansatz beschrieben, die Elastizität von moderner und komplexer Technologie-Architektur zu verbessern. Als Testverfahren werden mehrere Chaos- Experimente innerhalb eines geschlossenen Systems durchgeführt, um mögliche kleinere Probleme zu lösen und zu finden, bevor sie sich zu größeren Problemen entwickeln. Das hat den Vorteil, dass wichtige Dienste nicht irgendwann durch unwahrscheinliche, aber trotzdem mögliche, Randbedingungen in der Produktionsumgebung offline gehen.

Im modernen Web Streaming werden Video-Dateien üblicherweise in Segmente von zwei bis zehn Sekunden unterteilt, die dann per HTTP-Request angefragt und übertragen werden. Alle Requests der Folge-Segmente laufen automatisch ohne Interaktion zwischen Nutzer und System sequentiell weiter, wie es in [Su16] beschrieben wird. Während des Abspielens wird die Bandbreite des Nutzers bestimmt und dynamische Modifikationen angewandt, um Buffering zu reduzieren und für qualitativ angemessene Wiedergabe zu sorgen. Die Algorithmen dazu werden stetig weiterentwickelt und passen sich an moderne Gegebenheiten an, wie etwa die Diskrepanz zwischen Gebieten mit guter und schlechter Datenverbindung während einer Zugfahrt.

Ein bekanntes Beispiel sind NGINX Microservices [Ng20], welche die Workloads besser verteilen und für eine Optimierung der Effizienz sorgen. Microservices sind ein Ansatz in der Software-Architektur, die eine große, komplexe Anwendung aus vielen kleineren Komponenten bauen, die selbst jeweils nur eine Funktion ausführen. Die Kommunikation zwischen den unabhängigen Microservices läuft meist über Web APIs [Ng20].

4 Energie-Effizienz aktueller Video-Streaming-Plattformen

4.1 Aktuelle Video-Streaming-Plattformen

Heutzutage existiert eine große Bandbreite an Streaming-Plattformen mit dennoch unterschiedlichen Angeboten. Netflix, Hulu, Amazon Prime oder Disney Plus zählen dabei zu den größten und meistgenutzten Anbietern. In diesem Konkurrenzkampf müssen die Plattformen ihr jeweiliges Alleinstellungsmerkmal wahren, was die eigene Produktionsfirma wie Netflix Productions mit den Netflix Originals sein kann oder das stetig populärer werdende Apple TV+ mit bereits mehr als 33 Mio. Nutzern [Sp20]. Als deutscher Anbieter ist hier maxdome zu erwähnen, aber aufgrund der fehlenden verwertbaren Informationen zur Energie-Effizienz wird nicht weiter darauf eingegangen.

Neben diesen Aspekten ist es von ebenso großer Bedeutung, auf dem aktuellen Stand der Technik zu bleiben oder als Vorreiter zu gelten. Als anfänglicher Vergleichspunkt dient hierzu die Uptime der Plattformen, der unter [Up20] eingesehen werden kann. Im Betrachtungszeitraum vom 1. Januar 2020 bis zum 22. Juni 2020 ergab sich folgende Auswertung: Hulu hatte eine Downtime von 18 Minuten und Amazon Prime Video waren es 3 Minuten. Demgegenüber ist bei Netflix eine Uptime von 100% zu verzeichnen. Die Optimierung ist eng verknüpft mit der Energie-Effizienz und trägt einen großen Teil zur User Experience und flüssigen Interaktion mit dem User Interface bei.

4.2 Detaillierte Betrachtung des Energieverbrauchs und der Energie-Effizienz

Die Effizienz von Datencentern verdoppelt sich nach Koomey's Law fast alle zwei Jahre. Das Gesetz besagt, dass sich nach dem Jahr 2000 alle 2,6 Jahre die Leistungsaufnahme um den Faktor 2 vergrößert. Die längere Dauer im neuen Jahrhundert ist davon abhängig, wie stark Moore's Law sich entwickelt [KN15]. Letzteres beschäftigt sich mit der Möglichkeit, immer kleinere Transistoren zu bauen, und zwar immer effektiver und innerhalb einer bestimmten Zeitspanne.

In den letzten Jahren wurde von Streaming-Plattform vermehrt auf Hyperscale-Datencenter als Neuerung gesetzt. Das sind Big Data-Netzwerke, die durch Cloud- Computing entstehen und die sowohl hohe Zugriffszahlen als auch fluktuierende Nutzung ausgleichen können. Die Schwierigkeit bei der detaillierten Betrachtung des Energieverbrauchs und der Energie-Effizienz liegt vor allem in der Komplexität der direkten und indirekten Einflüsse, wodurch die Auswirkungen auf die Umwelt nicht leicht zu quantifizieren sind.

Eine interessante Beobachtung über die letzten Jahre hat gezeigt, dass die Intensität der Elektrizität logarithmisch fällt: Im Jahr 2000 lag der Verbrauch noch bei 6,8 kWh pro gestreamtem Gigabyte, wogegen es 2015 nur noch 0,06 kWh waren. Die damalige Prognose für 2020 lag bei 0,015 kWh, was am Ende des Jahres verifiziert oder falsifiziert werden kann. Dieser Wert ist der Durchschnitt aller Daten-getriebenen Anwendungen inklusive Video-Streaming, also er beschreibt nicht exakt deren Verbrauch. Aus diesem Grund wird das im nächsten Kapitel anhand von Netflix genauer betrachtet. Weitere Details sind unter [As18] zu finden.

Um den Verbrauch trotzdem ein wenig in Relation zu setzen, hat Andrew Sauber, Engineer bei Cloud Platform-Provider Linode herausgearbeitet, dass jedes CDN einen Link mit 40 Gigabit pro Sekunde zum Internet aufrechterhält. Jede Verbindung zu dem Server verbraucht eine Bandbreite von 7400 Kilobit pro Sekunde. Der durchschnittliche Energieverbrauch von jedem Server beträgt 321 W und bei voller Auslastung wären laut seiner Aussage 5333 simultane Streams möglich. Ein Server verbraucht 160,5 Wh, um einen laufenden Stream eine halbe Stunde an die mehr als 5000 Nutzer zu senden. Insgesamt ergibt sich dadurch ein Verbrauch von etwa 0,0301 Wh pro Nutzer innerhalb einer halben Stunde, in der er ein Video auf einer Plattform mit der Architektur wie Netflix streamt [Fu20].

Die durchschnittliche Lebensdauer von Endgeräten spielt eine ebenso wichtige Rolle in der Betrachtung des Energieverbrauchs, da sie zum Lebenszyklus dazuzählt. Diese Thematik wird in [Er16] genauer analysiert. Folgerungen daraus legen nahe, dass Geräte mit sehr langer Lebensdauer ressourcenschonender sind und sorgen dafür, dass nicht ständig neue Geräte gekauft werden, für deren Herstellung ein großer Energieaufwand betrieben werden muss. Geräte mit kurzer Lebensdauer haben demgegenüber den entscheidenden V orteil, dass ständig neue Geräte erworben werden müssen, die möglicherweise energie-effizienter sind als die Geräte vor ihrer Zeit. Dadurch wird die Energie-Effizienz auf lange Sicht verbessert. Eine bewusst implementierte kurze Lebensdauer von Geräten durch den Hersteller ist natürlich ein Ärgernis für den Kunden, aber auf der anderen Seite eine Art natürliche Auslese von schlechtem Energieverbrauch.

4.3 Fallbeispiel Netflix

Es gilt der Grundsatz, dass in jedem Schritt von Video-Streaming Energie benötigt wird. In der heutigen Zeit wird ein großer Teil von sauberen Energie-Ressourcen bezogen, aber es gibt immer noch einen bemerkenswerten Teil, der aus fossilen Brennstoffen auf Kohlenstoffbasis erzeugt wird [Da20]. Aufgrund der Tatsache, dass der letzte Teil zur Erderwärmung beiträgt, geht der Trend bei größeren Unternehmen immer mehr in Richtung „Grüne Energie“.

Internet Server verbrauchen Strom, Smartphones verbrauchen Strom, Netzwerke, die Dienste anbieten, verbrauchen ebenfalls Strom. Eine stetig steigende Nachfrage nach Video-Content bedeutet, dass stetig mehr Energie benötigt wird, um Infos zu speichern und zu verteilen. Dazu sind große Datencenter notwendig, die den Internetverkehr speichern, verarbeiten und verteilen. Zum kontinuierlichen Betrieb und zur zuverlässigen Funktionsweise benötigen diese groß angelegte Kühlungssysteme, welche ihren Teil zum Energieverbrauch beitragen. Diese Datencenter machen 1% des aktuellen jährlichen Energieverbrauchs weltweit aus und sind für 0,3% der CO2-Emissionen verantwortlich [DAD19]. Eine groß angelegte Studie, die sich mit diesem Thema beschäftigt, ist das EURECA Project: Darin wurde ermittelt, dass EU-Datencenter im Jahr 2017 eine Steigerung des Energieverbrauchs von 25% im Gegensatz zu 2014 verzeichnen. Das ist mit der steigenden Anzahl an Datencentern resultierend aus der wachsenden Nachfrage zu erklären.

Die Anzahl der Netflix-Abonnements ist 2019 um 20% auf 167 Mio. gestiegen, während der Energieverbrauch relativ gesehen um 84% gewachsen ist. Trotz diesen Zahlen bemüht sich Netflix stark, seine große Reichweite für Aufmerksamkeit zu Nachhaltigkeit nutzen, z.B. mit Content wie „Our Planet“ mit David Attenborough, was vom „World Wildlife Fund“ unterstützt wird. Diese und weiter Informationen sind unter [Ne20a] zu finden.

Das bereits erwähnte „Shift Project“ stellte die Behauptung auf, dass das Streamen von 30 Minuten Netflix 1,6kg CO2 verbraucht (also 3,2 kg pro Stunde), was einer Autofahrt von über 6 km entspricht. Im Folgenden sollen diese und weitere Behauptungen relativiert und korrigiert werden, wie es in [Ka20a] und [Ka20b] erfolgte:

Zu Beginn wird eine falsche Annahme der Datenrate genutzt: Es wird von einer Datenrate von 24 Megabit pro Sekunde (also 10,8 Gigabyte pro Stunde) ausgegangen, obwohl der Netflix-Durchschnitt 2019 weltweit bei 4,1 Megabit pro Sekunde (also 1,9 Gigabyte pro Stunde) liegt. Bei diesem Wert werden lediglich die Mobilfunknetze nicht mit einberechnet. Die Transfer-Rate von Video-Streaming in HD-Qualität beträgt in etwa 3 Gigabyte pro Stunde, die Transfer-Rate von Video-Streaming in UHD- bzw. 4K-Qualität dagegen 7 Gigabyte pro Stunde. Bei SD-Qualität sind es lediglich 0,7 Gigabyte pro Stunde. Allein aus diesen Zahlen ist ein Trend ersichtlich, dass selbst mit einer Einbeziehung von mobilen Endgeräten und deren niedrigerer Anzeigequalität niemals ein Wert von 24 Megabit pro Sekunde erreicht werden kann [Ne20b]. Die Vermutung liegt nahe, dass ein simpler Fehler bei der Unterscheidung zwischen Bit und Byte unterlaufen ist und dieser initiale Wert um den Faktor 8 verfälscht ist. Dieses anfängliche Problem sorgt dafür, dass alles, was darauf aufbaut, den Fehler multipliziert und alle weiteren Werte eine nicht korrekte Größenordnung aufweisen. Weiterhin wurde eine falsche Annahme über den Energieverbrauch von Datencentern und CDNs getätigt. Diese überstieg die realen Werte um das sieben- bis achtzehn-fache, was sich durch [Ne20a] und [DAD19] belegen lässt. Genauso wurde der Energieverbrauch von Netzwerken zur Datenübertragung um das sechs- bis siebzehn-fache überschätzt: Dabei war der angenommene Wert 0,9 kWh pro GB bei mobilem Streaming statt der eigentlichen 0,1 bis 0,2 kWh pro GB bei der Nutzung von 4G im Jahr 2019.

Ein weiterer bedeutender Faktor ist die Verteilung der Nutzung der Endgeräte, bei welcher das „Shift Project“ wieder zu optimistisch geschätzt hat. Laut ihnen streamen 50% der Nutzer von Netflix auf Smartphones und 50% auf Laptops. Vermutlich wurde diese Annahme zur Vereinfachung der Daten gewählt oder um die spätere Berechnung zu vereinfachen. In manchen Fällen kann das gewissen Vorteile bringen, aber gerade in diesem Beispiel ist das Resultat von zu vielen Faktoren abhängig, als dass es hilfreich wäre. Tatsächlich verteilt sich die Nutzung von Netflix folgendermaßen: 70% der Nutzer schauen über den Fernseher, 15% auf dem Laptop, 10% auf Tablet und 5% auf Smartphone [Ka18].

Generell ist Energieverbrauch durch Streaming stark abhängig vom benutzten Gerät, Netzwerkverbindung und Auflösung und gerade deshalb ist es umso wichtiger, dass man diese Variablen so exakt wie möglich bestimmt oder alternativ bewusst eine Wertspanne angibt, die einen gewissen Spielraum für Variationen ermöglicht. In der folgenden Grafik nach [St19] kann diese Varianz gut nachvollzogen werden:

Eine weitere Schwierigkeit liegt darin, dass sich die Werte selbst von Jahr zu Jahr stetig ändern und Studien aus vorherigen Jahren somit ständig angepasst werden müssen. Um auf den in diesem Zusammenhang wichtigen CO2-Fußabdruck (Carbon Footprint) zu sprechen zu kommen, muss man sich mit der Frage beschäftigen, welche Schritte von Streaming welchen Energieverbrauch verursachen. Dieser ist abhängig davon, wie weit entwickelt die jeweilige Generation der Geräteeinheit ist und wie viele CO2-Emissionen von dieser in ihrem Lebenszyklus erzeugt werden. Doch ein weiterer wichtiger Faktor, der bisher noch nicht angesprochen wurde, ist die regionale Komponente. Es ist sehr unterschiedlich, auf welche Art und Weise Energien in welchem Land erzeugt, importiert und genutzt werden.

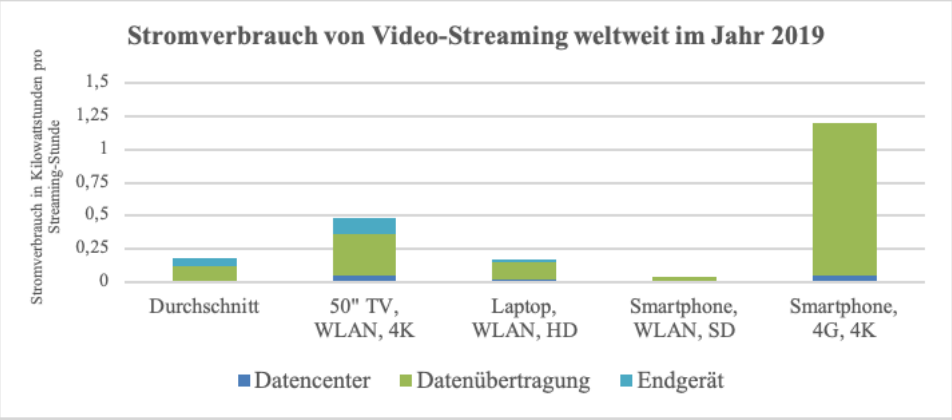


Abb. 1: Stromverbrauch von Video-Streaming weltweit im Jahr 2019

Jedes Land hat unterschiedliche Konzepte zu erneuerbaren Energien und der zukünftigen Orientierung. Aber letztendlich sind alle Werte, egal wie schlecht man sie schätzt, immer noch in einem realistischeren Rahmen als jene, die im „Shift Project“ präsentiert werden.

4.4 Streaming auf mobilen Geräten

Bereits zuvor wurde die Besonderheit von Video-Streaming auf Smartphones über mobile Netze erwähnt, worauf in diesem Kapitel genauer eingegangen wird. Video-Streaming ist der häufigste Anwendungsbereich für Smartphones und Tablets. Voraussichtlich werden Video-Streams 75% aller mobilen Daten bis 2021 ausmachen [Zh17].

Im Verlauf der letzten Jahre wurden Smartphones immer effizienter. Größere Bildschirme und höhere Auflösungen erforderten Akkus in annähernd gleichbleibender Größe mit besserer Energie-Effizienz. So wurde vermehrt auf LCD- statt CRT-Bildschirm gesetzt und das Smartphone häufiger statt Tablet oder PC genutzt. Um das Video-Erlebnis zu verbessern, was besonders bei nicht-konstanter mobiler Internetverbindung eine Rolle spielt, werden verschiedene Verfahren wie algorithmisches Prefetching eingesetzt, um ausreichende Performance-Verbesserungen bieten. Dadurch wird der Akkuverbrauch reduziert und die Kosten der Video-Bereitstellung auf Nutzer- und Anbieter-Seite gesenkt.

Generell ist der Akkuverbrauch mittels Prefetching über WLAN geringer als der während der Nutzung von mobilen Daten. Eine interessante Beobachtung ist, dass weniger Akkuverbrauch durch Prefetching nur dann erzielt werden kann, wenn das Gerät gerade aufgeladen wird. Außerdem kann durch algorithmisches Prefetching die User Experience verbessert werden, weil es seltener zu Ladezeiten durch Buffering kommt [GPN13].

Dabei ist zu beachten, dass zwei entgegenwirkende Quellen von Energieverbrauch bei Smartphones existieren: Zum einen ist das exzessives, aggressives Prefetching von Video-Content, den der Nutzer nicht schauen wird, wenn er die Session verlässt und zum anderen ist das eine exzessive Menge von Energie, die verbraucht wird, um die drahtlose Verbindung noch für eine kurze Zeitspanne aufrechtzuerhalten, nachdem ein Content- Block erhalten wurde. Der Grund für den hohen Energieverbrauch des zweiten Teils liegt darin, dass der Block zu klein ist und zu viele Blöcke ständig nachgeladen werden müssen. Eine Lösungsidee ist ein Algorithmus, der auf crowd-sourced Video-Nutzungs-Statistiken basiert, worauf dieses Paper aufgrund der Komplexität nicht genauer eingeht, was aber unter [HSN13] nachgelesen werden kann.

Eine gängige Technik zur Einsparung von Energie in integrierten Schaltkreisen nennt man „Race to sleep“. Hierbei nutzt der Chip die höchste Taktfrequenz, um den aktuellen Workload zu verarbeiten und schaltet dann wieder in den „Schlafmodus“ oder auf die niedrigste Frequenz um. Dieser Status des Schlafes kann durch „Batching of Frames“ und „Frequency Boosting“ verlängert werden, was in der Einsparung von Energie bei gleichzeitiger Vermeidung von Einbußen in der Framerate resultiert. Dabei wird das sogenannte Content Caching angewandt, welches die inhaltliche Ähnlichkeit von kleineren Video-Blöcken nutzt, um durch geschickte Cache-Organisation für eine Entlastung des Speichers zu sorgen. Durch Display Caching kann eine noch bessere Effizienz erzeugt werden, weil es die Ähnlichkeit der aktuell dargestellten Frames im Display Controller ausnutzt [Zh17].

5 Möglichkeiten zur Verbesserung

Im Verlauf dieses Papers wurden bereits ein paar Möglichkeiten zur Verbesserung der Energie-Effizienz kurz vorgestellt, vor allem für mobile Endgeräte. Dieses Kapitel konzentriert sich dagegen auf die Aufteilung in die Personengruppen, also den Nutzer und die Unternehmen, die Video-Streaming bereitstellen.

5.1 Nutzer

Obwohl der Nutzer in seinen Möglichkeiten stark eingeschränkt ist, hat er die Wahl des Endgerätes. Es ist bei Weitem energie-effizienter, ein Handy statt einem Computer, TV oder Laptop zum Streamen von Videos zu nutzen. Dabei wird davon ausgegangen, dass nur eine Person das Endgerät nutzt. Wenn man aber bedenkt, dass die durchschnittliche Anzahl von Nutzern vor einem TV größer ist als jene von Nutzern von einem Handy, würde sich der Stromverbrauch pro Nutzer geringer gestalten. Man sollte WLAN im Vergleich zu mobilen Netzen wie 3G/4G bevorzugen, weil damit eine stabilere Verbindung garantiert wird und nicht so viel Energie für die sich ständig ändernden Bedingungen und Verbindungsstärken verwendet werden muss. Eine niedrigere Auflösung fördert die Einsparung von Energie, da

nicht so viele Daten übertragen werden müssen. Auf den eigenen Geräten können die Nutzer auch verschiedene stromsparende Maßnahmen einhalten, wie etwa automatische Helligkeit oder automatische Stromsparmodi, die generell Helligkeit herunterregeln. Außerdem sollte der Benutzer eines Mobilgerätes darauf achten, dass er eine Balance bei der Erneuerung eines Gerätes findet. Man sollte das Gerät möglichst lange behalten und erst nach ein paar Jahren ein, effizienteres Gerät kaufen, sodass man nicht ein altes energie-ineffizientes Gerät zu lange unnötigerweise nutzt. Das ist umso wichtiger, wenn man bedenkt, dass die Produktion von Smartphones etwa zwei Drittel der CO2-Emission im Lebenszyklus eines Gerätes ausmacht. Je weniger Elektroschrott es gibt, desto besser, weil dieser mit einigen weiteren Energieverbräuchen wie Transport, Lagerung und Entsorgung gekoppelt ist.

5.2 Unternhmen

Datencenter sollten in naher Zukunft mit grüner Energie betrieben werden. Ebenso optimal wäre es, vermehrt Datencenter in kälteren Regionen oder unter der Erde aufzubauen, weil dadurch viel weniger Energie für die großflächige Kühlung benötigt wird. Für den Bau von neuen Datencentern sollte auch beachtet werden, größere und effizientere, statt kleinere zu bauen. Es sollte aber vermieden werden, neue Strukturen zu schaffen, wenn man auf bestehende zurückgreifen kann. So gibt es etwa Ansätze, Rechenzentren in Windkraftanlagen unterzubringen [RO18].

Als gutes Beispiel werden die globalen Datencenter von Apple schon zu 100% von erneuerbaren Energien betrieben. Facebook folgt demselben Ansatz mit ihrem neuen Datencenter in Singapur. Weiterhin gleichen viele Firmen den Verbrauch von nicht- erneuerbaren Energien durch Unterstützung von Projekten mit erneuerbaren Energien aus. So wird selbst den Firmen, die noch keine Möglichkeit zur Nutzung von grüner Energie haben, eine temporäre Möglichkeit geboten, sich nachhaltig und energie-effizient weiterzuentwickeln. Diese Unternehmen sollen als gute Beispiele weiter vorangehen, neue Technologien nutzen und in erneuerbare Energien investieren. Natürlich können kleine bis mittelständische Unternehmen nicht auf dieselbe Art agieren, weil sie nicht die Mittel und das Einkommen haben, möglichst schnell dafür zu sorgen. Hier wäre eine Subventionierung durch die jeweilige Regierung oder durch eine europäische Instanz förderlich.

Dennoch sollte weiterhin die Effizienz von Servern selbst verbessert und die Infrastruktur von Datencentern regelmäßig gewartet werden. Die steigende Popularität und Einsatzmöglichkeit von Künstlicher Intelligenz und Blockchain bietet noch viel Potenzial zur Optimierung der Energie-Effizienz in allen Bereichen (auf Server-Ebene, auf Endgeräten, für CDNs usw.). Beispielsweise können so laut [OS19] in Echtzeit Stromversorgung und Strombedarf koordiniert werden sowie die Energie-Effizienz in Echtzeit betrachtet und kontrolliert werden. Durch diese beiden Möglichkeiten können neben diesen Eigenschaften auch Netzwerkanomalien frühzeitig erkannt werden. Nachhaltiges Design und Coding ist schon länger von besonderer Bedeutung und gewisse Patterns wie Composites, Observers, Factory Methods oder Singletons sollten weiterhin eingesetzt und eventuell um neue Standards

mit neuen Technologien erweitert werden. Alle paar Jahre wird die Videokompression verbessert oder neue Codecs etabliert, wie es im Moment mit dem x265 HEVC Encoder und dem H.265 Video Codec der Fall ist.

Eine sinnvolle Idee auf YouTube könnte sein, dass bestimmte, als Musikvideos gekennzeichnete, Videos nur als Audio abgespielt werden, sobald sie sich im Hintergrund befinden. Dadurch muss immer nur der aktuelle Timestamp gemerkt werden, der dafür sorgt, dass das Video weiter abgespielt wird, sobald es wieder in den Vordergrund kommt [PSS19].

6 Fazit und Ausblick

Die International Energy Agency zeigt in ihrem neuen Bericht, dass sich trotz des wachsenden Workloads von Datencentern (2020 wird er sich verdreifachen) die Menge der verbrauchten Energie nur um 3% erhöhen wird [DAD19]. Kleinere Unternehmen müssen detaillierter zum Energieverbrauch untersucht werden, weil sich die vielen kleinen Einflüsse aufaddieren und zu einem großen Faktor werden. Noch ist es unbekannt, wie die gesamte Palette der Auswirkungen auf die Umwelt in den nächsten Jahren und Jahrzehnten unter der Verwendung von Künstlicher Intelligenz und Blockchain aussieht, aber die Vorhersagen sind bisher noch optimistisch. Einen Effekt, den es zu beachten gilt, ist der Rebound-Effekt, der beschreibt, dass das Einsparpotenzial von Effizienzsteigerungen nicht vollständig erfüllt wird. Dieses Phänomen wird weiter durch exponentielles Wachstum begünstigt, was bedeutet, dass es gerade in den letzten Jahren immer häufiger auftritt. Es muss stetig in Forschung und Entwicklung investiert werden, sodass die Energie-Effizienz in Zukunft mit dem Datenverbrauch mithalten kann und die Unternehmen sich immer an der Spitze der Forschung bewegen und diese effizient nutzen können. Künstliche Intelligenz könnte gegen die Erderwärmung und den Klimawandel helfen, aber ohne einheitliche Politik besteht die Gefahr, dass sie für Erdölerzeugung und Kohlekraftwerke zweckentfremdet wird.

Ein Appell an die Nutzer von Streaming-Plattformen besteht darin, dass man sich nicht auf Fehlinformationen durch Medien verlassen soll, wie es etwa bei der Berichterstattung durch verschiedene Medien zum „Shift Project“ der Fall war, sondern gründliche Analysen durchführen und sich flächendeckend informieren soll. Der Einsatz von Corporate Leadership und einheitlicher Politik, zumindest EU-weit würden schon einen positiven Einfluss darauf haben. Der Optimalfall wäre eine zweigleisige Herangehensweise, die sowohl den Energieverbrauch minimiert als auch den Ausbau der Grünen Energie fördert.

Der letzte und interessanteste Punkt ist die große Schwankung der verschiedenen Studien zur Energie-Effizienz. Durch die Vielzahl an Variablen und deren stetige Änderung ist es ungemein schwierig, zuverlässige Aussagen zu treffen. Es ist sehr leicht, einen Teil außer Acht zu lassen, was dann aber dafür sorgt, dass eine mögliche Prognose falsche inkorrekte Aussagen beinhaltet. Aus diesem Grund ist es verständlich, dass eine solch große Varianz herrscht. Die logische Konsequenz ist es, dass möglichst viele Studien durchgeführt werden sollten und man sich nicht nur auf eine eingeschränkte Menge verlassen sollte,

um die Energie-Effizienz aus allen möglichen Perspektiven zu betrachten. Diese wird nämlich auch in den nächsten Jahren noch genügend Bandbreite für Diskussionen und Weiterentwicklungen bieten.

Literatur

[AE15] Andrae, A.; Edler, T.: On Global Electricity Usage of Communication Technology: Trends to 2030. Challenges 6/, S. 117–157, Apr. 2015.

[As18] Aslan, J.; Mayers, K.; Koomey, J. G.; France, C.: Electricity Intensity of Internet Data Transmission: Untangling the Estimates. Journal of Industrial Ecology 22/4, S. 785–798, 2018, eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/jiec.12630, URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/jiec.12630.

[As19] Ashely John: The Technology That Powers Your Favorite Video Streaming Services, 2019, URL: https://hackernoon.com/technology-that-powers-your-favorite-video-streaming-services-481i32di, Stand: 30. 05. 2020.

[Da20] Daigle, Thomas: 'Completely unsustainable': How streaming and other data demands take a toll on the environment, 2020, URL: https://www.cbc.ca/news/technology/data-centres-energy-consumption-1.5391269, Stand: 22. 06. 2020.

[DAD19] Dulac, J.; Abergel, T.; Delmastro, C.: Tracking Buildings, 2019, URL: https://www.iea.org/reports/tracking-buildings/data-centres-and-data-transmission-networks, Stand: 30. 05. 2020.

[Er16] Ercan, M.; Malmodin, J.; Bergmark, P.; Kimfalk, E.; Nilsson, E.: Life Cycle Assessment of a Smartphone. In. Jan. 2016.

[Fu20] Fulton III, Scott: How Much Is Netflix Really Contributing to Climate Change?, 2020, URL: https://www.datacenterknowledge.com/energy/how-much-netflix-really-contributing-climate-change, Stand: 30. 05. 2020.

[GPN13] Gautam, N.; Petander, H.; Noel, J.: A Comparison of the Cost and Energy Efficiency of Prefetching and Streaming of Mobile Video. In: Proceedings of the 5th Workshop on Mobile Video. MoVid '13, Association for Computing Machinery, Oslo, Norway, S. 7–12, 2013, ISBN: 9781450318938, URL: https://doi.org/10.1145/2457413.2457416.

[HSN13] Hoque, M. A.; Siekkinen, M.; Nurminen, J. K.: Using Crowd-Sourced Viewing Statistics to Save Energy in Wireless Video Streaming. In: Proceedings of the 19th Annual International Conference on Mobile Computing & Networking. MobiCom '13, Association for Computing Machinery, Miami, Florida, USA, S. 377–388, 2013, ISBN: 9781450319997, URL: https://doi.org/10.1145/2500423.2500427.

[Ka18] Kafka, Peter: You can watch Netflix on any screen you want, but you're probably watching it on a TV, 2018, URL: https://www.vox.com/2018/3/7/17094610/netflix-70-percent-tv-viewing-statistics, Stand: 30. 05. 2020.

[Ka20a] Kamiya, George: Factcheck: What is the carbon footprint of streaming video on Netflix?, 2020, URL: https://www.carbonbrief.org/factcheck-what-is-the-carbon-footprint-of-streaming-video-on-netflix, Stand: 30. 05. 2020.

[Ka20b] Kamiya, George: The carbon footprint of streaming video: fact-checking the headlines, 2020, URL: https://www.iea.org/commentaries/the-carbon-footprint-of-streaming-video-fact-checking-the-headlines, Stand: 30. 05. 2020.

[KN15] Koomey, J.; Naffziger, S.: Moore's Law Might Be Slowing Down, But Not Energy Efficiency, 2015, URL: https://spectrum.ieee.org/computing/hardware/moores-law-might-be-slowing-down-but-not-energy-efficiency, Stand: 30. 05. 2020.

[Ne20a] Netflix: Environmental Social Governance 2019 Sustainability Accounting Standards Board (SASB) Report, 2020, URL: https://s22.q4cdn.com/959853165/files/doc_downloads/2020/02/0220_Netflix_EnvironmentalSocialGovernanceReport_FINAL.pdf, Stand: 30. 05. 2020.

[Ne20b] Netflix: ISP Speed Index, 2020, URL: https://ispspeedindex.netflix.com/global, Stand: 30. 05. 2020.

[Ng20] Nginx: Microservices, 2020, URL: https://www.nginx.com/learn/microservices, Stand: 30. 05. 2020.

[OS19] Omezzine, F.; Schleich, J.: The future of blockchain according to experts in the energy sector, 2019, URL: https://theconversation.com/the-future-of-blockchain-according-to-experts-in-the-energy-sector-111780, Stand: 22. 06. 2020.

[PSS19] Preist, C.; Schien, D.; Shabajee, P.: Evaluating Sustainable Interaction Design of Digital Services: The Case of YouTube. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. CHI '19, Association for Computing Machinery, Glasgow, Scotland Uk, S. 1–12, 2019, ISBN: 9781450359702, URL: https://doi.org/10.1145/3290605.3300627.

[RO18] Rüdiger, A.; Ostler, U.: Das erste Rechenzentrum in einer Windmühle läuft und die Kundschaft ist happy, 2018, URL: https://www.datacenter-insider.de/das-erste-rechenzentrum-in-einer-windmuehle-laeuft-und-die-kundschaft-ist-happy-a-752383/, Stand: 22. 06. 2020.

[Sp20] Spangler, T.: Apple TV Plus May Have More Than 33 Million Users But 'Vast Majority' Aren't Paying for It, Researcher Says, 2020, URL: https://variety.com/2020/digital/news/apple-tv-plus-33-million-users-free-year-subscribers-1203478683/, Stand: 22. 06. 2020.

[St19] Statista: Electricity consumption of video streaming worldwide in 2019, 2019, URL: https://www.statista.com/statistics/1109623/electricity-consumption-video-streaming-by-device-globally/, Stand: 22. 06. 2020.

[Su16] Summers, J.; Brecht, T.; Eager, D.; Gutarin, A.: Characterizing the workload of a netflix streaming video server. In: 2016 IEEE International Symposium on Workload Characterization (IISWC). S. 1–12, 2016.

[Ta15] Taufick, R.: Over-the-Top Content and Content Regulation, 2015, URL: http://dx.doi.org/10.2139/ssrn.2708278, Stand: 22. 06. 2020.

[Up20] Uptime: Uptime & performance monitoring made easy, 2020, URL: https://uptime.com/, Stand: 22. 06. 2020.

[Zh17] Zhang, H.; Rengasamy, P. V.; Zhao, S.; Nachiappan, N. C.; Sivasubramaniam, A.; Kandemir, M. T.; Iyer, R.; Das, C. R.: Race-To-Sleep + Content Caching + Display Caching: A Recipe for Energy-efficient Video Streaming on Handhelds. In: 2017 50th Annual IEEE/ACM International Symposium on Microarchitecture (MICRO). S. 517–531, 2017.

Betriebliche Umweltinformatik und nachhaltige Entwicklung: ein grenzüberschreitendes Praxisbeispiel aus Ägypten

Reimund Lepiorz[1] Slim Abdennadher[2] Ralf Klischewski[3] Volker Wohlgemuth[4]

Abstract: Die betriebliche Umweltinformatik als anwendungsorientierte Disziplin kann gerade durch die Kombination mit IT-gestützten Verfahren im Unternehmenskontext zur umweltgerechten Modernisierung sowie zu nachhaltigen und ressourceneffizienten Lösungen beitragen. Mit diesem Ziel setzt das Transferprojekt SUSTAIN Impulse für Innovationen zur Bewältigung wirtschaftlicher und ökologischer Probleme in Ägypten, oder potenziell auch in vergleichbaren Ländern. Das Projekt unterstützt den Aufbau von angewandter Forschung und unternehmensorientierten Dienstleistungen an der German University in Cairo sowie die Entwicklung von Bildungsangeboten zur betrieblichen Qualifizierung und in Form eines Masterprogrammes. Bislang wurden mehrere Train-the-Trainer-Workshops durchgeführt und die erworbenen Kompetenzen in einer Reihe von Fallstudien zur Anwendung gebracht. Die beispielhaften Ergebnisse sowie Dienstleistungen zur Verbesserung von Ressourceneffizienz und Nachhaltigkeit in der Produktion wurden einem betrieblichen Fachpublikum vorgestellt. In weiteren Schritten sind die Bildungsangebote als professionelle Module vorzubereiten und ein Netzwerk aufzubauen, das auch langfristig ein Forum für einen wissenschaftlichen und gesellschaftspolitischen Austausch zu den zentralen Themen des Projektes bietet.

Keywords: Betriebliche Umweltinformatik; Nachhaltige Produktion; Digitale Transformation; SDGs; Agenda 2030; Entwicklungszusammenarbeit; Ägypten; Wissens- und Technologietransfer

1 Die Agenda 2030 und nachhaltige Produktionsprozesse in Ägypten

Mit der Verabschiedung der Agenda 2030 im Jahr 2015, den Sustainable Development Goals (SDGs), hat sich die Weltgemeinschaft ein Konzept für eine globale Entwicklung gegeben. Die zurückliegende Zeit seit den siebziger Jahren war durch die Millennium Development Goals (MDGs) geprägt. Die SDGs haben im Unterschied zu den MDGs universellen Charakter und der wesentliche Aspekt, der die Sichtweise der vorherigen Ziele ergänzt, ist Nachhaltigkeit. Sie ist zum zentralen Prinzip globaler Entwicklung erhoben worden. Mit der Agenda 2030 ist damit ein weltumspannender Transformationsprozess anvisiert, ein Modernisierungsprozess, der wesentlich auch eine klimaverträgliche Gesellschaft zum Ziel hat [WB19].

[1] Hochschule für Technik und Wirtschaft Berlin, Wilhelminenhofstraße 75a, 12459 Berlin, Deutschland, reimund.lepiorz@htw-berlin.de
[2] German University in Cairo (GUC), 11835 New Cairo City, Ägypten, slim.abdennadher@guc.edu.eg
[3] German University in Cairo (GUC), 11835 New Cairo City, Ägypten, ralf.klischewski@guc.edu.eg
[4] Hochschule für Technik und Wirtschaft Berlin, Wilhelminenhofstraße 75a, 12459 Berlin, Deutschland, volker.wohlgemuth@htw-berlin.de

doi:10.18420/inf2020_23

Die SDGs umfassen 17 Punkte, die sich in 169 Zielvorgaben gliedern [Na15]. Mit ihnen sollen unter anderem Umweltschutz und wirtschaftlicher Fortschritt in Einklang gebracht werden bei gleichzeitiger Achtung sozialer Interessen. Im Zentrum des SDG 12 stehen so zum Beispiel nachhaltige Konsum- und Produktionsmuster. Von einer Umstellung auf nachhaltige Produktionsmethoden verspricht man sich positive Umwelteffekte. Damit steht auch nachhaltiges Unternehmenshandeln im Fokus. Sie sind nicht nur aufgefordert, nachhaltige Verfahren einzuführen, sondern auch verstärkt über Nachhaltigkeit zu informieren.

Obwohl digitale Lösungen bereits zum industrieökologischen Wandel beitragen und besonders unter dem Aspekt Industrie 4.0. breit diskutiert werden, kommen sie in den SDGs im Zusammenhang mit den Transformationsprozessen nur am Rande vor. Ressourceneffizienz und optimierte Produktionsabläufe durch digitale Methoden sind hingegen ein Kriterium innerhalb der Nachhaltigkeitsdebatte und werden beispielsweise auch im Zusammenhang mit Entwicklungszusammenarbeit diskutiert [DE19].

Ägypten hat sich wie alle 193 Mitgliedsstaaten der Vereinten Nationen zu den SDGs als universeller Zielvorgabe bekannt [Na18]. Das Land steht vor spezifischen ökologischen Herausforderungen. Ein hohes Bevölkerungs- und auch industrielles Wachstum haben massive Umweltbelastungen zufolge, z.B. durch unsachgemäße Entsorgung von Industrieabfällen. Auch ein nachhaltiges Wasser- und Abwassermanagement ist für Ägypten, einem der wasserärmsten Länder der Welt, von essentieller Bedeutung [GI20]. Gleichzeitig ist man bemüht, den eigenen Energieverbrauch zu senken und erneuerbare Energien auszubauen. Um etwa die Klimaschutzziele zu erreichen, die für das Land auch von strategischer Bedeutung sind, soll zum Beispiel der Anteil der erneuerbaren Energien an der Stromerzeugung von derzeit neun Prozent bis 2022 auf 20 Prozent steigen [LI20]. Es fehlt allerdings an praxisorientierter Ausbildung für die entsprechenden Berufsbilder. Relevant in einem industrieökologischen Kontext sind die knapper werdenden Ressourcen, für die Ägypten Lösungen durch eine gesteigerte Ressourceneffizienz im Industriesektor finden will. Für die Implementierung innovativer Lösungen im Bereich der Energie- und Ressourceneffizienz sowie nachhaltigen Produktion fehlt es allerdings ebenso an Fachkräften, auch adäquate Bildungsangebote und berufsspezifischen Qualifizierungen sind in diesem Sektor noch rar [GI20].

2 Das Transferprojekt SUSTAIN

Das Projekt SUSTAIN - „Sustainable Production & Digital Transformation“ (SUSTAIN) - ist ein Gemeinschaftsprojekt der Hochschule für Technik und Wirtschaft (HTW) Berlin und der German University in Cairo (GUC). SUSTAIN wird vom Auswärtigen Amt (AA) im Zeitraum von März 2019 bis Ende 2020 gefördert und vom Deutschen Akademischen Austauschdienst (DAAD) im Rahmen der deutsch-ägyptischen Fortschrittspartnerschaft betreut. Das Projekt verfolgt im Kern die Idee, den in Deutschland seit mehreren Jahrzehnten etablierten Ansatz des Stoffstrommanagements [He98, vgl.] nach Ägypten zu transferieren, aber auch mit modernen IT-Ansätzen zu digitalisieren und so Transformationsprozesse in einem industrieökologischen Zusammenhang anzuregen.

Die GUC wurde 2002 gegründet und bietet rund 70 Studiengänge, darunter eine Reihe von ingenieurswissenschaftlichen Studiengängen, für derzeit etwa 13.000 Studierende. Ressourceneffizienz und nachhaltige Produktionsmethoden mit dem Fokus auf betrieblicher Umweltinformatik stellen ein innovatives Forschungsgebiet in ägyptischen Hochschulen und anwendungsorientiert auch auf dem ägyptischen Markt dar. Es gibt derzeit keinen Studiengang an ägyptischen Hochschulen, der dieses Thema ins Zentrum stellt. In Hinblick auf die SDGs und die Zielvorgaben in Ägypten besteht daher Interesse an einer Zusammenarbeit, um dieses Forschungsfeld zu etablieren.

SUSTAIN adressiert oben genannte Herausforderungen in Ägypten und verfolgt multidimensionale Ziele, die sich im Wesentlichen um Forschung, Qualifizierung und Transfer gruppieren. Den Rahmen bildet der Aufbau struktureller Forschungs- und Serviceaktivitäten zu „Sustainable Production and Digital Transformation“ an der GUC. Unter dem Dach dieser Forschungsaktivitäten ist einer der Forschungsschwerpunkte die betriebliche Umweltinformatik. Es handelt sich um eine junge und innovative Disziplin, die den scheinbaren Widerspruch zwischen ökonomischen, ökologischen und sozialen Zielen entkräftet. Sie entwickelt und setzt IT-Systeme so ein, dass von einem Betrieb geringstmögliche Umweltbelastungen ausgehen. Gleichzeitig wird ein effektives Umweltmanagement im Betrieb gefordert. Darüber hinaus unterstützt die betriebliche Umweltinformatik die Planung, Kontrolle und Steuerung betrieblicher Nachhaltigkeitsaktivitäten. Durch das Aufdecken von Energie- und Materialeinsparpotentialen lassen sich sowohl Kosten reduzieren, als auch natürliche Ressourcen schonen. Der Beitrag, den diese Disziplin zur Harmonisierung insbesondere von Ökologie und Ökonomie leistet, scheint bezogen auf Ägypten besonders relevant zu sein.

Die multidimensionale Ausrichtung des Projektes sieht neben Qualifizierungsangeboten an der GUC auch ein erweitertes Ausbildungsangebot in Form eines Masterprogrammes zur verbesserten Beschäftigungsfähigkeit für Absolventen der Universität vor. Die Serviceleistungen selber sind transferorientiert und wollen mit ihren Angeboten nachhaltiges Handeln in die Region fördern und Anreize für eine langfristige wirtschaftliche und soziale Innovationsfähigkeit schaffen. Das Forschungs- und Servicecenter ist damit als zentrale Anlaufstelle für Akteure aus Wirtschaft, Politik und Gesellschaft in Fragen der Ressourceneffizienz und nachhaltiger Produktionsmethoden konzipiert und könnte langfristig ein Forum für einen wissenschaftlichen und gesellschaftspolitischen Austausch zu den zentralen Themen des Projektes sein.

In Verlauf von mehreren gegenseitigen Exkursionen im Jahr 2019 erarbeiteten die Projektpartner ein strukturelles Gerüst für den Aufbau der Forschungs- und Serviceleistungen an der GUC. Dabei standen neben Gastvorlesungen an der HTW Berlin auch industrielle Anwendungsfälle auf der Agenda, die Ressourcen- und Energieeffizienz zum Gegenstand hatten. Zentral war die Entwicklung eines praxistauglichen Trainingskonzepts, das die Projektpartner als Train-the-Trainer-Workshops für die zunächst interne Schulung an der GUC konzipierten. Eine qualifizierte Ausbildung des akademischen Personals der GUC schafft erst die Voraussetzungen, das Know-how in einen anwendungsorientierten Kontext

mit Akteuren aus der regionalen Wirtschaft zu bringen und langfristig beispielsweise als Fortbildungskonzept zu konsolidieren. In die Workshops sind Nachwuchswissenschaftler und Studierende eingebunden, die das erworbene Know-how in unterschiedlichen Konstellationen anwenden sollen. Angedacht sind Inhouse-Seminare in den Unternehmen, Praktika oder die beratende Funktion in oder für Unternehmen.

2.1 Methoden

Im Zentrum steht der Know-how-Transfer zu Methoden des Energie- und Stoffstrommanagements. Das Verfahren beschreibt die tiefgreifende Analyse und gezielte Optimierung von Material- und Energieströmen, die bei der Herstellung von Gütern oder bei Dienstleistungen entstehen [En18]. Es dient somit der zielorientierten, ganzheitlichen Analyse sowie Steuerung und Kontrolle von Stoff- und Energieströmen in betrieblichem Kontext, um Effizienzsteigerungen von Produktionsprozessen zu erreichen und schließt international etablierte Methoden ein. Im Vordergrund stehen im Zusammenhang mit dem Projektverlauf Life Cycle Assessment (LCA) und Material Flow Cost Accounting (MFCA).

LCA, zu Deutsch Lebenszyklusanalyse oder auch Ökobilanz, ist eine systematische Analyse und Bewertung der Umweltwirkungen von Produkten für deren gesamten Lebenszyklus („cradle-to-grave“). Sie wurden in den ISO-Standards 14040:2006 und 14044:2006 international festgelegt und in das deutsche Normenwerk übertragen [Um18]. Eine ausgeführte Ökobilanz besteht zum einen aus der Analyse der Stoff- und Energieströme des gesamten Produktsystems inklusive aller beteiligten Prozesse entlang des Lebensweges eines Produktes, also von der Rohstoffgewinnung bis zur Entsorgung. Außerdem werden Emissionen in Luft, Wasser und Boden sowie der Natur entnommene Ressourcen systematisch erfasst. Im Anschluss erfolgt im Rahmen der Wirkungsabschätzung die Auswertung der potenziellen Umwelteffekte.

Die Materialflusskostenrechnung (engl. Material Flow Cost Accounting, MFCA) bewertet die Effizienz von Produktionsprozessen im Hinblick auf Materialverluste. MFCA ist unter der Norm ISO 14051 ebenfalls standardisiert [No11] und will Kosten aufdecken, die aus Verlusten während der Erzeugung von Produkten in der Prozess- und Fertigungsindustrie entstehen. Die klassische Kostenkalkulation eines Produktes basiert darauf, dass alle in die Produktion fließenden Rohstoffe letztlich im Produkt landen und in die Preisgestaltung einbezogen werden [NM04, S.3-5], [Or14, S.3-6]. Die Menge an Rohstoffen, die als Abfall nicht im Produkt landen, wird dabei aber vernachlässigt. Jedoch macht es nur die Kenntnis dieser Mengen und Kosten möglich, Abfall zu reduzieren. Die Ergebnisse der MFCA bilden damit eine Entscheidungsgrundlage, um Materialkosten zu reduzieren. Die Methode kann somit negative unternehmerische Umweltwirkungen verringern und gleichzeitig mehr ökonomische Effizienz erreichen.

MFCA eignet sich zudem für eine systematische Softwareunterstützung. Die IT-gestützte Anwendung bringt Zeit- und Kostenersparnisse in der Produktion mit sich, indem sie das

Verarbeiten von Daten zu Stoff- und Energieströmen vereinfacht und die Ergebnisse zudem visualisiert, beispielsweise in Sankey-Diagrammen. Der Digitalisierung kommt damit Bedeutung bei der effizienteren Umsetzung des Stoffstrommanagements zu. Die Verbindung aus nachhaltiger Produktion und der Investition in IT-gestützte Optimierungsprozesse trägt außerdem zur Innovationsfähigkeit der Unternehmen bei. Die Idee ist nun, diese Methoden in digitalisierter Form in Ägypten zum Einsatz zu bringen.

Eine auf dem Markt erhältliche Software, die eine MFCA im betrieblichen Kontext ermöglicht, ist Umberto Efficiency+. Deswegen war im Rahmen der Schulungen für das akademische Personal der GUC an der HTW Berlin der Umgang mit diesem IT-unterstützenden Software-Paket in den ersten Projektphasen grundlegend. Neben dem Einsatz gängiger Softwaretools verfolgte das Projekt auch den ehrgeizigen Plan, ganze Prozessabläufe, von der Erfassung der Stoffstromdaten über die Analyse und Auswertung bis zur Ergebnisdarstellung zu digitalisieren. Diese Idee musste jedoch mangels Ressourcen eingeschränkt werden und umfasst zurzeit nur noch die Digitalisierung der Erfassung von Stoffströmen.

2.2 Wissens- und Technologietransfer

Nach bisherigem Kenntnisstand gibt es in Ägypten und der Region keine nennenswerten Angebote im Bereich der Nachhaltigkeit und des betrieblichen Umweltschutzes zur Verbesserung der Energie- und Ressourceneffizienz von produzierenden Unternehmen. SUSTAIN ist in diesem Sinne ein Pilotprojekt, das sukzessive entwickelt wird und im Projektverlauf auch stetige Justierungen erfordert.

Die einzelnen Phasen im Projektverlauf sind an die jeweiligen personellen und fachlichen Ressourcen der beiden Hochschulen angepasst. In der ersten Projektphase von April bis Ende des Jahres 2019 wurde das akademische Personal der GUC in mehreren Schritten in einem Portfolio an Fachwissen zu Begriffen, Methoden und Maßnahmen im Bereich der nachhaltigen Produktion und der Ressourceneffizienz geschult. Das vermittelte Know-how wird in ersten Feldstudien mit lokalen Unternehmen zur Anwendung gebracht und evaluiert. In diese Zusammenarbeit – ebenso wie in die Train-the-Trainer Workshops – sind akademische Nachwuchskräfte der GUC eingebunden, die sich in vorherigen Bewerberscreenings an der GUC für die Teilnahme bewarben.

An den Feldstudien sind derzeit drei ägyptische Unternehmen aus dem verarbeitenden Gewerbe beteiligt. Die untersuchten Kernprozesse sind in der Möbelherstellung, der Glasverarbeitung und der Herstellung von Sicherheitssystemen angesiedelt. Das Ziel der Studien ist, erste Prozessdaten aus der laufenden Produktion mithilfe von Umberto Efficiency+ zu erfassen und die Daten in dem Software-Tool zu modellieren. Nach Analyse und Visualisierung der Daten konnten bisher bei zwei der drei Unternehmen konkrete Modifikationen im Prozessablauf vorgeschlagen werden. Zum Beispiel ist das glasverarbeitende Gewerbe sehr energieintensiv, und einer der Prozessschritte, die effizienter gestaltet werden können, betrifft

die Abwärme des Schmelzofens: Ein Recycling der Abwärme durch eine Rückführung in einem Tunnelsystem kann laut erster Analyse signifikante Energieeinsparungen bringen. Das auf Sicherheitssysteme und Schlösser spezialisierte Unternehmen ist im Wesentlichen ein metallverarbeitender Betrieb und einer der größten und ältesten Betriebe dieser Branche in Ägypten. Auch hier wurde der gesamte Herstellungsprozess mit der Software abgebildet, vom Metallzuschnitt über Phosphatierung, Trocknung, Lackierung und Aushärtung der Farbe, um auf dieser Basis Optimierungspotenziale zu identifizieren. Diese betreffen den Einsatz höherwertiger Aluminiumkomponenten, ein verbessertes Werkzeugdesign zur Verringerung des Ausschusses sowie eine Neuordnung der Lagerung, so dass noch kalte Rohmaterialen die Abwärme gerade verarbeiteter Komponenten aufnehmen können. Die ggf. umgesetzten Empfehlungen werden in einem weiteren Schritt erneut modelliert, um die Einsparpotenziale konkret beziffern zu können.

Zu Beginn der zweiten Phase fand im Februar ein Symposium mit ca. 50 Teilnehmenden aus Industrie und Wissenschaft statt, auf dem die Methoden und damit verbundenen Potenziale anhand der erarbeitenden Fallstudien vorgestellt und die Umsetzung vor allem mit Vertretern kleinerer und mittlerer Unternehmen diskutiert wurden. Im Zentrum des Interesses stand dabei die Frage, wie der anstehende Digitalisierungsschub in ägyptischen Unternehmen mit den Methoden des Stoffstrommanagements sinnvoll verbunden werden kann. Besondere Bedeutung kommt über den gesamten Projektverlauf hinweg der Netzwerkarbeit zu, in die die Gesellschaft für internationale Zusammenarbeit (GIZ) in Kairo eingebunden ist. Gleichzeitig ist die Institutionalisierung des Wissensgebietes in Form eines Masterstudienganges für betriebliche Umweltinformatik an der GUC geplant.

3 Ausblick

Die betriebliche Umweltinformatik als anwendungsorientierte Disziplin hat im Unternehmenskontext gerade durch die Kombination mit IT-gestützten Verfahren Potenziale, zur umweltgerechten Modernisierung sowie nachhaltigen und ressourceneffizienten Lösungen beizutragen. Als Transferprojekt setzt SUSTAIN Impulse für Innovationen zur Bewältigung wirtschaftlicher und ökologischer Probleme in Ägypten, oder potenziell auch in vergleichbaren Ländern. Eine nachhaltige Wirkung hängt allerdings wesentlich von lokalen und politischen Rahmenbedingungen ab.

Ein übergeordnetes Ziel des Projektes ist die Entwicklung eines gemeinsamen Masterprogramms, das über den Projektzeitraum hinaus zur Institutionalisierung der Wissensdisziplin beitragen soll. Inhalte und Ausrichtung des Masterprogramms ergeben sich aus den Erkenntnissen der bisher noch laufenden Projektphase. Eine solche Institutionalisierung kann den wissenschaftlichen Nachwuchs in einem interkulturellen, erweiterten und praxisbezogenen Kontext fördern. Und sie fördert den Transfer über Köpfe, indem insbesondere Absolventen Zugang zu einem transdisziplinären Netzwerk erhalten, das um den zentralen Ankerpunkt des Forschungs- und Servicecenters an der GUC entstehen soll.

SUSTAIN wird gefördert vom Deutschen Akademischen Austauschdienst (DAAD) aus Mitteln des Auswärtigen Amts (AA).

Literatur

[DE19] DEval: Nachhaltigkeit gestalten: Die Agenda 2030 in der Entwicklungszusammenarbeit, Themenschwerpunktbericht, Deutsches Evaluierungsinstitut der Entwicklungszusammenarbeit, 2019, URL: https://nbn-resolving.org/urn:nbn:de:0168-ssoar-64920-5, Stand: 27. 04. 2020.

[En18] Enquête-Kommission: Schutz des Menschen und der Umwelt: Umweltverträgliches Stoffstrommanagement Band I. Deutscher Bundestag, 2018.

[GI20] GIZ, G. f. I. Z.: Ägyptisch-Deutsches Komitee zur Förderung der Erneuerbaren Energien, der Energieeffizienz und des Umweltschutzes (JCEE), 2020, URL: https://www.giz.de/de/weltweit/16274.html, Stand: 30. 04. 2020.

[He98] Henseling, K.: Grundlagen des Managements von Stoffströmen. In: Das Management von Stoffströmen. Springer, Berlin, Heidelberg, S. 17–85, 1998.

[In20] International, K.: Länderbericht Ägypten, 2020, URL: https://www.kooperation-international.de/laender/afrika/aegypten/laenderbericht/?tx%5C_contentaggregation%5C_pages%5C%5Baction%5C%5D=list%5C&tx%5C_contentaggregation%5C_pages%5C%5Bcontroller%5C%5D=AggregatePages, Stand: 28. 04. 2020.

[LI20] LIPortal: Länderinformationsportal, 2020, URL: https://www.liportal.de/aegypten/ueberblick/, Stand: 28. 04. 2020.

[Na15] Nationen, V.: Transformation unserer Welt: die Agenda 2030 für nachhaltige Entwicklung, Resolution der Generalversammlung, 25. Sep. 2015, URL: http://www.un.org/Depts/german/gv-70/band1/ar70001.pdf', Stand: 19. 02. 2019.

[Na18] Nationen, V.: Sustainable Development Goals, Egypt, 2018, URL: https://sustainabledevelopment.un.org/memberstates/egypt, Stand: 28. 04. 2020.

[NM04] Nakajima; Michiyasu: On the Differences between Material Flow Cost Accounting and Traditional Cost Accounting - In Reply to the Questions and Misunderstandings on Material Flow Cost Accounting. Kansai University Review of Business and Commerce/, 2004.

[No11] für Normung, D. I.: DIN EN ISO 14051: Environmental Management – Material Flow Cost Accounting – General Framework (ISO 14051:2011). Berlin, 2011.

[Or14] Organization, A. P.: Manual on Material Flow Cost Accounting ISO 14051, Tokyo, 2014.

[Um18] Umweltbundesamt: Ökobilanz, 17. Okt. 2018, URL: https://www.umweltbundesamt.de/themen/wirtschaft-konsum/produkte/oekobilanz, Stand: 27. 03. 2020.

[WB19] WBGU: Unsere gemeinsame digitale Zukunft. Wissenschaftlicher Beirat der Bundesregierung Globale Umweltveränderungen, Berlin, 2019.

Analyse von Heizungs- und Lüftungsverhalten mit Data Mining Methoden

Erste Ergebnisse aus den Messwerten der ersten Heizperiode

Lutz Westhäusser[1] David Nickel[2] Grit Behrens[3] Klaus Schlender[4]

Abstract: In dem hier beschriebenen Projekt wird interdisziplinär mit Psychologen zusammen gearbeitet. Ziel der Arbeit ist es, Modelle zu entwickeln, um das Umweltverhalten von Hausbewohnern positiv zu beeinflussen und zu verstetigen. In der hier beschriebenen Arbeit werden die ersten Daten aus dem ‚Reallabor' Sennestadt genutzt, die in den Wohnungen von freiwilligen Studienteilnehmern zu ihrem Heizungs- und Lüftungsverhalten erhoben werden. Mittels Machine Learning Technologien werden diese Daten analysiert.

Keywords: Data Mining; umweltbewusstes Verhalten; Heiz- und Lüftungsverhalten; Machine Learning

1 Einleitung

In dem Forschungsprojekt ‚ENVIRON' (Förderkennzeichen 01UT1703A) sollen die Bewohner von Wohngebäuden begleitet werden, deren Wohnungen in Rahmen einer energetischen Sanierung renoviert werden. Im Zuge von verbesserten Isolierungen an Hauswänden und Fenstern ist die Begleitung der Bewohner bei der Umsetzung von neuen, umweltbewussten Verhaltensweisen ein wichtiger Faktor.

2 Erkennen von Phasen im Heizungs- und Lüftungsverhalten

2.1 Maschinelles Lernen von Phasen des Umweltenergieverbrauchs

Zu Beginn der Umsetzung in den einzelnen Wohnungen, wurden die Bewohner mit einem Telefoninterview zu ihrem Verhalten befragt. Aufgrund einer Definition von den in dem Projekt beteiligten Psychologen konnten die Bewohner in verschiedene Phasen für ihr Umweltverhalten eingeteilt werden.

[1] FH-Bielefeld University of Applied Sciences, Artilleriestr. 9, 32427 Minden, Lutz.Westhaeusser@fh-bielefeld.de
[2] FH-Bielefeld University of Applied Sciences, Artilleriestr. 9, 32427 Minden, David.Nickel@fh-bielefeld.de
[3] FH-Bielefeld University of Applied Sciences, Artilleriestr. 9, 32427 Minden, Grit.Behrens@fh-bielefeld.de
[4] FH-Bielefeld University of Applied Sciences, Artilleriestr. 9, 32427 Minden, Klaus.Schlender@fh-bielefeld.de

doi:10.18420/inf2020_24

Die vier Typen sind:

- Pre-decisional (1): Die Person hat noch keine konkrete Zielintension entwickelt, das eigene Heizverhalten zu ändern.
- Pre-actional (2): Die Person hat eine Zielintension entwickelt, diese aber noch nicht angefangen umzusetzen.
- Actional (3): Die Person ist dabei, sein Ziel umzusetzen.
- Post-actional (4): Die Person hat ein umweltbewusstes Verbrauchsverhalten.

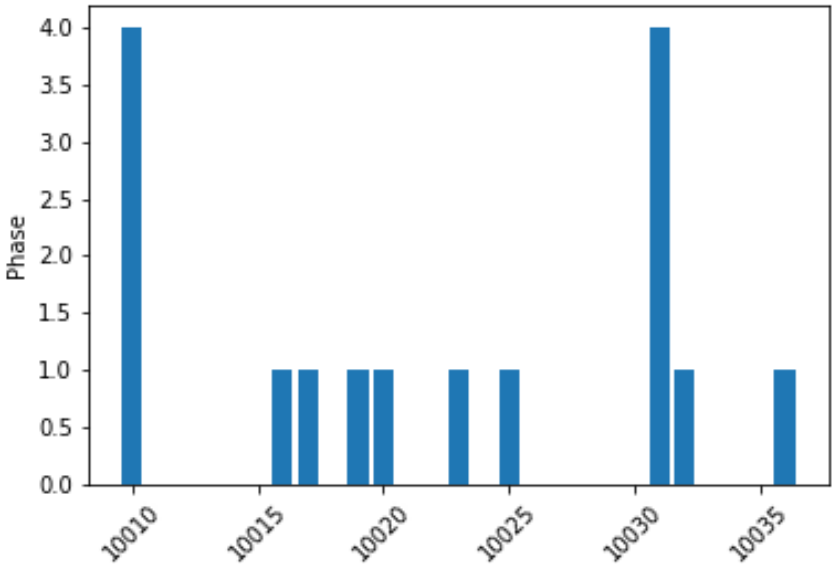


Abb. 1: Phasenanalyse, 2 Wohnungen in Phase 4, 8 Wohnungen in Phase 1

In der Abbildung 1 sind die Ergebnisse der Fragebogenauswertung grafisch dargestellt. Die Einteilung der Wohnungen erfolgte nach den Vorgabe der Psychologen.

2.2 Ziele für die einzelnen Gruppen

Ziel ist es das die Personen immer in die nächsthöhere Gruppe mit dem verbesserten Umweltverhalten aufsteigen. Im Zuge des Verhaltensübergangs zu einem umweltfreundlicheren Verhalten wird jeder Bewohner in jeder einzelnen Phase mit spezifischen Problemen konfrontiert, die gelöst werden müssen, um das Verhalten erfolgreich zu ändern. Um einen

näheren Einblick zu geben, werden die vier Stufen, die Bamberg [Ba13] entwickelt hat, im Folgenden kurz aufgelistet.

Dazu sind für jeden Typ Unterziele definiert:

- Pre-decisional (1): Ziel ist es eine Zielintention zu wecken, indem die Diskrepanz zwischen dem tatsächlichen Ergebnis des eigenen Verhaltens und gewünschten aufgezeigt wird. Wichtig ist in dieser Gruppe die Machbarkeit der Änderung des Verhaltens aufzuzeigen.
- Pre-actional (2): Ziel ist es eine Verhaltensintention zu wecken, indem ein spezifisches Ziel gebildet wird. Es Abwägen verschiedener Strategien und wählen der für einen persönlich geeignetsten.
- Actional (3): eine Implementierungsintention zu wecken. Hilfen sind: Bewältigungsplanung, Bewältigungs-Selbstvertrauen, Handlungsplanung.
- Post-actional (4): Intention, nicht ins alte Verhalten zurück zu fallen. Wecken des Vertrauens in die Fähigkeit der eigenen Person, nicht in das Verhalten aufrecht zu halten.

3 Sensoranalyse

Mit den von den Psychologen vorgegeben Definitionen konnten dann anonym die Daten der Sensoren zugeordnet werden.

3.1 Sensorvergleich bei zwei Wohnungen

Für die ersten Vergleiche sind zwei Wohnungen ausgewählt werden, die sich in zwei unterschiedlichen Phasen befinden. Durch unterschiedliche Installationszeiten sind auch die Datenreihen in der Datenbank erst nacheinander gefüllt worden, so das es im ersten Schritt notwendig war, die Daten nach Zeiträumen zu analysieren, in denen die Basis mit ausreichend vielen Einträgen gefüllt ist. In der Grafiken befindet sich die Wohnung in Phase 4 immer auf der linken Seite, die Wohnung in der Phase 1 auf der rechten Seite. Beispielhaft wurde ein Sensor gewählt, der in den Wohnzimmer der Wohnungen die Raumtemperatur, die Luftfeuchtigkeit und den Luftdruck misst.

Es ist schon bei dem ersten Betrachten der Grafik eine Unterschied zu bemerken. In der linken Wohnung (Phase 4) ist die Raumtemperatur und die Luftfeuchtigkeit der gemessen Raumes erheblich niedriger.

Zum Vergleich der Werte ist hier die Grafik für einen weiteren Messtag aufgeführt. An diesem Tag ist die Temperatur nicht so unterschiedlich wie in der vorherigen Grafik, allerdings ist der Unterschied in der Luftfeuchtigkeit höher.

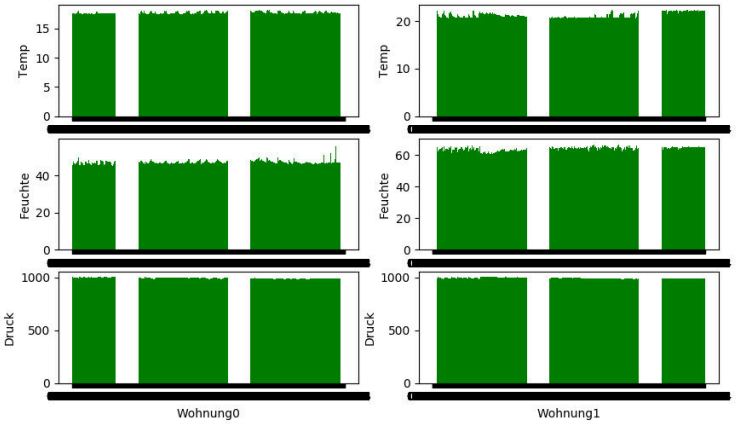


Abb. 2: Sensorauswertung für den 13.02.2020

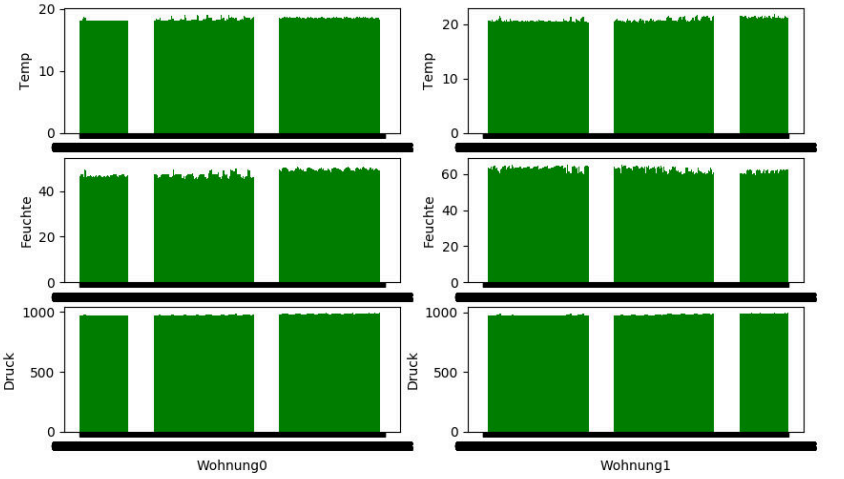


Abb. 3: Sensorauswertung für den 06.03.2020

Auffällig bei den Graphen ist auch, das die Datensätze nicht vollständig sind. Im Laufe der Messungen ist es immer zu Ausfällen in der Kommunikation gekommen, die sich in allen Teilen des Messaufbau wiederfinden.

3.2 Sensorvergleich bei allen Wohnungen

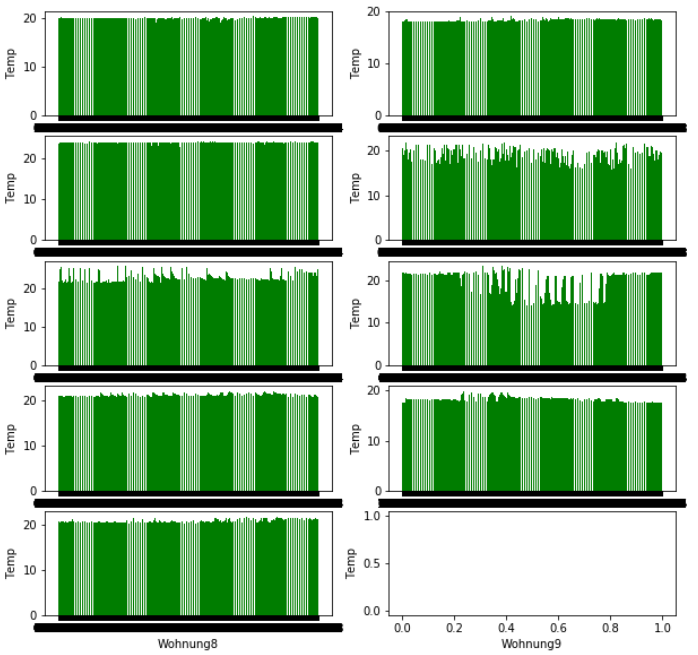


Abb. 4: Sensorauswertung für alle Wohnungen

Für einen Vergleich ist hier die Raumtemperatur des Wohnzimmers aller untersuchten Wohnungen grafisch dargestellt. Auch hier sind in den Messwerten deutliche Unterschiede in dem Verhalten der einzelnen Wohnungsbewohner zu erkennen.

4 Datenanalyse

Für die Datenanalyse sind, wie schon erwähnt, zwei Wohnungen ausgewählt worden. Dazu kommt aus einem vorherigen Projekt schon eine vorhanden Datenbasis hinzu, bei der schon Berechnungen für einen Vergleich mit Daten aus der Nachbarschaft durchgeführt wurden. Bei allen Grafiken ist die Wohnung in der Phase 4 auf der linken Seite, die Wohnung in der Phase 1 auf der rechten.

4.1 Vergleich über 30 Tage

Beispielhaft ist hier eine Analyse über die Leistung eines Heizkörpers in einem Zeitraum von 30 Tagen.

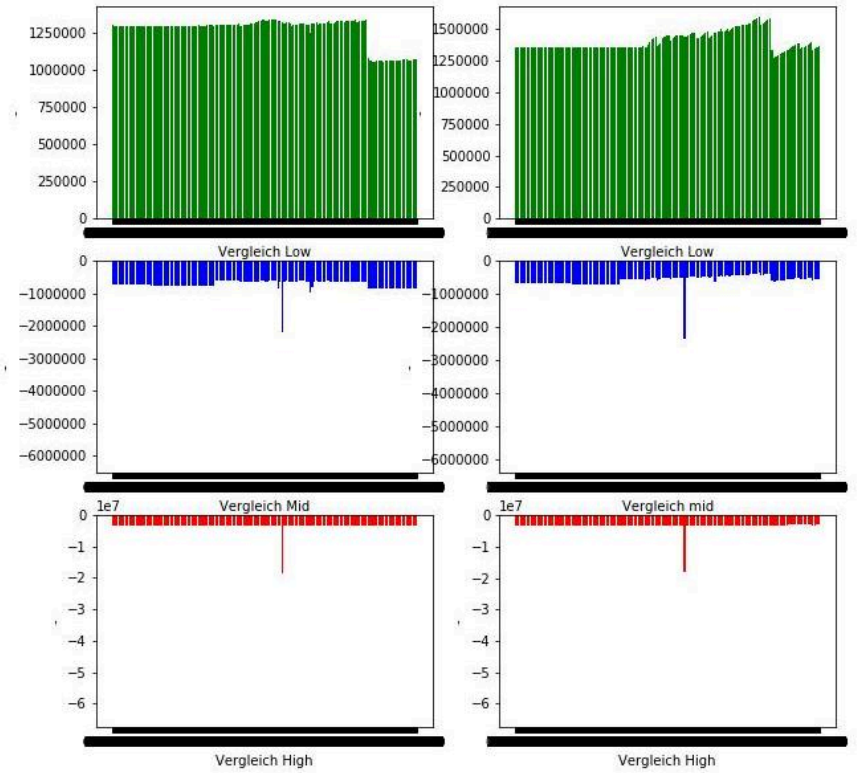


Abb. 5: Heizkörpertemperatur über 30 Tage

In diesem Fall wurde zum Vergleich aus der Nachbarschaftstabelle die berechneten Daten der Heizkörperleistung hinzugezogen. Es wurden drei Vergleiche berechnet:

- der Vergleich Low ist die Berechnung der Heizkörperleistung der Wohnung mit der als kleinsten Wert angenommen Heizleistung.
- der Vergleich Mid ist die Berechnung der Leistung mit dem errechneten Mittelwert der Heizkörperleistungen
- der Vergleich High ist die Berechnung der Heizkörperleistung der Wohnung mit der als größten Wert angenommen Heizleistung.

Bei der Analyse der Ergebnisse ist ein erkennbarer Unterschied in der Heizleistung sichtbar.

4.2 Vergleich über 24 Stunden

Für den Vergleich über 24 Stunden ist die gleiche Berechnung angestellt worden, wie bei dem Vergleich über 30 Tagen.

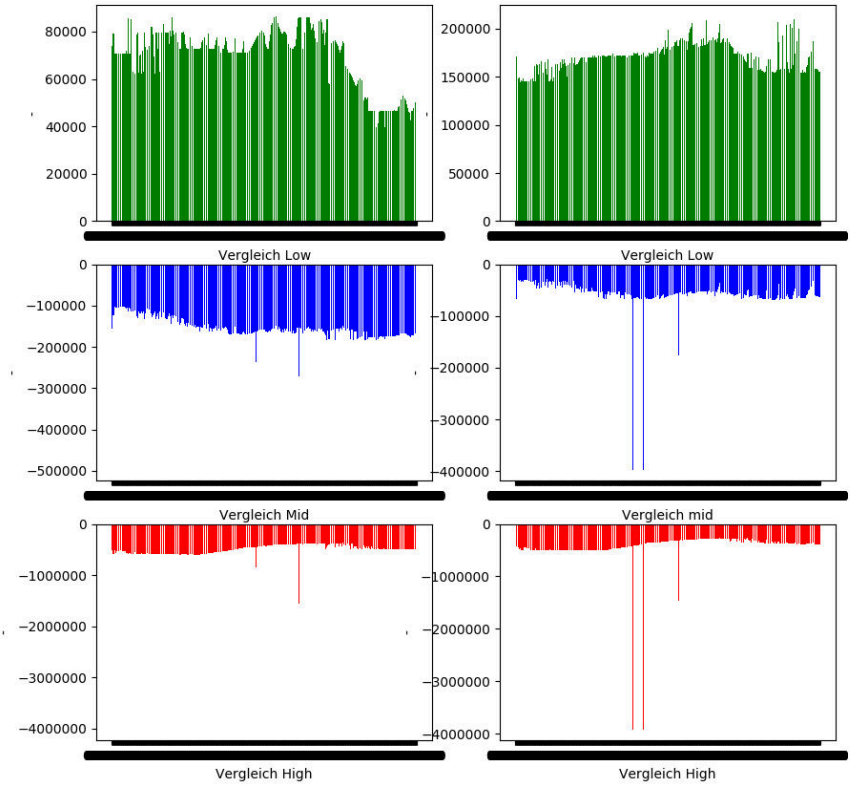


Abb. 6: Vergleich 24 Stunden

Da bei beiden Vergleichen immer die Summen der Werte der davor liegenden 24 Stunden gebildet werden, ist der Vergleich über 24 Stunden differenzierter und die Unterschiede sind deutlicher zu erkennen. Wenn in dieser Grafik z.B. der Bereich des Vergleichs Low betrachtet wird, sind unterschiedliche Kurven zu sehen. Da auch die Skalen der Grafiken Unterschiede aufweisen, ist hier schon die geringere Nutzung der Heizung in der Wohnung der Phase 4 ersichtlich.

4.3 Datenkorrelation

Um einen Zusammenhang deutlicher zu machen, wurden auch schon die Korrelationen gemacht.

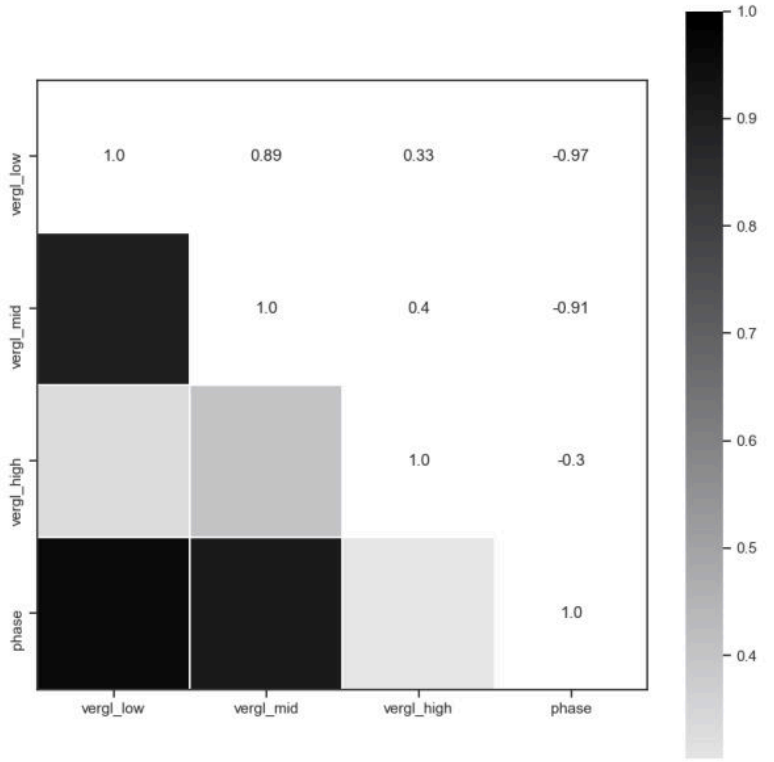


Abb. 7: Korrelationsmatrix des Heizkörpervergleichs

In dieser Matrix sind die Zusammenhänge in Graustufen dargestellt. Je dichter der Wert an dem Wert 1 liegt, desto dunkler wird er dargestellt. Dadurch ist es auch leicht ersichtlich, das es hier einen hohen Zusammenhang zwischen der Phase und der mittleren Heizkörperleistung gibt. Weitere Auswertungen in dem Bereich sollen folgen.

5 Fazit und Ausblick

Diese Arbeit ist zu dem jetzigen Zeitpunkt nur eine erste Grundanalyse der Daten in dem Projekt ‚ENVIRON'. Weitere Auswertungen der Datensätze werden nun erfolgen und auch die Fehlerkorrektur wird noch implementiert.

Als Zwischenfazit des Projektes kann jetzt schon gesagt werden, das es noch ein großes Potenzial in dem Punkt des effektiven Heizverhaltens gibt. Dieses zu aktivieren würde zu einer größeren Senkung in der CO2-Erzeugung führen und der Klimaerwärmung entgegenwirken.

Literaturverzeichnis

[Ba13] Bamberg, S.: Applying the stage model of self-regulated behavioral change in a car use reduction intervention. Journal of Environmental Psychology. S. 33, 68-75, 2013.

Wir danken dem Bundesministerium für Bildung und Forschung und FONA Sozial-ökologische Forschung BMBF für die finanzielle Unterstützung im Projekt ‘Environ - Entwicklung und Evaluation einer Intervention zur Vermeidung von durch energetische Sanierung ausgelösten Rebound-Effekten’.

Umweltinformatik – Alles Geschmackssache?

Hans-Knud Arndt[1]

Abstract: Der Beitrag befasst sich mit dem Einfluss des kurzfristigen (Massen-) Geschmacks auf die Nachhaltigkeit der Informations- und Kommunikationstechnik (IKT). Konzepte wie das Usability Engineering Prozessmodell setzen u.a. auf das Gesetz der großen Zahlen in der Statistik, um so den Geschmack und damit eine geeignete Usability der Nutzer zu treffen. Es wird aber an ausgewählten Anwendungsbeispielen aufgezeigt, dass diese Vorgehensweise zu Problemen und in der Regel auch nicht zu einer Nachhaltigkeit in der IKT führt.

Keywords: Nachhaltigkeit; Informations- und Kommunikationstechnik; Geschmack; Ästhetik; Design; Usability

1 Geschmack, Ästhetik, Design und Usability

Der Begriff „Geschmack“ hat vielfältige Bedeutungen und ist damit auch einem Kontext zu den Begriffen „Ästhetik“ und „Usability“ zu sehen. Auf der einen Seite steht „Geschmack“ für „etwas schmecken“ (diese Pizza schmeckt sehr gut) oder auch für den „Geschmacksinn“ beim Essen (heute bin ich erkältet und habe überhaupt keinen Geschmack). Auf der anderen Seite wird „Geschmack“ auch als „Fähigkeit zu ästhetischem Werturteil“ verstanden, objektiv wie subjektiv [Dud20]. D. h., Geschmack ist damit die Grundlage, die Ästhetik von Gegenständen, Gebäuden und auch von Informations- und Kommunikationstechnik (IKT) einschätzen zu können. In einem engen Zusammenhang zum Geschmack steht der Begriff der Urteilskraft. Der Philosoph Hans-Georg Gadamer versteht den Begriff „Urteilskraft“ als den „gesunden Menschenverstand“, der einen „Dummkopf“ von einem „klugen Menschen“ unterscheidet. Dabei gilt, dass sich Urteilskraft nicht erlernen lässt, sie kann sich allenfalls entwickeln. Und vom grundsätzlichen Urteilsvermögen ist wiederum auch der Geschmack abhängig [Get20]:

„Guter Geschmack ist sich seines Urteils stets sicher, d. h. er ist seinem Wesen nach sicherer Geschmack, ein Annehmen und Verwerfen, das kein Schwanken, Schielen nach dem Anderen und kein Suchen nach Gründen kennt.“[Gad10]

Philosophisch gesehen kann durch Geschmack und Urteilskraft „Ästhetik“ bestimmt werden. Ästhetik bezeichnet die „Wissenschaft, die allg. Probleme der Kunst und i. e. S. des Schönen (Erhabenen, Häßlichen, Tragischen, Komischen usw.) behandelt. Sie

[1] Otto-von-Guericke Universität Magdeburg, Fakultät für Informatik, Arbeitsgruppe Wirtschaftsinformatik – Managementinformationssysteme, hans-knud.arndt@iti.cs.uni-magdeburg.de

doi:10.18420/inf2020_25

untersucht erkenntnistheoretisch, teils mit empir. Methoden, zum einen die Bedingungen der Konstruktion von Kunstwerken, die Strukturen des ästhet. Gegenstandes in Kunst und Natur, das Verhältnis von Kunst und Wirklichkeit, zum anderen die Bedingungen und Formen ästhet. Rezeption durch den einzelnen wie durch die Gesellschaft. (...) Daß auf dem Gebiet der Ä[sthetik] eine systematisch aufgebaute, begründete Theorie möglich sei, wird vielfach bestritten“[Mey93]. Fragen der Ästhetik sind in jedem Fall aber ein wichtiger Erfolgsfaktor für Produkte jeglicher Art, also auch für digitale Produkte [Wil19]. Und Ästhetik ist gerade bei der IKT in einem engen Zusammenhang mit dem (visuellen) Design zu sehen.

„Der Begriff Design bezeichnet die Gestaltung von Gegenständen aller Art nach den Kriterien von Funktionalität (z.B. (..) Ergonomie) und Ästhetik.“[HSZ96a] Der deutsche Industriedesigner Dieter Rams hat sich lange mit der Frage eines guten Designs beschäftigt und aus seiner langjährigen praktischen Erfahrung „Orientierungs- und Verständnishilfen“[Ram96] für ein gutes Design formuliert. Analog zur Design-Definition sieht Dieter Rams Ästhetik als wesentliches Kriterium eines guten Designs an. Darüber hinaus präzisiert er (zunächst für das Industriedesign) die (spezielle) Forderung nach Ergonomie – „Ergonomie beschäftigt sich als Teilwissenschaft der Arbeitswissenschaften mit der Anpassung der Arbeit an den Menschen“[HSZ96b] – hin zu (allgemeineren) Forderungen nach Gebrauchstauglichkeit von Produkten [Ram96].

In der IKT gibt es ebenfalls eine lange Diskussion über Fragen der Gebrauchstauglichkeit, die auch ihren Niederschlag in der internationalen Normung gefunden hat. Gebrauchstauglichkeit (englisch usability) wird im Kontext der IKT sehr abstrakt betrachtet und bezeichnet nach DIN EN ISO 9241-11 das „Ausmaß, in dem ein Produkt durch bestimmte Benutzer in einem bestimmten Nutzungskontext genutzt werden kann, um bestimmte Ziele effektiv, effizient und zufriedenstellend zu erreichen“[DIN98]. Die Normungsinstitution versteht in diesem Zusammenhang unter [DIN98]:

- Nutzungskontext: „Die Benutzer, Arbeitsaufgaben, Arbeitsmittel (Hardware, Software und Materialien) sowie die physische und soziale Umgebung, in der das Produkt genutzt wird.“
- Effektivität: „Die Genauigkeit und Vollständigkeit, mit der Benutzer ein bestimmtes Ziel erreichen.“
- Effizienz: „Der im Verhältnis zur Genauigkeit und Vollständigkeit eingesetzte Aufwand, mit dem Benutzer ein bestimmtes Ziel erreichen.“

Aufgrund der guten Erfahrungen mit dem Software Engineering wird bezogen auf die Gebrauchstauglichkeit häufig von einem Usability Engineering gesprochen, „einen parallel zur klassischen Software-Entwicklung laufenden Prozess, der auch eng mit diesem verzahnt sein sollte.“[SaB16]

2 Suche nach dem guten Geschmack: Usability Engineering

Die Qualität eines Entwurfs ist der Schlüssel zu mehr Umweltfreundlichkeit und Nachhaltigkeit von Produkten. Dies gilt auch für umweltfreundliche bzw. nachhaltige IKT-Produkte [Arn13]. Unter einem „Usability Engineering“ wird „der methodische Weg zur Erzeugung der Eigenschaft Usability [verstanden]. Es ist ein Teilprozess der Entwicklung und Gestaltung technischer Systeme und ergänzt das klassische Engineering, beispielweise Software-Engineering, um ergonomische Perspektiven.“[SaB16]

Das Prozessmodell „Usability Engineering“ (siehe auch Abbildung 1) kann aus folgenden Phasen bestehen [SaB16]:

- Analysephase (1): In dieser Phase sollen die Rahmenbedingungen des zukünftigen Systems erfasst und Anforderungen an das zukünftige System abgeleitet werden.
- Konzeptphase (2): Aufbauend auf den Ergebnissen der Analysephase soll in dieser Phase über die Funktionalität des neuen Systems entschieden sowie das Konzept für das neue System entwickelt werden. Die Grundlage bilden dabei die Arbeitsgestaltung und die Prozessdefinition, welche noch unabhängig vom neu zu gestaltendem System sind.
- Entwicklungsphase (3): In dieser Phase wird das in der 2. Phase entwickelte Konzept umgesetzt. Ausgangspunkt sind dazu in der Regel Prototypen. Erweisen sich einzelne Prototypen als positiv, werden diese Ansätze im eigentlichen neu zu gestaltendem System umgesetzt (Systemintegration).
- Einführungsphase (4): In dieser Phase wird das neue System zunächst in ausgewählten Bereichen pilothaft eingeführt. Nach ggf. erforderlichen Änderungen (Arbeitsgestaltungsmaßnahmen) kann dann die Einführung des neuen Systems in der Fläche erfolgen.
- Evaluation: Die sach- und fachgerechte Bewertung hat einen zentralen Stellenwert im Usability Engineering, stellt aber keine eigenständige Phase dar, sondern ist vielmehr als begleitende Aktivität aufzufassen.
- Projektplanung und -management: Da die Entwicklung eines neuen Systems eine hochgradig interdisziplinäre Aufgabe darstellt, bedarf es eines entsprechenden Projektmanagements.

Die Schwierigkeiten einer solchen ingenieurtechnischen Vorgehensweise – gerade auch im Hinblick auf eine nachhaltige Gestaltung – lassen sich beispielhaft am Office-Paket „Microsoft Office 2007“ aufzeigen, welches am 30. Januar 2007 vorgestellt wurde. Eine wesentliche Neuheit dieser Version war, die klassische Menüstruktur und die üblichen Symbolleisten, die die vorangegangenen Office-Versionen kennzeichneten, durch die neue Multifunktionsleiste „Ribbon“ zu ersetzen. Dazu hatte der für die Gestaltung dieser Oberfläche verantwortliche Microsoft Mitarbeiter Jensen Harris auf verschiedenen Präsentationen

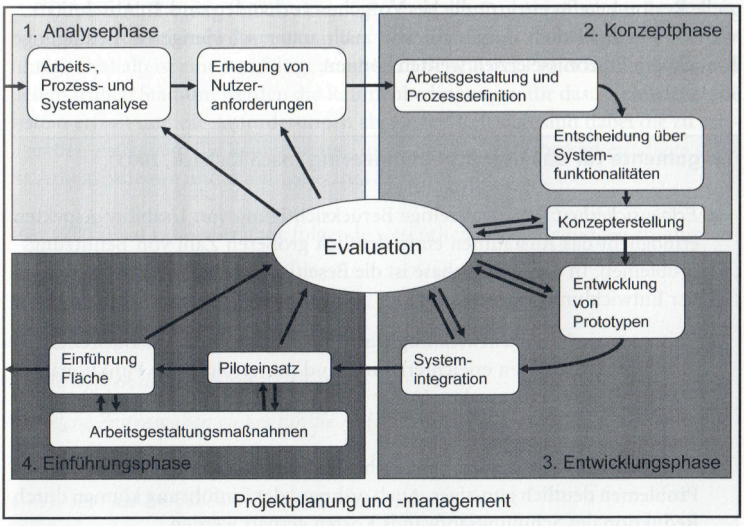


Abb. 1: Prozessmodell „Usability Engineering“ von Sarodnick und Brau (Quelle: [SaB16])

zum Office-Paket „Microsoft Office 2007“ erläutert, welche „zahlreiche Prototypen und (...) viele Experimente die Microsoft unternahm, um schlussendlich mit einer neuen Benutzeroberfläche für Office aufwarten zu können.“[WiF08] In der Praxis zeigte sich dann aber, dass die „neue Benutzeroberfläche ‚Ribbon‘ (..) gemischte Reaktionen aus[löst], das Design wird geliebt und gehasst. Trotz ausführlicher Betatests erscheint der Release inkonsequent“. [HDi19] Schlussendlich wurde bezogen auf die Multifunktionsleiste „Ribbon“ und die neue Benutzeroberfläche von „Microsoft Office 2007“ folgende Meinung vertreten: „Wer umsteigt, wird sich damit am Anfang schwertun. Für Computer-Anfänger soll das neue Konzept hingegen einfacher sein.“[SWi07]

Zusammenfassend lässt sich damit festhalten, dass eine ingenieurtechnische Vorgehensweise wie beim Usability Engineering nicht zwangsläufig zu einem allgemeingültig guten Geschmack und damit zu einer guten Ästhetik und Usability führt. Und diese Schwierigkeiten haben wiederum einen großen Einfluss auf die Nachhaltigkeit der IKT-Produkte. Die Süddeutsche Zeitung spricht deshalb auch 2011 von „30 Jahre Microsoft-Horror“ und davon, dass Microsoft „für Effizienz [stand], das immer, aber eben nicht für Eleganz und für das, was man mit Ease-of-Use bezeichnet. Microsoft hatte eher das Image seines eigenen Wortungetüms: Es war ein unerwartet schwerer Ausnahmefehler, was Design und Anwender-Erfahrung angeht.“[Gra11]

Die Komplexität der Beziehung „Geschmack“ und „Nachhaltiges Design“ soll deshalb in den folgenden zwei Anwendungsbeispielen „Apple Airpods“ und „Apple Watch“ näher betrachtet werden.

3 Anwendungsbeispiele

3.1 Anwendungsbeispiel „Apple AirPods“

Das Unternehmen Apple traf mit der Vorstellung des ersten iPhone am 9. Januar 2007 traf sowohl in technischer als auch in gestalterischer Hinsicht nahezu uneingeschränkt den Geschmack der Menschen und „veränderte (..) quasi über Nacht unser Verständnis und unsere Erwartung bezüglich Smartphones“ [Zec11]. Als das Unternehmen Apple dagegen am 7. September 2016 die Kopfhörer „AirPods“ vorstellte, war die Zustimmung in geschmacklicher Hinsicht nicht so eindeutig gegeben.

Als „AirPods“ bezeichnet das Unternehmen Apples die ersten kabellosen („wireless“) Gehörgang-(„In-Ear“)-Kopfhörer unter eigenem Markennamen (vorher wurden nur die kabellosen Kopfhörer von Apple‘s Tochtergesellschaft „Beats Electronics“ angeboten). Das Unternehmen Apple unternahm damit einen weiteren Schritt in Richtung der Strategie einer kabellosen Zukunft digitaler Endgeräte, denn erstmals wurde auch bei dem Smartphone „iPhone“ mit der Versionsnummer 7, welches gleichzeitig mit den AirPods vorgestellt wurde, der Kopfhöreranschluss weggelassen. Das Design der AirPods nimmt Bezug auf die (2012 von Apple eingeführten) Gehörgang-Kopfhörer „EarPods“. Zur Umsetzung einer kabellosen Verbindung wird der Industriestandard „Bluetooth“ eingesetzt. Die AirPods verfügen über einen W1-Chip und können deshalb weitestgehend automatisiert mit einem iPhone, dem iPad, den Mac-Computern, der Apple Watch und dem Apple TV bei entsprechend dafür vorgesehenen Betriebssystemversionen verbunden werden. Bei anderen Betriebssystemen, wie zum Beispiel Android, können die AirPods wie übliche (Bluetooth-)Funkkopfhörer eingesetzt werden. Darüber hinaus ist in den Gehörgang-Kopfhörern jeweils ein Mikrofon integriert, damit mit diesen Kopfhörern auch kommuniziert werden kann. Die beiden Akkus in den Kopfhörern sollen einen Einsatz von über 5 Stunden ermöglichen, bevor ein Wiederaufladen im mit ausgelieferten „Ladecase“ notwendig ist. Das Ladecase wiederum kann die Kopfhörer mit Strom für zusätzliche 24 Stunden Akkulaufzeit versorgen und verfügt über ein Nahfeldkommunikation-(Near Field Communication, NFC)-Modul, welches vor allem für die automatisierte Kommunikation mit den digitalen Apple-Endgeräten genutzt wird. Die Aufladung des Ladecase erfolgt über einen Lightning-Anschluss. Werden die AirPods aus dem Ladecase genommen, verbinden sich die Kopfhörer automatisch mit dem in Reichweite befindlichen und autorisierten digitalen Endgerät. Der Entwurf der AirPods umfasst einige optische Sensoren, damit u.a. automatisch erkannt werden kann, ob die AirPods sich gerade in Nutzung befinden: Wird z.B. einer der Kopfhörer aus dem Ohr genommen, wird die Wiedergabe von Musik bzw. Videos auf dem digitalen Apple-Endgerät angehalten, werden beide Kopfhörer den Ohren entnommen, wird die Wiedergabe beendet. Bei der ersten Generation der AirPods kann durch ein Doppeltippen auf einen der beiden Kopfhörer wahlweise die Spracherkennungssoftware Siri von Apple aktiviert, die Wiedergabe gestoppt oder der nächste Musiktitel ausgewählt werden [MaL20].

Während der technische Fortschritt der Apple „AirPods“ bereits bei der ersten Präsentation allgemein anerkannt wurde, traf das Design der AirPods (zunächst) nicht den Geschmack

der Allgemeinheit: Die Vorstellung der ersten Apple AirPods „sorgte im Netz aber schnell für viel Häme und Spott“ [Dre16]. Auch zwei Jahre nach der ersten Präsentation der AirPods wird festgehalten, dass für dieses Apple-Produktdesign gilt: „Kein Produkt hat seit seiner Vorstellung im September 2016 für derart viel Häme und Spott gesorgt wie Apples Drahtlos-Kopfhörer Airpods. Bilder von Zahnbürsten und Tampons im Ohr machten schnell die Runde.“[Ste18] Gerade der Vergleich der AirPods mit den Braun Oral-B Aufsteckzahnbürsten für elektrische Zahnbürsten (siehe Abbildung 2) ist bis heute üblich.

Abb. 2: Braun Oral-B Aufsteckzahnbürsten und Apple AirPods 1. Generation (links)/Gags mit Köpfen von elektrischen Zahnbürsten (rechts)
(Quellen: Eigene/[Dre16])

Im Sinne des oben aufgezeigten Usability Engineering-Prozessmodells hätte also diese Form, dieses Design der kabellosen Gehörgang-Kopfhörer niemals zur Marktreife gebracht werden dürfen. Denn die Reaktionen aus dem Internet sind stellvertretend dafür zu sehen, dass die pilothafte Einführung in ausgewählten Bereichen (Phase 4) zwangsläufig zu Änderungen im Design (das Design war ja schließlich beim Kunden mehrheitlich durchgefallen) hätte führen müssen: „Mit den AirPods verkauft Apple erstmals eigene Bluetooth-Kopfhörer. Die klingen nicht schlecht, sind teilweise sogar richtig clever – und trotzdem nichts für die Masse“[Küh16].

Auf lange Sicht aber lagen die (kurzfristigen) Geschmacksbekundungen des Internets daneben und die ersten Apple AirPods erwiesen sich nicht nur als zukunftsweisendes Produkt in Technik *und* Design, sondern auch als stilprägend für Konkurrenzmodelle von kabellosen Gehörgang-Kopfhörern:

> „Rückblickend betrachtet war es ein historischer Fehler, sich derart lustig gemacht zu haben über die sogenannten Airpods, über diese winzigen kabellosen Kopfhörer, die Apple 2016 auf den Markt geschmissen hatte. All der Hohn und Spott, die Memes und Tweets und das Gelächter über den Unsinn einer solchen Erfindung hatten am Ende nichts gebracht. Vor allem in den Großstädten dieser Welt stecken sie nun in den Ohren einer immer rasanter anwachsenden Zahl von Menschen.
> Dabei war es gleichermaßen vorschnell wie nachvollziehbar, über die so

> seltsamen Geräte erst einmal zu lachen, ihnen haftete etwas Defizitäres an, etwas Entstelltes, als hätte man das Auto der Zukunft entwickelt und die Räder vergessen. Befeuert wurde diese Wahrnehmung dadurch, dass die neuen Kopfhörer im Vergleich zu ihrem Vorgängermodell in puncto Design so gut wie unverändert geblieben waren, es fehlte nun lediglich das Kabel, als hätte es jemand Ungeschicktes aus Versehen abgeschnitten und behauptet, das müsse so sein.“ [Kal19]

Dass sich das ursprüngliche und trotz der negativen Rückkoppelung bezüglich des (anfänglich) artikulierten Geschmacks unveränderte Design der AirPods als sehr nachhaltig erwiesen hat, zeigt auch der Erfolg am Markt: „Apple dürfte 2019 mehr als 60 Millionen Stück von seinen Airpods verkaufen. Dies berichtet Bloomberg in einem aktuellen Artikel in Berufung auf interne Quellen bei dem Hardwarehersteller. Sollte sich diese Prognose bewahrheiten, würde Apple damit nicht nur die Absätze des Vorjahrs verdoppeln, sondern auch einen Marktanteil von 50 Prozent bei echten drahtlosen Kopfhörern erreichen.“[DSt19].

3.2 Anwendungsbeispiel „Apple Watch“

Das Produkt „Apple Watch“ wurde in der ersten Generation am 9. September 2014 als ein „unglaublich präziser Zeitmesser, ein persönliches und direktes Kommunikationsgerät und ein bahnbrechender Begleiter für Gesundheit und Fitness“ [App15a] vorgestellt (siehe auch Abbildung 3). Technische Merkmale der Apple Watch sind u.a. eine digitale Krone (Digital Crown, ein innovativen Weg „zum flüssigen Scrollen, Zoomen und Navigieren ohne das Display zu versperren“ [App15a]), ein Retina Display mit Force Touch („Force Touch nutzt winzige Elektroden rund um das flexible Retina Display, die zwischen leichtem Tippen und stärkerem Drücken unterscheiden können und sofortigen Zugriff auf kontextspezifische Steuerungen erlauben“ [App15b]) sowie die Taptic Engine (um eine haptische Rückmeldung zu geben - die Apple Watch tippt dem jeweiligen Träger immer auf das Handgelenk, wenn er eine Benachrichtigung oder Mitteilung erhält oder er auf das Display drückt [App15b]).

Auch die Apple Watch schien zunächst nicht den Geschmack der Allgemeinheit zu treffen. Dies vor allem im Hinblick auf die Technik. Es wurde von einem „Flop“[bag20, Web20] gesprochen: „Als die Apple Watch dann 2015 vorgestellt wurde, waren die Ansprüche entsprechend hoch. Es folgte schnell die Enttäuschung, denn die Uhr konnte in ihrer Version den Vorstellungen der Kunden nicht gerecht werden.“ [bag20] „Auch die Schweizer Uhrenindustrie brachte viele kluge Argumente gegen die Neuerscheinung vor: Ihr Akku müsse jeden Tag geladen werden, sie funktioniere nur mit einem Apple-Smartphone, und sie sei kein langlebiges Produkt“[Spe20].

Gleichfalls für die Apple Watch gilt, dass sich das grundsätzliche und trotz der negativen Rückkoppelung bezüglich des (anfänglich) artikulierten Geschmacks unveränderte Design der Apple Watch ebenfalls als sehr nachhaltig erwiesen hat. Dies spiegelt sich auch im Erfolg am Markt wider:

Abb. 3: Apple Watch 1. Generation
(Quelle: [App15c])

> „Wie erfolgreich Apple mit seiner Uhr ist, wird offiziell nicht kommuniziert. Auch die Verkaufszahlen bleiben geheim. Folgt man den Analysten von Strategy Analytics, hat Apple im vergangenen Jahr 31 Millionen Geräte verkauft, ein Drittel mehr als im Jahr 2018. Die gesamte Schweizer Uhrenindustrie kommt 2019 zusammen auf 21 Millionen Uhren, ein Rückgang von 13 Prozent im Vergleich mit 2018. Auch wenn der Vergleich zwischen einer Smartwatch und einer traditionellen Luxusuhr hinkt, scheint Apple in diesem Segment doch einiges richtig zu machen.“[Spe20]

Zudem gilt für das Segment der Smartwatches, dass es dem „Android-Lager (. . .) unterdessen nicht [gelang] - ganz anders als bei Smartphones - mit dem Plattform-Ansatz Apple unter Druck zu bringen. Obwohl das Google-Betriebssystem den Smartphone-Absatz mit einem Anteil von mehr als 80 Prozent dominiert und auch viele Anbieter von Modeuhren Android-Modelle im Angebot haben, hielt Apple nach IDC Berechnungen im vergangenen Jahr den Spitzenplatz mit rund 29 Prozent Marktanteil.“[bag20]

Bereits durch diese kurze Analyse der beiden Anwendungsfälle von Apple einschließlich des Beispiels des Office-Pakets „Microsoft Office 2007“ lässt sich festhalten, dass der Indikator eines aktuellen, kurzfristigen Massengeschmacks wenig geeignet ist, nachhaltige IKT zu entwerfen und langfristig am Markt unterzubringen. Aber genau den Einsatz von u.a. diesem Indikator fördert das Usability Engineering-Prozessmodell, denn es setzt auf das Gesetz der großen Zahlen in der Statistik. Um eine Nachhaltigkeit in dem Entwurf von IKT zu erreichen, sollte deshalb der Erkenntnis des Industriedesigners Dieter Rams gefolgt werden: Es „geht bei ästhetischer Qualität um Nuancen, um feine Abstufungen, um den Gleichklang und das subtile Gleichgewicht einer Vielzahl von visuellen Elementen. Man braucht ein Auge, das durch jahrelange Erfahrung geschult ist, um hier ein fundiertes Urteil zu haben“[Ram96], es zählt also das Können und die Kompetenz des/der einzelnen Designer und nicht der kurzfristige Geschmack einer breiten Masse.

4 Zusammenfassung und Ausblick

In der Philosophie wird die Ansicht vertreten, dass durch Geschmack und Urteilskraft „Ästhetik“ bestimmt wird. Und Ästhetik wiederum ist eine der Voraussetzungen sowohl für die Usability als auch für die Nachhaltigkeit von IKT. In der Informatik wird häufig eine ingenieurtechnische Vorgehensweise wie beim Usability Engineering vorgeschlagen. Aber eine solche Vorgehensweise führt nicht zwangsläufig zu einem allgemeingültig guten Geschmack, führt nicht zwangsläufig zu einer guten Ästhetik und Usability und führt schließlich auch nicht zwangsläufig zu einer Nachhaltigkeit von IKT. Es zeigt sich, dass weiterhin das Können und die Kompetenz des/der einzelnen Designer notwendig ist. Damit stellt sich auch ganz konkret die Frage einer Automatisierbarkeit des nachhaltigen Entwurfs von IKT z.B. durch Konzepte der Künstlichen Intelligenz.

Literaturverzeichnis

[App15a] Apple Inc.: Apple Watch ab 24. April in neun Ländern verfügbar, Apple Presseinformation 09. März 2015, http://www.apple.com/de/pr/library/2015/03/09Apple-Watch-Available-in-Nine-Countries-on-April-24.html [2015-04-20].

[App15b] Apple Inc.: Apple Watch Technologie, 2015 https://www.apple.com/de/ watch/technology/ [2015-04-20].

[App15c] Apple Inc.: Designing for Apple Watch, 2015-03-09, https://developer.apple.com/library/prerelease/ios/documentation/UserExperience/Conceptual/WatchHumanInterfaceGuidelines/index.html#//apple_ref/doc/uid/TP40014992-CH3-SW1 [2015-04-20].

[Arn13] Arndt, H.-K.: Umweltinformatik und Design - Eine relevante Fragestellung? In: Horbach, M. (Hrsg.): INFORMATIK 2013: Informatik angepasst an Mensch, Organisation und Umwelt (16.–20. September 2013, Koblenz, Germany), GI-Edition-Lecture Notes in Informatics (LNI), P-220, Gesellschaft für Informatik e.V., Bonn, 2013, S. 931-939.

[bag20] bagre: Jubiläum: Fünf Jahre Apple Watch: Vom Flop zum Marktführer, Die Presse, 23.04.2020, https://www.diepresse.com/5804411/funf-jahre-apple-watch-vom-flop-zum-marktfuhrer [2015-04-20].

[DIN98] Deutsches Normungsinstitut e.V. (Hrsg.): Ergonomische Anforderungen für Bürotätigkeiten mit Bildschirmgeräten, Teil 11: Anforderungen an die Gebrauchstauglichkeit — Leitsätze, (ISO 9241 -11 :1998) Deutsche Fassung EN ISO 9241 -11 :1998, Beuth Verlag, Berlin/Wien/Zürich, Januar 1999.

[Dre16] Drees, C.: Apple AirPods: So hämisch reagiert das Netz, mobilegeeks.de, 08.09.2016, https://www.mobilegeeks.de/artikel/apple-airpods-so-haemisch-reagiert-das-netz/ [2020-05-15].

[DSt19] Der Standard (Hrsg.): IT-BUSINESS: Airpods: Apples "Zahnbürsten-aufsatz" verkauft sich wie wahnsinnig, derstandard.de, 22.11.2019, https://www.derstandard.de/story/2000111382580/apples-airpods-der-zahnbuerstenaufsatz-verkauft-sich-wie-wahnsinnig [2020-05-15].

[Dud20] Dudenredaktion (Hrsg.): Geschmack, Bibliographisches Institut, Berlin, https://www.duden.de/rechtschreibung/Geschmack [2020-05-15].

[Gad10] Gadamer, Hans-Georg: Wahrheit und Methode: Grundzüge einer philosophischen Hermeneutik, Verlag Mohr Siebeck, Tübingen, 2010, S. 42.

[Get20] getabstract.com (Hrsg.): Zusammenfassung von Wahrheit und Methode: Hans-Georg Gadamer, https://www.getabstract.com/de/zusammenfassung/wahrheit-und-methode/6677 [2020-05-15].

[Gra11] Graff, B.: Microsoft entdeckt die Kraft des Designs, Süddeutsche Zeitung, 29.12.2011, https://www.sueddeutsche.de/digital/it-konzern-im-umbruch-microsoft-entdeckt-die-kraft-des-designs-1.1246227 [2020-05-15].

[HDi19] Hülsbömer, S., Dirscherl, H.-C.: Die Geschichte von Microsoft Office: Word, Excel, Access, PC Welt, 05.10.2019, https://www.pcwelt.de/ratgeber/Microsoft-Office-Mit-einer-Maus-fing-alles-an-6091613.html [2020-05-15].

[HSZ96a] Heider, T., Stegmann, M., Zey, R.: Design, Lexikon Internationales Design: Designer, Produkte, Firmen, Rowohlt Taschenbuch, Rowohlt Verlag, Reinbek bei Hamburg, 1996, S. 88.

[HSZ96b] Heider, T., Stegmann, M., Zey, R.: Ergonomie, Lexikon Internationales Design: Designer, Produkte, Firmen, Rowohlt Taschenbuch, Rowohlt Verlag, Reinbek bei Hamburg, 1996, S. 101.

[Kal19] Kaleyta, T. K.: Airpods und ihre Äquivalente: Ich höre doch zu, faz.net, 24.07.2019, https://www.faz.net/aktuell/feuilleton/airpods-von-apple-haben-laecherlichen-ruf-ueberstanden-16294886.html [2020-05-15].

[Küh16] Kühl, E.: Apple AirPods: Besser als 'ne Zahnbürste im Ohr, Zeit Online, 30.12.2016, https://www.zeit.de/digital/mobil/2016-12/apple-airpods-kopfhoerer-test [2020-05-15].

[MaL20] MacLife (Hrsg.): Bluetooth-Koprhörer von Apple: Technik, Infos & Fakten: Apple AirPods: Alles Wissenswerte zu den kabellosen Apple-Kopfhörern für iPhone & Co., MacLife, https://www.maclife.de/thema/airpods [2020-05-15].

[Mae17] Maehner, J.: Apple AirPods günstig: Kult-Ohrhörer zum kleinsten Preis - Drahtlos glücklich, chip.de, 15.08.2017, https://www.chip.de/artikel/Apple-AirPods-guenstig-Kultige-Wireless-Kopfhoerer-zum-besten-Preis_117345015.html [2020-05-15].

[Mey93] Meyers Lexikonredaktion (Hrsg.): Ästhetik, Meyers neues Lexikon in 10 Bänden, Bd. 1 A – Ben, Meyers Lexikonverlag, Mannheim/Leipzig/München/Zürich, 1993, S. 339.

[Ram96] Rams, D.: Zehn Thesen zum Design. In: Dieter Rams (Hrsg.): Weniger, aber besser – Less, but better, 5. Aufl., Jo Klatt Desig+Design Verlag, Hamburg, 2016, S. 6–7.

[SaB16] Sardonick, Florian/Brau, Henning: Methoden der Usability Evaluation: Wissenschaftliche Grundlagen und praktische Anwendungen, 3. Aufl., Verlag Hofgrebe, Bern, 2016, S. 93ff.

[Spe20] Spehr, M.: Fünf Jahre Apple Watch: Mit dem Design der Eleganz, Frankfurter Allgemeine Zeitung, 24.04.2020, https://www.faz.net/aktuell/technik-motor/digital/fuenf-jahre-apple-watch-mit-dem-design-der-eleganz-16738289.html [2020-05-15].

[Ste18] Stepanek, M.: Apple AirPods: Apple Airpods im Test: Hässlich, aber gut, futurezone.at, 05.08.2018, https://futurezone.at/produkte/apple-airpods-im-test-haesslich-aber-gut/400070837 [2020-05-15].

[SWi07] Spehr, M., Wiseman, R.: Software: Mit Office 2007 wird alles anders, Frankfurter Allgemeine Zeitung, 08.01.2007, https://www.faz.net/aktuell/technik-motor/digital/software-mit-office-2007-wird-alles-anders-1410805.html [2020-05-15].

[Web20] Weber, V.: Fünf Jahre Apple Watch: Wie Apples Smartwatch vom Flop zum Klassiker wurde, spiegel.de, 22.04.2020, https://www.spiegel.de/netzwelt/gadgets/wie-die-apple-watch-vom-flop-zum-klassiker-wurde-a-9f13b205-cb0f-4e44-9085-3c6341a03661 [2020-05-15].

[WiF08] WinFuture (Hrsg.): Office 2007: Der lange Weg zur neuen Oberfläche, 12.03.2008, https://winfuture.de/news,38052.html4 [2020-05-15].

[Wil19] Wilhelm, T.: Visuelle Designer sind Künstler und Handwerker, Nutzerbrille 18.10.2019, https://www.nutzerbrille.de/visuelles-design/ [2020-05-15].

[Zec11] Zec, P.: „All New Design“: Das Geheimnis eines magischen Unternehmenserfolgs. In: Schulze, S./Grätz, I. (Hrsg.) Apple Design, Publikation zur Ausstellung „Stylectrical – Von Elektrodesign, das Geschichte schreibt“, Museum für Kunst und Gewerbe Hamburg (MK&G) 26. August 2011 – 15. Januar 2012, Hantje Cantz Verlag, Ostfildern, 2011, S. 88–103.

5th GI/ACM I4.0 Standardization Workshop on Industrial Automation and Control Systems

The 5th GI/ACM Workshop 2020 Scope and Draft Programme on Standardization of Secure and Safe Smart Manufacturing Systems with respect to IEC 62443 IACS

Jan deMeer[1] Karl Waedt[2] Axel Rennoch[3] Hans-Joachim Hof[4]

Abstract: The 5th GI/ACM Workshop Programme on Standardization of Secure and Safe Production within Industrial Automation and Control Systems (IACS) took place virtually at September 28, 2020 at the Karlsruhe Institute of Technology (KIT) that hosted the 50th GI's yearly assembly (GI Informatik 2020 Jahrestagung): https://informatik2020.de/programm/workshops/

Keywords: I4.0; Security & Safety; Industrial Automation and Control Systems; Digital Twin; Production Ontologies; Smart Manufacturing; Asset Administration Shell; OPC-UA; AutomationML; OT/IT Security; Syntactic and Semantic Interoperability; Edge Computing

1 Organization of the IACS Standardization Workshop 2020

The IACS workshop call has been closed August 31 2020 due and 10 up-to-date papers from the realms of Standardization, Best Practice and I4.0 Research have been presented. The preparation of the IACS Workshop has been achieved by means of the platform: https://easychair.org/conferences/?conf=i40acsws20 in co-operation with the IACS Standardization WS Programme Committee. All accepted contributions (with the exception of abstracts) that coincide to Springer Publisher's author guidelines of LNI: https://gi.de/service/publication/ini/ are listed below and are ready for publication in the Conference Proceedings of the GI Informatik Jahrestagung 2020.

The PC decided on the WS' Programme that comprises the following presenters and IACS topics. Notice the 1st (‚IACS Scope and Semantics') and the 10th (‚Quo Vadis IACS') presentation embrace the WS' Programme of Work in an introducing and concluding part:

1. *Jan deMeer et al.: Introduction into IEC 62443 IACS Scope and Semantics*
2. *Vitaly Promyslov et al.: Validation of Control Systems with Heterogeneous Digital Models and Virtualization Technologies*

[1] smartspacelab.eu GmbH, Berlin, 12205, demeer@smartspacelab.de
[2] FRAMATOME GmbH, Erlangen, 91058, karl.waedt@framatome.com
[3] Fraunhofer FOKUS, Berlin, 10589, axel.rennoch@fokus.fraunhofer.de
[4] Technische Hochschule Ingolstadt, hans-joachim.hof@thi.de

doi:10.18420/inf2020_26

3. *Mithil Parekh et al.: Aligning with Cyber Security Framework by modeling OT Security*
4. *Yuan Gao et al.: Operational Security Analysis and Challenge for IoT Solutions*
5. *Axel Rennoch et al.: Edge Computing Standardization and Initiatives*
6. *Vanesa Watson et al.: MAC-layer Security for Time-sensitive Switched Ethernets*
7. *Joseph Schindler et al.: Gossip Protocol Approach for a Decentralized Energy Market with OPC-UA Client Server Communication*
8. *Nikolas Mühlbauer et al.: Feature-based Comparison of Open Source OPC-UA Implementations*
9. *Deeksha Gupta et al.: Simulation Model for Threat and Impact Analysis on Modern Electrical Power Systems*
10. *Pierre Kobes: Quo Vadis IEC 62443 IACS?*

The Board of GI/ACM WS Co-Chairs and the WS Programme Committee appreciated the technical support from:

2 Scope of IACS Standardization

The scope of IACS Standardization and thus of the Workshop included but is not limited to the full bandwidth of the current 13 parts of the IEC 62443 IACS Standards series, i.e.:

1. IACS Modeling and Concepts
2. System Security Conformance Metrics
3. Security Lifecycle and Use Cases
4. Patch and Security Management Systems
5. IACS Security Risk Assessment and Security Levels
6. Product Development
7. Security Requirements for IACS Components,

which includes requirements to security und safety measures to be applied during the full IEC62443 IACS live cycle comprising:

1. Rules and procedures for operation and maintenance of IACS
2. Planning and installation of Basic Process Control Systems (BPCS), Safety Instrumented Systems (SIS), other hard and software of IACS
3. Development and implementation of IACS components comprising
    1. embedded devices
    2. network components
    3. host devices
    4. applications

Additionally Security and Safety Requirements of IEC 62443 production sites and devices comprise the following measures and concepts:

1. to identify security contexts in order to condemn threats;
2. to identify security aims in order to enable automatization of production plants with the base line of ‚Safety first';
3. to identify production safety based on minimal right restrictions without influencing IACS availability too much;
4. to organize staggered defense measures in order to harden the IACS assets against attacks;
5. to perform Risk Analysis in order to evaluate Threats and Vulnerabilities of Assets;
6. to identify Guidelines and Processes of enterprises for the purpose of a holistic view on the enterprise's and plant's security;
7. to invent concepts of security into the whole chain of production in order to avoid vectors of threats and vulnerabilities.

3 Semantics of IACS Standardization

3.1 Complexity in Standardization

Industrial Standards such as the multi-part IEC 62443 standard on Security of Industrial Automation and Control Systems that is currently elaborated by a couple of standardization committees such as ISA99, ISO JTC1/SC27/WG4, IEC TC65/WG23 etc. become more and

more complex in the sense of yielding a common understanding with respect to a unique interpretation, e.g. for an implementation of a production system to be conform to the given set of complex standards.

3.2 Semantic Classification

With respect to the issue of yielding a common interpretation of standardization texts the so-called ‚System Committee on Smart Manufacturing (SyC SM)' has started the task force ‚ISO/IEC Joint Smart Manufacturing Standards Map (TF SM2)' to solve the issue of a common understanding by inventing the methodology of classification supported by a platform of integrated tools comprising visualization and a central repository of classified text passages.

The process of SM Standards Classification comprises three steps:

to collect formats and characteristics of products or of processes of production prescribed and constrained in related standards (Notice: In future this step needs to be supported by a SM2 Vocabulary that is do-day not available);

to actualize parameters of SM standards by assigning specific values to the characteristics of standards identified in the SM2 catalogue;

to perform tool-supported semantic analysis in 2D or 3D graphic representations to standards contained in the SM2 catalogue.

The method of graphic analysis means the mapping of product or production characteristics to two, three or more dimensional axes of a standardized reference model, e.g. life-cycle phases of product types or of production systems.

When inspecting the SM2 catalogue for retrieving features of PLC (Production Life Cycle of IEC 61131-4) Languages then you may get the following ‚answers' depending of the used classification scheme:

```
product class := control> & <production system phase := design |
implementation> & <product usage := functional layer>;
```

where ‚product class', ‚production system phases', ‚product usage' are dimensional components and ‚control', ‚design', ‚implementation', ‚functional layer' are values that are assigned to the dimensional components.

3.3 Semantic Interoperability

The standardization documentation series of JTC1/SC41 of the IIoT project ‚ISO 21823 - IoT Framework, Transport Interoperability, Semantic and Syntactic Interoperability',

distinguishes explicitly between syntactic and semantic 'Interoperation among Industrial Things' - whereby 21823 part 4 describes syntactic and 21823 part 3 describes semantic "Interoperability between IoT Models".

Thus syntactic interoperability means syntactic data exchange among entities with multiple options of information representation of IoT data - semantic interoperability means data exchange among multiple and different IoT device ontologies.

In the IEC white paper(2019) ‚Semantic Interoperability' it is defined that 'semantics and semantic interoperability' comprises 'linked (data) structures onto which data is mapped and then it is propagated across these structures to produce new data; the latter operation is called inferencing'.

The Standardization Committee SC42/WG3 on AI in the documents of ISO SC42/WG3 24029-2 and SC42/WG3 TR24029-1 does not explicitly address semantics even not in the realm of validation and verification of robustness of *Neural Networks (NN)*. However Formal Description Techniques (FDT) should help 'to determine strong (robustness) properties that are proven true on a whole domain of inputs to a NN and not just isolated ones' of ISO SC42/WG3 24029-2.

The essence is on proving properties with formal methods - which is also applied by the DKE/VDI DINCONNECT:2020 SemNorm Approach. In SemNorm the properties to be proven are directly represented by semantic Graph artifacts and indirectly by language term artifacts (e.g. such as *CSlang* of ETSI in GS ISI006). Here the formal method is the mathematics of many-sorted Algebraic Theories and because of mathematics it is designed to be computational. The textual language terms share the semantics of the given formal methods. Thus a language term may have the same semantical meaning as a programmed piece of behavior of a Digital Twin (Model) in the semantic domain.

4 Workshop Conclusions

More information about various technical aspects of I4.0/IACS industrial semantics have been presented in the submitted paper work to the workshop. The contained publication ‚Semantics for I4.0 Smart Manufacturing' gives a rough overview of standardization activities in the realms of I4.0 language design, declarative semantics and operational Digital Twin simulations.

The next 6th Workshop on IACS is to be announced for Berlin at GI INFORMATIK2021!

Semantics for I4.0 Smart Manufacturing

Jan deMeer[1]

Abstract: In the realm of ‚Smart Manufacturing'[IEC20a] the ‚SemNorm' Project [DKE20a] addresses the question of how to at derive an executable Digital Twin (DT) [DTC20] from standards. A Digital Twin is a virtual representation that embodies an asset of any type [IOSB18]. In that sense a DT is compared to the Asset Administration Shell (AAS). Smart Manufacturing is a real thing of a factory represented by its structure and behavior of inter-connected things that generate real-time data [IOSB18]. By combining Smart Manufacturing processes with a Digital Twin it is intended to validate operations of a production systems in real-time. In general the properties of inter-operating things respectively systems, and especially the properties of energy transportation between systems are considered to be the ‚Prove of Concepts' (PoC) of semantics. When The Information Technology (IT) that processes data and up to some extend information, is compared to the technology that enables communication among things or objects then the technology is called Operation Technology (OT). Whereas the semantics of IT is straight-forward, namely the interpretation of data objects in different contexts of a sending and a receiving environment, the semantics of OT is achieved on two levels. The the first (informal) level explains semantics as a narrative of how things are processed in a smart manufacturing plant, whereas the second (formal) level defines semantics more formally, i.e. by means of graph manipulations. Graph Manipulations represent sequences of events that are related to the narrative of talking about inter-operations among things. At same time a graph is a computational representation (cp. [IOSB18]) in terms of sequences of events (so-called runs) that are executable by appropriate tools from the shelve. Thus graph computations and told narratives are said to be ‚similar', respectively ‚comparable' since they are related to each other by a morphism i.e. the formal relationship between artifacts of graphs, artifacts from the standard ontology and artifacts of the technical asset domain. The OT narration validated by a graph semantics analysis is finally to be transformed into a standard's document which is then called to be a Semantic Standard. This process is a backward transformation of interoperation properties from a semantic representation into an English text that describes the requirements of these properties. In a forward transformation it is started with the textual standards together with the derived guidelines to transform the standard into a semantic representation respectively a Digital Twin.

Keywords: I4.0; Security & Safety; Industrial Automation and Control Systems (IACS); Digital Twin; Production Ontologies; Smart Manufacturing (SM); Asset Administration Shell (AAS); OT/IT Syntax and Semantics; Interoperability; Graph Manipulation Theory; Morphism

1 Semantic Artifacts of Smart Manufacturing

1.1 Smart Manufacturing

[CoW20] on its page defines Smart Manufacturing (SM) as „the approach of control of distributed production machines for the purposes of automatization and optimization . . . "

[1] smartspacelab.eu GmbH, Berlin, 12205, demeer@smartspacelab.de

doi:10.18420/inf2020_27

of interconnected production processes, whereas automatization is based on safe and secure interconnection systems and optimization is based on production data analyses.

Similar [WIKI20a] defines Smart Manufacturing (SM), as a „category of manufacturing based on computer-integrated manufacturing... SM is characterized by high degree of adaptability, rapid design changes, optimized supply chains, efficient production etc.“. Making manufacturing smart may include New Technologies (NT) like „big data processing capabilities, industrial connectivity of devices and services, robotics, Internet of Things, Machine Learning, Human-to-Machine, Machine-to-Machine Communication etc.“ and last not least Syntactic and Semantic Interoperability. etc.

The Reference Architecture Model of I4.0 (RAMI4.0) [DKE20b] represents a ‚three-dimensional roadmap‘ that guides through the I4.0 landscape comprising the three dimensions of (1) IT layers having a view on assets, of (2) life cycle of assets comprising product type and instance management, of (3) the production hierarchies in a distributed plant beginning with the basic product up to a connected world of enterprises.

1.2 Asset and Asset Administration Shell

AAS and Asset are related to each other in the same way as IT and OT are about. AAS is so-to-say the bridge from the world of device-based to information-based interoperation. Whereas the devices are the assets that generate information and the information is operated by the Digital Twin. If semantics would be assignable to the DT, the DT would be able to do more than just operate information, i.e. the DT could reason about and check it for correctness with the asset. This is exactly what the SemNorm Project [DKE20a] is addressing and which is presented by the Prove-of-Concept (see sections 2, 3).

An asset is the central element of RAMI4.0 since each asset obeys a value to the enterprise with a certain function and specific properties which are to be viewed semantically consistent on every IT layer, in every phase of the life cycle and on each level of the plant hierarchy.

An example of a ‚thing' according to RAMI4.0 is a product that is operated by a plant using smart manufacturing technologies, e.g. robots (to be the asset). Another example would be a plant that generates electrical energy (to be the asset) for a remote production plant and manages the energy transportation by automatized control processes.

The properties of things' production, respectively the properties of energy transportation are to be defined in a declarative style of semantics. The declarations are represented on two levels. The the first (informal) level *explains* semantics as a narrative of processing things respectively transporting energy, whereas the second (formal) level *defines* semantics in terms of graph theory. Narratives and Graphs both represent sequences of events being observable in a plant which means that a morphism exists between the narrative of processing things and the sequence of events gained from graph manipulations. Thus the graph is computational using appropriate simulation tools from the shelve. Computations and

narratives are ‚similar', i.e. ‚comparable' to each other, since they are bound by a morphism that prescribe the formal relationship between artifacts of graphs to the artifacts of narration.

1.3 Activity and Interoperability

In part 1 of the 'Digital Factory Framework' [IEC19a] it is said, an 'activity is a group of tasks that are classified as having a common objective' e.g. to perform interoperability. Notice it will be differentiated between the declaration of a task and its successful execution. Thus two different semantic artifacts activity and event are identified whereas the activity is described by a relation comprising a set of unordered node pairs, and the event is described by an ordering of two adjacent nodes. In Graph Theory ‚nodes' are denominated as vertices and ‚activities' respectively ‚events' are denominated as undirected respectively directed edges.

The difference between Activity and an Event, is like 'tossing a coin' and 'observing the rebound of the coin', at which actively tossing and passively observing makes the difference. Speaking in formal terms: the activity is a declaration of measures under certain conditions and that have not yet happened, and the event is the observed impact of that activity after it has happened. Thus the semantic artifacts of activity and event in graph theory both are represented by relations among vertices, i.e. an activity is an unordered pair of vertices, and an event is an ordered pair of vertices whereas the former has the meaning of a dormant transition and the latter has the meaning of an activated transition from vertex A to vertex B.

Since a graph is a network of vertices interconnected either by activated (i.e. directed) or by dormant (i.e. undirected) edges, trajectories through the graph are possible to be identified.

In a sampling taken from Power Management System (IEC TC57) interconnected to a Smart Manufacturing System (IEC TC65) comprising also volatile energy generation, an event of interest that is represented by the graph manipulation (G-e->G') has multiple dependencies, i.e. the volatile weather conditions, the electrical generation conditions and the conditions of the distributed energy resources (DER) of the electrical network and finally the end-user consumption conditions in homes or smart factories. This is a multidimensional event with multiple contextual inter-operations that forms a trajectory from generation to consumption that is considered to be a gradient obeying a value and direction through the graph.

For example: *when* (wind blows in a predefined windspeed-window & a set of wind rotators in operation do not violate the stability criteria of the electrical DER network & end-users request for energy that is in balance to energy offered by the network) *then* the graph gradient specifies the compound system transition from e.g. 'stand-by(G)' into ‚in-operation(G')' graphs.

Notice from gradients useful information respectively metadata can be derived by observation: the flow of energy from 'stand-by(G)' to 'in-operation(G')' is represented by the differential

variable e^{dot}. In the considered system configuration it has to fulfill at least three contextual condition of weather, of electrical network and of energy consumption. Hence all have their specific constraints from which information can be learned after happening or proven in advance.

The graph of energy flow, i.e. the set of sequences of possible transitions through the various contexts of weather, wind rotators, DER networks, SM sites or home consumers is described by:

graph:: (volatile_kinetic_energy || (mechanical_energy —>electrical_energy —>energy_consumption_demand)) —>graph:: ;

where the operator '||' symbolizes independence of coupled system components like wind and rotator being from different contexts, e.g. when the wind blows too strong the rotators must be switched-off to avoid damage.

The operator ‚—>' symbolizes dependency of coupled components like a rotator must at first be brought into operation before electrical energy can be spent into the DER network. Similar electrical energy can be consumed provided the energy demanded from an SM plant can be made available by the compound system.

Notice the coupling by ‚||' is a loose coupling of an autonomous stakeholder i.e. the weather with valuable assets i.e. rotator, electrical network, consumers etc. whereas the coupling by ‚—>' is a strong coupling because of the asset stakeholders' dependency from available interoperation capabilities.

The information learned from the behavior of the compound system is firstly manifested by the invariants of the energy distribution through assets which shall maintain devices' entirety and system's stability during the whole life cycle of operation and secondly by coupling components through a feed-back link that signals current load conditions of the network or current consumption conditions of the end-user. The conditions in the energy network and the end-user (home/SM) conditions are modeled by the pair of continuous variables (g^{dot}, h^{dot}) which are declared in pairs of graph vertices.

2 Toward Semantic Standards?

2.1 Interoperability Semantics

Semantics is a formal theory [MW85] by which sentences of an SM ontology language can be built that become true when its interpretation is restricted to that theory, e.g. Abstract Data Types or Graph Manipulation Theory or some Predicate Logic etc. A formal theory comprises a vocabulary with symbols for constants, functions and predicates, and a set of axioms based on the vocabulary of a certain SM ontology.

In figure 1 the theory that implements the DT is marked with blue color, the technical SM Asset domain is marked in grey color, and the descriptive standards domain is marked in yellow color. All these three domains are extended from the semiotic triangle to an isomorphic rectangle with the four domains of syntax and semantics (blue), technical assets (grey) and descriptive standards also called ontology (yellow).

Thus the diagram shows the anticipated isomorphy between the branches of digitalization (in figure 1 ‚technical translation‘) and modeling (in figure 1 ‚formal translation‘) that should yield same semantics of a Digital Twin. Whereas *digitalization* is the process of translating the physical asset into the semantic representation of a digital twin, *modeling* is the process of translating specified asset characteristics described by a formal description technique into a semantic representation.

From a given standard a ‚guided derivation‘ shall lead to an implementation that is conform to the standard. Similar by applying rules of formalization applied to the standard a standard conformant representation can be achieved. Finally if the two paths of ‚digitalization - guided derivation‘ and ‚modeling - formalization‘ can be proven to be isomorph then the considered standard is called to be a semantic standard because the resultant digital twins are ‚semantic similar‘ and thus the semantic standard is be proven to be correct.

Isomorphy ensures that both mapping paths, i.e. the technical and the formal transformation paths, result in the identical digital twin.

The formal translation comprises the contexts (i.e. graph vertices) of:

A. (English-written) Standards using Narratives (standardized Ontology)

B. Signatures derived from Ontologies (Syntax)

C. Computational Model derived from Signatures (Semantics)

The technical translation comprises the contexts of:

A. (English-written) Standards using Narratives (standardized Ontologies)

B. Sequential Programs derived from Narratives (SW&HW, Programming&Technique)

C. Computational Model derived from sequential Programs (Semantics)

Notice isomorphy means that both paths the formal transformation and the technical transformation have in common the starting points (A) and the end points (C) of the above schema. However the intermediate nodes (B) are different. The standard prescription and semantic description of specifying an asset production are fixed by definition since the semantics must exactly represent the standard, whereas the signatures and programs can freely be chosen provided isomorphy is not violated.

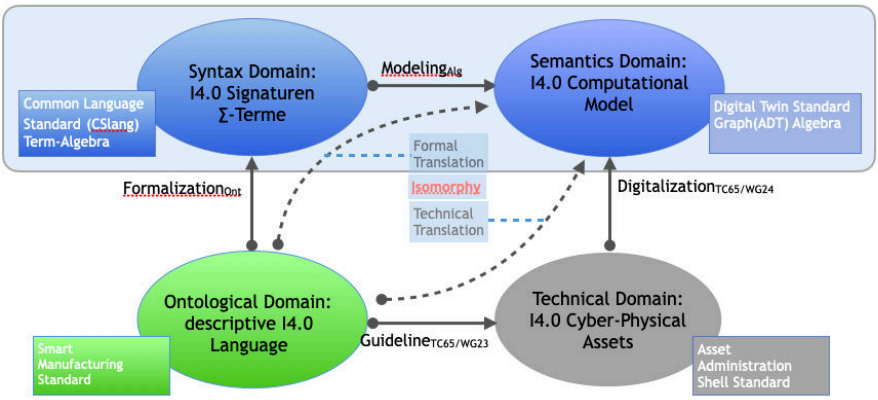


Fig. 1: I4.0 Methodology: Isomorphy between Technical and Formal Transformation

The isomorphy property is useful for validation or verification purposes of the free derivation of a digital twin from a set of standards by comparing the results of both transformations i.e. the one of using guidelines and digitalization and the other one of using formalization and modeling. The degrees of freedom which are proven are the programs and the signatures.

3 The Interoperability Paradigm of Energy Transformation

According to the anticipated PoC the semantics based on formal graph theory shall be demonstrated to be suitable for the interoperability between two big systems and their resources of a smart grid power management system (according to IEC TC57) on one side and a smart manufacturing fabric (according to IEC TC65) on the other side. Both systems interoperate at their system interfaces by exchanging energy from different contexts each one of which follow their own rules and conditions such as weather conditions, mechanic construction constraints, electrical constraints, DER constraints, City/Home consumption conditions.

In the same manner data is exchanged from one context to another one and after translation the data becomes re-interpreted. It is said that the interacting contexts A and B are at an instant of time are bound together by the artifact of a directed edge, represented by the ordered pair (A, B). This translation of data from one context to another one is also an observable event, certain information of interoperation conditions can be derived.

A data (IT) oriented example is taken from privacy domain of defining personal obligations of persons (according to the German DBSG Law §5). The artifacts of directed edges are related to activities respectively edges of authorizing entities or persons, of providing warrants, defining the Personal Identification Information (PIIs) and dealing with personal

obligations of entities or persons. Thus in the Graph domain trajectories can be identified that define a process either of energy transformations or of gaining information from successfully executed activities, i.e. events.

Another OT-based example from technical energy distribution would be that the weather conditions are just like that that the construction of generators will withstand the loads of wind, water, sun etc. hence energy is able to be produced, i.e. transformed from kinetic into mechanical energy representations.

The semantics of energy transformations between sequential contexts shall be considered in more details:

The vector of variables of modeling energy flow through 5 contexts is declared by:

$$<c, A = const, ((m, v), (p, v), (\boldsymbol{M}, \Omega), (cos(phi), outages), (f(\Delta T))>$$

that contains the context variables of the 5 interoperating contexts represented by a network of graph vertices:

1. $c(\mathbf{m\ v^2})^{\mathbf{dot}}$ means the variable kinetic energy of the wind being part of the volatile weather conditions, with m:wind mass, v:wind speed ;
2. $cA(\mathbf{p\ v^3})^{\mathbf{dot}}$ means the earned power from rotators of a certain construction of wind rotators, with A:rotator plane, v:wind speed, p:wind density;
3. $c(\mathbf{M\ \Omega^2\ })^{\mathbf{dot}}$ means the electrical power earned from the rotator, with M:torque, velocity:Ω;
4. (**cos(phi), outages, . . .**) DER form factors, where the 1st one is a metric that influences the impedances of DER branches and the 2nd one is the redistribution of electrical power, when branch outages occur etc.
5. **f(ΔT)** means the provision of transported electrical energy according to the consumer/end user behavior model [cp. RJ. Agüero IEEE PEM];

Notice the two stakeholder contexts ‚Wind' and ‚Rotator' interoperate via their continuous variables, i.e. functions, indicated by x^{dot} of by metrics of wind mass, wind speed, rotator impulse defined by the following vector of metrics:

$$<[kp\ s^2/m]^{dot}, v[m/s]^{dot}, p[kp\ s]^{dot}>$$

4 Conclusion

It has been shown that semantics based on graph theory makes sense to smart manufacturing because semantics is a necessity to invent the Digital Twin providing declarative and operational correctness.

Declarative correctness is provided by a formal style of specifying system behavior, characteristics and properties as parts of an ontology respectively of a textual standard. The formal style of specification is derived from algebraic style of specification where the term algebra derived from a signature (i.e. the syntax) of an Abstract Data Type (ADT) represents the algebra in the semantic mathematical sense. An ADT signature comprises the three parts of declarations, i.e.

- the Sorts (i.e. sets of terms) with Variables and Constants, (Notice a sort could also be understood as single basic context.),
- the Operations representing the system functions,
- the Axioms and Rules representing the constraints and invariants of Interoperation.

The whole of sorts, operations and axioms define a ‚system context' represented by a complex graph, a vertex or a single sort. A graph edge, (i.e. a pair of ordered vertices) combines the axioms and rules of two or more communicating contexts. A model checker analyzes possible transitions between contexts which become inserted into the follow-up graph as a directed edge provided the resources of the transition-in-consideration are made available.

Operational correctness can be shown by applying graph manipulation tools to graph models representing Digital Twins, by which forward and backward simulations are possible. Forward simulation enables predictive maintenance of a production system. Backward simulation enables reasoning on failures, unstable behaviors, necessary system improvements or safeguarding of assets during life cycle processes.

The principles of *syntactic* interoperation among IoT devices that pertain to a certain context comprise for each device type an own (ADT meta) model each containing a distinguished specification of the device. The interoperation among ADT meta models is achieved by a framework including a set of rules of interoperation between different devices.

The principles of *semantic* interoperation among IoT ontologies (i.e. set of asset characteristics) require cross-domain knowledge (e.g. about transportation, electricity and e-mobility domains) for which two general views exist, i.e. a process and a usage view (in the document called models). Whereas the process view focuses on “semantic interoperability” the usage view focuses on “semantic information”.

It has been shown by the chosen Prove-of-Concept from standardization domains of IEC TC57/TC65 that Graph Theory combined with the Theory of ADT is a suitable tool to

model Digital Twins representing complex assets. By the presented semantics approach a Digital Twin is associated with declarative and operational semantics which allows analytic and operative analyses of SM systems.

The relationship between the domains of standardization, technique, syntax and semantics is represented by the semiotic triangle respectively isomorphic rectangle.

Acknowledgement

The work on developing guidelines for establishing semantic standards for Industry I4.0 is supported by the national standardization organization DKE and VDE having launched in February 2020 the DINCONNECT 2020 project #602668 ‚SemNorm'. The authors of this work and of the paper which is published at the LNI proceedings of INFORMATIK 2020 are grateful to the authorities of the DINCONNECT 2020 Organization: Without their support this work on semantics would not have been possible.

Abbreviations

AAS	Asset Administration Shell
ADT	Abstract Data Type (Term Algebra)
BDSG	German Bundes-Datenschutz-Gesetz
DER	Distributed Energy Resources
DF	Digital Factory
DT	Digital Twin
IACS	Industrial Automation and Control System
IoT	Internet of Things
IT	Information Technology
NT	New Technologies
OT	Operational Technology
PII	Personally Identifiable Information
PoC	Prove of Concepts
RAMI4.0	Reference Architecture Model for Industry 4.0
SM	Smart Manufacturing

Bibliography

[CoW20] https://www.computerweekly.com/search/query?q=smart+manufacturing

[DTC20] Digital Twin Consortium® coalesces Industry, Government, Academia on Vocabulary, Architecture Security Interoperability of DT Technology, Milford USA, https://www.digitaltwinconsortium.org/press-room/index.htm

[DKE20a] VDE/DKE2020 (DINCONNECT 2020) 'SemNorm- Requirements to the Formal Digital Twin' ProjectID#602668, February 2020 - Januar 2021; Semantics specified by the Semiotic Triangle (as presented to TC65 WG23 Meeting May 2020); https://www.din.de/de/din-und-seine-partner/presse/mitteilungen/din-connect-gewinner-stehen-fest-706880

[DKE20b] RAMI4.0 The Reference Architecture Model as the baseline for I4.0 Interoperation https://www.dke.de/de/arbeitsfelder/industry/rami40

[IEC20a] IEC TC65/WG23 Smart Manufacturing Framework and Concepts for Industrial Process Management, Control and Automation - Status Report of the Standardization Project ‚Guidelines for Terminology for Information Models'; Web Meeting September 21-23, 2020, Koji Demachi Japan

[IEC19a] IEC 62832-1 ed1 - 65/776/CDV - 2019-11-29 TC65 Industrial Process Management, Control and Automation Digital Factory Framework part 1: General Principles;

[WIKI20a] https://en.wikipedia.org/wiki/Smart_manufacturing

[IOSB18] Industrial IoT - Digital Twin, FhG-IOSB Jürgen Beyerer et al. Publisher, 2018, https://www.iosb.fraunhofer.de/servlet/is/14330/visIT_1-26-03-2018_web.pdf

[MW85] Zohar Manna StanfordU, Richard Waldinger SRI International ‚The Logical Basis for Computer Programming Volume I Deductive Reasoning', Addison-Wesley 1985.

Validation of Control Systems with Heterogeneous Digital Models and Virtualization Technologies

Kirill Semenkov [1], Vitaly Promyslov[2], Alexey Poletykin[3]

Abstract: A modern instrumentation and control system is a cyberphysical system that combines hardware and software components in a network aware environment. The virtualization technologies become common in the development and operation of control systems: they can be used as a tool for system life-cycle extension or for the creation of a comprehensive digital model of the system, a digital twin. The paper deals with the problem of the design of digital twins of cyberphysical systems of an Industry 4.0 era. It analyses and compares the properties of the model in implementation to the modeling of the digital and physical components of a cyberphysical system. A heterogeneous digital model combining virtual machines and emulators and some real software and hardware is recommended for the purposes of testing and verification of complex cyberphysical systems. The role and tradeoff of virtualization environment for control system simulation are discussed. A real use-case of digital modeling is described and discussed. The practical aspects of assessing the performance characteristics in virtual environment and technologies for keeping the synchronous operation of the components the digital twin and the problems of timekeeping within the heterogeneous digital model are discussed.

1 Introduction

The concept of Industry 4.0 and the term itself were phrased at the first time in 2011, by the Germany working group on the vision of industry development prospects [KLW11]; the other countries had performed researches of the same kind [HPO15] as well. In a short period, the concept and the term have become widespread. The authors of the concept [KWH13] consider the industrial environment of the future as a flexible and adaptable cyberphysical system (CPS) that unites manufacturing, warehousing, and logistics through the medium of Internet of Things (IoT). The list of digital technologies that must be used by a manufacturing company of Industry 4.0 includes (see, for example [HPO15], [SS16]) cloud and fog computing, virtualization, artificial intelligence, new data transmission protocols for the IoT and many others.

Such systems are very complex, and often have an expensive and extensive life cycle. A way of decreasing the complexity of the design and verification and validation (V&V)

[1] V.A. Trapeznikov Institute of Control Sciences of the Russian Academy of Sciences, Moscow, Russia, semenkovk@mail.ru, https://orcid.org/0000-0003-0865-9072
[2] V.A. Trapeznikov Institute of Control Sciences of the Russian Academy of Sciences, Moscow , Russia, vitalionics@gmail.com
[3] V.A. Trapeznikov Institute of Control Sciences of the Russian Academy of Sciences, Moscow, Russia, poletik@ipu.ru

doi:10.18420/inf2020_28

of such systems is the creation of their complete digital model In practice, a problem occurs: a digital twin built on a top of pure digital technology often does not well fit to the expectations due to incomplete formal description of the CPS or bad compromise between discrete nature of the computer model and continuous time in actual system [Ok19]. The possible solution is building of a heterogeneous model that combines original and simulated digital parts with some actual analog components. The heterogeneous digital model itself is actually a CPS. Those heterogeneous models have significant advantages [Qu16] compared to purely computer or purely imitation models. Compared with the first ones, they have the ability to more accurately simulate the object behaviour, especially its timing characteristics; compared with the second ones, they are more flexible and easier to use. The heterogeneous digital models are applicable to many aspects of the industrial system development as performance validation, safety, and security assessment.

In the work, we discuss several questions and technical problems in development and application of digital models for the Instrumentation and Controls (I&C) system: the selection of the proper architecture for the digital model and balance between the original digital and analogue parts of the model. We argue the question of having simultaneously discrete and continuous time scale and share the experience received in synchronizing the components of the digital twin using NTP protocol [Mi91].

2 The Review of I&C Modelling Approaches

In this section, we briefly review the main types of model and describe our approach for the design of a digital model for a real I&C system. In general, an I&C system consists of a set of sensors, actuators and the computers with software implementing the control algorithms and human-machine interfaces (HMI). So, an instrumentation and control system itself is a CPS. The validation of the CPS is usually performed in the conditions as close as possible to the operational conditions. The test cases should cover both normal and stress scenario of system functioning. The validation process has some difficulties:

- The absence of the actual control object or its components during design and partially in other stages.
- The impossibility to validate some modes due to risk of the object physical destruction or high costs of testing.

The modelling in many cases resolves these problems. The usability of the modelling depends on the used framework. Tab. 1 shows the comparison of different models and the regard on their potential of use for the physical (Phys.) or digital (Dig.) component assessment.

The most comprehensive and rigorous type of the models is full-scale model. The full-scale test prototypes can be built to try-out and validate the interaction of system elements. There

Tab. 1: Comparison of properties of models. The "+" or "±" signs mean whether the model suits or does not suit for a specific purpose

	Type of model	**Motivation**	**Phys.**	**Dig.**
1	Analytical	Description of a physical object behavior	+	-
		Verification and validation of algorithms (system grey/white-box design)	-	+
		Verification of timing characteristics[4]	+	±
		Staff training	+	±
		Design of the control (system black box design)	+	+
2	Statistical	Reliability and stability estimation	+	+
3	Functional	System design	+	+
4	Data and data flow	Data representation and system logics without regard to real-time system behavior	-	+
5	Full-scale	Validation of system design	+	+
		Validation of models (same as 1–4)	+	+
		Validation of time behavior	+	+
		Validation of system safety and security	+	+
		Staff training	+	+
6	Digital twin (virtual model, pure digital model)	Validation of system logical structure and interfaces	+	+
		Validation of discrete (state-by-state) time behavior	+	+
		Validation of system security	-	±
		Staff training	+	+
7	Heterogenous: virtual and some real components	Combines all advantages of the models of types 5 and 6	+	+

the preliminary check-out is performed with the real equipment; the final integration with the physical object is carried out at the commissioning stage. The process consumes time and resources. With the progress of the computational facilities, a concept of digital twins of CPS has been gaining popularity as alternative of the full-scale models. For example, Lemay et al. ([LFK13]), using a number of virtual machines running within a computational cluster, created a digital model for a SCADA (supervisory control and data acquisition) system of a power plant. They used some simulations of physical processes and PLCs, sensors, and actuators. The model gave the possibility to imitate a bunch of cyberattacks to a reference SCADA system, albeit not showed high productivity. Alves et al. ([Al18]) for security measures tests applied modularity principles to the design and construction of a digital model of a SCADA system. They implemented every element of the SCADA (server,

workstation, PLC, sensor et cetera) as a separate software module, that allows model scaling in a wide range.

However, it is necessary to understand and take into account the limitations of digital models. First, with a digital model, we hardly (if ever) can obtain any data about the productivity of a real system. The system productivity depends on specific models of installed computers and controllers, network capacity, and many other conditions. The model allows getting just some productivity estimates like algorithm performance. Second, the hardware and equipment of any specific manufacturer have its features and restrictions, some internal details of equipment functioning are company secrets, so they cannot be implemented entirely in a digital model. It means the digital model will use some "average," "neutral" models of the equipment, which also does not allow getting an accurate model. Even if we have a perfect model of a hardware, we shell have in mind the problems of time correlation between the dynamics of an actual CPS and its digital model. Indeed, the processor time of program execution flow is not the real physical time, because a computer represents the change of time by incrementing a hardware-dependent counter. Therefore, the time-related properties of the device digital model can differ from the properties of the real device. Lee and Seshia ([LS17]), in chapter 1 of their book, describe the problem in detail.

So, a digital twin for an I&C system should unite the three kinds of models: models of physical object itself, models of control tools (e.g. actuators), and control software or its model. We summarize these considerations in Tab. 2.

Tab. 2: The principles of I&C digital twin design.

	Physical object	**Controllers/sensors**	**Software**
Type of model	analytical (equations)	emulation; black-box software models	—
Representation of time	an abstraction: an argument of the equations	a counter of ticks (hardware dependent)	a counter of ticks (hardware dependent)
Implementation within a digital twin	the equations are solved separately, the twin uses the results and can approximate them	a piece of software running on a real or virtual computer	a piece of software running on a real or virtual computer
Technologies	numerical simulation, clusters, supercomputers	emulation, simulation	virtualization, cloud computing

The maintaining of a single timescale is of great importance for I&C systems. The digital twins are often implemented in a cloud environment with a large number of virtual machines, they have multiple concurrent context-switching events between the processes and virtual machines. It gives rise to the problem of time synchronizations for the components within a model. Currently, the issue of synchronization of a system of virtual machines has little been studied, but the problem is recognized. Thus, VMware [Vm11] says that differences

between virtual and real machines "can still sometimes cause timekeeping inaccuracies and other problems in software running in a virtual machine." Therefore, there is always a risk that the necessary synchronization accuracy would never be achieved in the digital model of the cyberphysical system.

We suppose the restrictions described above can be partially removed in heterogeneous models that combine a digital model with some real parts. The hybrid model compare to the only digital, or full scale model has advantages: better simulation of the functionality of the actual I&C system and timing characteristics correspond to the purely digital model and reduced costs and simplified maintenance compare to the full scale prototype.

For example, a researcher can include in the model a few real sensors and PLCs. The sensors would transmit their output to the virtual medium; the model of the physical process would calculate input signals for PLCs and pass the signals to the standard interfaces of the PLCs. Partial integration of that kind would allow us to check the temporal characteristics of some control signals and facilitate the commissioning process.

3 The Practice of Digital Twin Design and Operation

We designed and implemented a heterogeneous digital model of the upper-level control system (ULCS) of I&C system of a nuclear power plant ([Po17]). The ULCS is a part of the I&C system and is used to provide HMI functions, gather and integrate information from I&C subsystems and perform self-diagnostics. The ULCS consists of servers, active and passive network equipment, workstation, auxiliary equipment of the cabinets (uninterruptible power sources, printers et cetera). The information is transferred over Ethernet LAN; all key nodes and data paths are redundant and work in parallel providing hot-spare redundancy. The ULCP software works under industrial Linux-based operating system LICS OS ([LI19]). The precise time sources provide a unified time scale within the system.

We implemented a heterogeneous digital model (Fig.1) that includes some real hardware of ULCS system and a few dozens of virtual components representing ULCS computers and network devices running in virtual environment. The real hardware includes a timeserver, an Ethernet switch cabinet, a workstation cabinet with operator terminal, server cabinet. The real timeserver acts as the time synchronization source via the Network Time Protocol (NTP). Switch cabinet and server cabinet hardware elements are used, first, to integrate the model with some real hardware and software running on a "bare-metal" computer and, second, for I&C self-diagnostics subsystem verification. The virtual machines for every server and workstation physically run on a host server under LICS OS and QEMU/KVM hypervisor ([KY20], [QE20]); the workstation physical computer runs under LICS OS as well. The software imitators represent the adjacent subsystems of the NPP I&C system. The modelling of network interaction was a significant part of the task. The I&C is a distributed system, the components communicate with one another via TCP/IP and UDP/IP protocol. As we have told, the operational servers, workstation and Ethernet network are redundant,

and the model must reflect both logical and physical redundancy of the system. As the model is a heterogeneous one, it is assumed to interact with external real hardware via some network protocols. We implemented the ULCP network structure and topology in the virtual environment, but made the interaction with the real hardware possible. Virtual switch software OpenVSwitch [Op20] is a tool for network topology construction within the model: every real switch is mapped to a software switch, and the commutation of virtual machines and virtual switches within the virtual network corresponds to the ULCS network topology. VLAN assignment within software switches allows to model physical separation of redundant Ethernet channels. Thus, we model network redundancy where any single failure in the network path does not break the system connectivity.

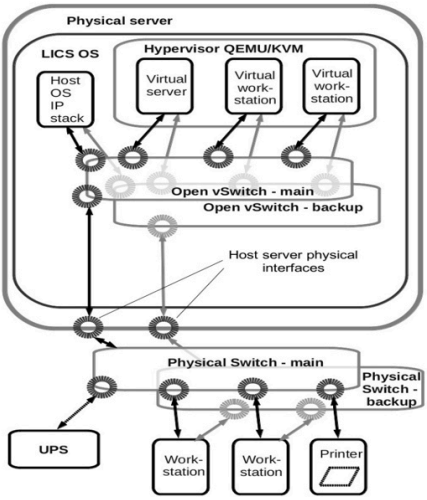


Fig. 1: The outline of heterogeneous digital model for the upper-level control system of an I&C system of NPP

To work with operator graphical environment, we pass the graphics via Spice network protocol [Sp20] to the real workstation; both single and multiple monitor workstation was modelled. Spice client software was also installed onto the virtual machines.

The created heterogeneous digital model allows us to test the interaction of ULCS software components, prepare software deployment to the real complex, test software updates before the deployment, verify security controls. However, the operation process showed some drawbacks in the heterogeneous model design. In the beginning, the computer hardware emulated by the hypervisor (like controllers, network adapters) was chosen to be maximally close to the real one. But the performance of software emulation of the devices did not satisfy the needs of I&C system, the model showed poor network throughoutput. So we had to switch over to paravirtualized devices (for the description of software emulated and paravirtual devices see e.g. [SMP19], [Go11]) and to give up the idea of maximal similarity between the emulated and real hardware. Graphical mode and HMI is another weak point

of virtual models. Virtual environment usually does not allow achieving the same graphics productivity as bare-metal solutions. During the modelling, we did not intend to achieve the graphical productivity of a real system but we were going to test and verify the HMI. We managed to achieve acceptable graphics throughoutput and latency for our purposes. In spite of the restrictions and drawbacks, the designed and constructed model became a good platform for the software complex testing and verification of an I&C system. The heterogeneous digital model allows significantly decrease duration of the validation for the ULCS. The cloud-based platform for the model provides effective and secure remote access for the personal involved in system design during the pandemic lockdown restriction imposed by government officials.

4 Some Aspects of Clock Synchronization in a Virtual Environment

The synchronization of virtual system clock and timekeeping in virtual models is a problem of great importance and presently investigated poorly. The developers of virtualization systems and timekeeping hardware limit themselves to general vague recommendations (see, for example [Bu17). Below we are going to describe our approach to synchronization that we used during creation of the heterogeneous digital model. A virtual server or workstation is a virtual machine (VM) of x86-64 architecture, the virtual machines and hypervisor work under operating system (OS) LICS, a specialized Linux distribution; QEMU/KVM is used as the virtualization software. A GPS/GLONASS NTP timeserver serves as a synchronization source for the virtual model.

Computer clocks usually should conform to the following requirements: they should not stop; they should not go backwards; their resolution must be sufficient for the application purposes; application software must be able to read the clock data; clock access time should be small. Historically the computers of x86 architecture used a set of times (PIT, RTC, HPET and others), but now the main timer is the TSC, Tick Step Counter, a counter of CPU cycles. In modern processors, this register increases evenly and does not depend on the dynamically changing CPU frequency; in addition, in multiprocessor systems, the TSC registers in all processor cores and processors within the same motherboard are synchronized.

TSC emulation is available in Linux-based virtual machines; however, for the stable timer operation during virtual machine migration and CPU frequency changes, major Linux developers (see e.g. [Re20]) recommend using the paravirtual (that is, interacting directly with the host clock) clock driver “kvm-clock”. This is the default clock driver for QEMU/KVM.

However, the experiment shows that the paravirtual driver does not allow to achieve acceptable synchronization accuracy for our virtual model of the I&C system. The experiments show that, first, the virtual machine clock is constantly lagging behind the timeserver clock and, second, the PLL (phase-locked loop) cannot adjust the virtual machine clock. The saw-tooth kind of the curve means the NTP protocol after a period of linear growth sets

the virtual machine clock forcibly. NTP statistics also shows that for this case the clock rate exceeds 500 • 10^{-6}500 · 10^{-6} sec/sec, or 43 sec/day. We assume that simultaneous use of paravirtual clock driver on a number of competing VMs causes delays in hypervisor response to a VM request, so the VM clock latency increase up to 100 times.

After the abandon of the paravirtual driver and switching to the TCS clock in emulation mode in the VM (i.e., refusing to hard link the VM clock to the hypervisor clock), the situation becomes normal (see Fig. 2).

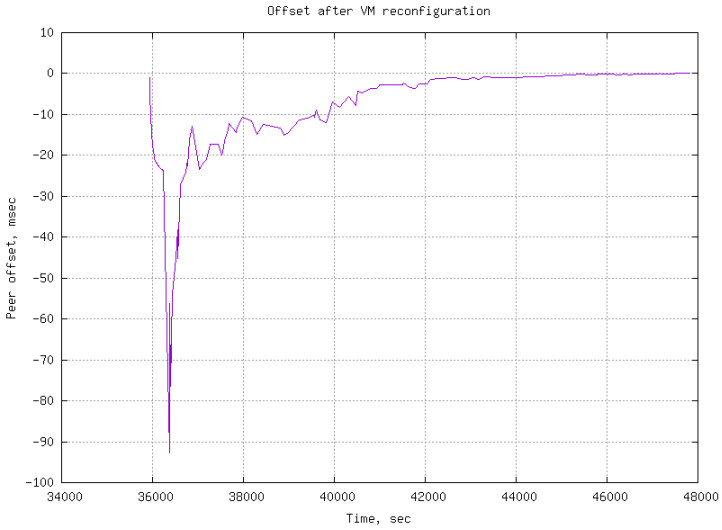


Fig. 2: Clock offset between a virtual machine and NTP server when the virtual machine uses TSC-emulation clock driver

As we can see, in this case, NTP managed to synchronize the timeserver and the VM within a few milliseconds. So, NTP seems to be suitable for application in virtual models but for now its application needs further researches.

5 Conclusions

In the paper, we deal with the problem of the designing the architecture and practical implementation of the digital model of the instrumentation and control systems (I&C systems). I&C system, being a cyberphysical systems (CPS), includes many hardware digital and analogue components. It makes time landscape of the I&C system comprehensive because it combines and mixes discrete and continuous time scale.

The two problems of modelling are considered: the extension of the model life span over whole life cycle of the real objects and accurate simulation of the performance and timing

characteristics of the I&C system. The first problem is solved in a framework of the digital twin approach and heterogeneous model. The digital twin is a useful tool for working out the interaction of system elements, their functional relations and software quality and hardware reliability estimation. To achieve high reference level between the I&C system and model the actual software components of I&C system are wrapped using the virtualization layer. To mitigate the second problem, real hardware devices are used in a time critical functions simulation. However, such digital twin, being deployed in virtual environment, do not provide the ability to measure the temporal characteristics of digital components I&C system and that question still challenging. A useful approach dealing with I&C system performance assessment using deterministic queuing theory is provided in [BP19].We demonstrated the efficiency of a proposed heterogeneous digital model of I&C systems in the course of works on the test and verification of and upper-level control system for a nuclear power plant. The digital components of the model are about 40 virtual machines for servers and workstations, software switches; the real components of the model are some devices (servers, workstations, auxiliary hardware). This solution allows us to significantly reduce the amount of hardware at the test site, the complexity and labour intensity of configuration change during the development, the total cost of test site construction while maintaining a high reference of obtained results. The cloud-based platform used for system's digital model successfully provides secure and moderate development environment for the engineers and system designers. The existence of the model allows uninterrupted development and verification process during the lockdown period without violation of social distance restrictions.

In the process of model implementation, we put attention to the timekeeping problem. For the modelled system, the timekeeping it performed by means of the NTP (Network Time Protocol). We conclude that paravirtual clocks for virtual machines don't fit for timekeeping in high load environment with intensive exchange between virtual machines. The use of emulated TSC (Tick Step Counter) processor register allows achieve time synchronization. However, the search of optimal way of timekeeping is under research.

That heterogenous digital model solves problems of an accurate performance validation of the CPS and verification of the components with incomplete mathematical description. The balance between original software and hardware parts and simulated ones of the digital model is a challenging question as well as architecture solutions of such systems. The experience gained during the creation of the digital model in cloud-based environment can help to solve the problem of the life span prolongation of the software when hardware is outdated and not more available on the market. This already have been done for the upper level components of an I&C system [SMP19] and might be feasible for the controllers and even smart sensor level.

The reported study (Sections II) was partially funded by RFBR, project number 19-29-06044.

Bibliography

[Al18] T. Alves, R. Das, A. Werth, and T. Morris, “Virtualization of SCADA testbeds for cybersecurity research: A modular approach”, Computers and Security, 77, pp. 531–546, 2018.

[Bu17] Burnicki, M., Time Synchronization in Virtual Machines // https://kb.meinbergglobal.com/kb/time_sync/time_synchronization_in_virtual_machines, 2017.

[BP19] A.A. Baybulatov, and V.G. Promyslov “Control System Availability Assessment via Maximum Delay Calculation”, Proceedings of the 2019 International Conference on Industrial Engineering, Applications and Manufacturing (ICIEAM). Sochi: IEEE, 2019.

[Go11] Y. Goto, “Kernel-based Virtual Machine Technology,” FUJITSU Sci.Tech. Journal., vol. 47, # 3, pp. 362–368, 2011.

[HPO15] M. Hermann, T. Pentek, and B. Otto, “Design Principles for Industrie 4.0 Scenarios: a Literature Review”, https://www.researchgate.net/publication/307864150_Design_Principles_for_Industrie_40_Scenarios_A_Literature_Review, 2015.

[KLW11] H. Kagermann, W-D. Lukas, and W. Wahlster, “Industrie 4.0: mit dem Internet der Dinge auf dem Weg zur 4. industriellen Revolution,” VDI Nachrichten, No. 13, 2011 (in German).

[KWH13] H. Kagermann, W. Wahlster, and J. Helbig, “Recommendations for implementing the strategic initiative INDUSTRIE 4.0: securing the future of German manufacturing industry; final report of the Industrie 4.0 working group”, Forschungsunion, Berlin, 2013.

[KY20] KVM Contributors, https://www.linux-kvm.org/page/Documents (last access on Apr. 26 of 2020).

[LFK13] A. Lemay, J. Fernandez, and S. Knight, “An isolated virtual cluster for SCADA network security research”, ICS-CSR, 2013.

[LI19] LICS, RF registration number 2019618036, https://www1.fips.ru/publication-web/publications/document?type=doc&tab=PrEVM&id=07B0B75D-B08F-4A7B-BF76-011ED855B976, 2019. (in Russian).

[LS17] E.A. Lee, and S. Seshia, “Introduction to Embedded Systems – A Cyber Physical Systems Approach”, Second Edition, MIT Press, 2017.

[Mi91] Mills, D.L. Internet time synchronization: the Network Time Protocol. IEEE Trans. Communications COM-39, 10 (October 1991), 1482-1493.

[Ok19] Oks, Sascha Julian & Jalowski, Max & Fritzsche, Albrecht & Moeslein, Kathrin. Cyber-physical modeling and simulation: A reference architecture for designing demonstrators for industrial cyber-physical systems. Procedia CIRP. 84. 257-264. 10.1016/j.procir.2019.04.239, 2019

[Op20] OpenVSwitch Team, http://docs.openvswitch.org/en/latest/, (last access on Apr. 26 of 2020).

[Po17] A. Poletikin, E. Zharko, N. Mentgazetdinov, and V. Promyslov, “The new generation of upper levels systems and industry 4.0 conception in NPP APCS”, Proceedings of the 10th International Conference "Management of Large-Scale System Development" (MLSD). Piscataway, USA: IEEE, vol. 1. pp. 1-5, 2017.

[QE20] QEMU Contributors, https://wiki.qemu.org/Main_Page (last access on Apr. 26 of 2020).

[Qu16] Quadri, Imran Rafiq et al. “Modeling Methodologies for Cyber-Physical Systems : Research Field Study on Inherent and Future Challenges.”, 2016.

[Re20] RedHat, KVM Guest Timing Managemeng // https://access.redhat.com/documentation/en-us/red_hat_enterprise_linux/7/html/virtualization_deployment_and_administration_guide/chap-kvm_guest_timing_management, 2020.

[SBK17] M. Schütze, S. Bondorf, and M. Kreider, “Verification of the FAIR Control System Using Deterministic Network Calculus”, ICALEPS2017 Proceedings, pp. 238-245, DOI: 10.18429/JACoW-ICALEPCS2017-TUCPL06, 2017.

[SMP19] K.V. Semenkov, N.E. Mengazetdinov, and A.G. Poletykin, “Extending Operation Lifespan of Instrumentation and Control Systems with Virtualization Technologies”, Proceedings 2019 International Russian Automation Conference (RusAutoCon) https://ieeexplore.ieee.org/document/8867595, 2019.

[Sp20] Spice project team, https://www.spice-space.org/, (last access on Apr. 26 of 2020).

[SS16] A. Schumaher, S. Erol, and W. Sihn, “A maturity model for assessing Industry 4.0 readiness and maturity of manufacturing enterprises”, Procedia CIRP, 52, pp. 161–166, 2016.

[Vm11] Vmware Inc., “Timekeeping in VMware Virtual Machines. Information Guide”, https://www.vmware.com/content/dam/digitalmarketing/vmware/en/pdf/techpaper/Timekeeping-In-VirtualMachines.pdf, 2011, (last access on Apr. 26 of 2020).

Aligning with cybersecurity framework by modelling OT security

Mithil Parekh[1] Karl Waedt[2] Asmaa Tellabi[3]

Abstract: Before the last decade, production units and its related systems were considered nearly as island systems and were managed as an air-gaped in their daily operations. Information and network security was not an issue because their plant's safety and continues operations have the highest priority. In the recent years, many initiatives like smart factories, adopting Industry 4.0, complex mesh of connected devices and data privacy have shifted paradigm of value chain and trust model in the production environment. By this means, state-of-the-art manufacturing environment demands for the comprehensive framework and holistic approach to address cybersecurity that affects reliability of plant operations. Therefore, few standards are gradually evolving and are extended in to this field. The ISA/IEC 62443 is one of the standard series addresses the Security of Industrial Automation and Control Systems (IACS) throughout their lifecycle. On the other hand, NIST Special Publication 800–82 is a Guide to Industrial Control Systems Security and follows NIST CSF to address OT security. As with Operational Technology (OT) requirements in general, also considering to security-related requirements as per ISA/IEC 62443, ask for more effort to deal with it later. Accordingly, bearing in mind, the need for security from the beginning of the system engineering processes reduces the overall effort and complexity during the lifecycle of OT systems. The corresponding paradigm is called Security by Design. This paper proposes on how high level foundational security requirements from ISA/IEC 62443 can be modelled using AutomationML (AML) tool and consequently explains on how easy is to integrate seamlessly that model during the design phase of engineering process.

Keywords: OT security; AutomationML; ISA/IEC 62443; NIST; Security modelling

1 Introduction

Several trends have made cybersecurity as an essential property of IACS, along with safety, integrity, and reliability. First, over the last two decades, IACS technologies have migrated from vendor-proprietary to commercial off-the-shelf technologies. Second, the value of data residing in the IACS for the business has significantly increased the interconnectivity of IACS both internal and external to the organization. The combination of these trends has made IACS more vulnerable to cyberattack [Qu20]. To deal with it, it requires executing cybersecurity program to guide plant owner safeguarding their OT assets against

[1] Otto von Guericke University Magdeburg, Fakultät für Informatik, Universitätsplatz 2, Magdeburg, 39106, mithil.parekh@ovgu.de
[2] Framatome GmbH, Henri-Dunant-Strasse 50, Erlangen, 91058, karl.waedt@areva.com
[3] University of Siegen, Chair of Data Communication Systems, Hoelderlinstr. 3, Siegen, 57076, asmaa.tellabi@student.uni-siegen.de

doi:10.18420/inf2020_29

cyberattacks. Finally, the means, resources and skills are required that run such security programs efficiently that results in time consuming and very expensive.

The current trend to secure manufacturing environment has gone through its early stage of evolution, reaching an initial level of maturity. Security experts are trying to characterize security in manufacturing environment is good and bring a balanced approach between tackling technical and nontechnical aspects. Technical aspects further characterize into IT and OT while nontechnical aspects characterize in to process and regulatory requirements. The most crucial part here, particularly as cultural issues and potential clashes, between OT and IT/security departments, can jeopardize efforts to tackle security problems [MG19].

However, this is a situation not comfortable especially for small and medium size companies. They should be able to provide and focus its special skills in in operational and safety related cases rather to shift their effort significantly in to OT security issues. Thus, already integrated technical capabilities are required during design phase supporting them to engineer, integrate and use security controls based on technology independent engineering support. The basic requirement is a security model for systems in all its facets, i.e. covering all fundamental security requirements from ISA/IEC 62443. Thereby, the necessary properties and conditions on all levels have to be expressible.

Different approaches are possible based on such a basic model. At first, modeling tool that can be applicable for existing OT systems. Such a tool can provide at least the device configurations and documentations for system installation and maintenance but this paper does not discuss more on this approach. Second, the design process can be supported by the provision of security model enabling security for the system from the beginning.

2 Current trends in OT security

With IT security being tasked with coordinating OT security, in most cases, OT is typically involved in the reviewing of the security processes and the controls to be deployed. This is key in enabling organizations to take into consideration safety and reliability concerns, which remain paramount [MG19].

Several products and services are emerging targeting security in the field of operational environment that can be divided into the general categories along with the core functions. Some of these category and functions may be delivered on-site as appliances (real and virtual), cloud services, or hybrids of on-site and the cloud [MG19]. Here, few categories are explained.

Continuous network monitoring requires gaining the visibility of assets and the means to profile, tracking and managing OT assets. Other than security, this capability also helps engineering team have in-depth insight and properties of OT assets e.g. continuous tracking of configuration of OT assets helps engineering team in the change control management for their Good Manufacturing Practices (GMP) environment. Further, this category includes

any capability that detects anomalies, threats and/or incidents, and provides functions to respond to them.

Network segmentation includes any capability that manages data flow between IT networks and OT environments or OT network segmentation. Unidirectional gateways (data diodes) technology used to compel traffic to travel only in one direction, thereby protecting highly critical environments. Related functionality keeps the data flow secure, particularly for smaller companies that use unified threat management (UTM) solutions. This may also include intrusion prevention system (IPS) functionality [MG19].

Remote access includes specialized solutions that allow for secure, third-party partners and employees' remote access.

Endpoint security for OT assets could be as anti-malware, personal firewall, port and device control, encryption, memory protection, configuration and security-related patch management, continuous assessment, portable media management, application control and allow-listing, integration with other security management systems, and forensic investigation of OT security compromises and impacts [MG19].

Other professional services represent capabilities to deliver risk assessments, strategic planning, policy development, architecture and design skills, as well as software and system integration across multiple technologies and processes. It also represents those managed or functional services that can be delivered via platform, infrastructure and/or software as a service (SaaS) from the cloud [MG19].

All described categories comparatively cover technical capabilities required for ISA/IEC 63443 and NIST Cybersecurity Framework (CSF). Several vendors provides their services in the different categories of OT security and they claim to be the best in the market. However, general implementation guidance and example proof-of-concept solutions is available that demonstrates how open-source and commercial off-the-shelf (COTS) products that are currently available today can be implemented in manufacturing environments to satisfy the requirements in the Cybersecurity Framework (CSF) [CF19].

3 Modelling of system

In order to reduce the complexity of highly sophisticated security requirement and modern production systems, the architecture of OT systems and its process is broken down into various phases. This leads to different and specialized engineering tools for each phase. Naturally, a broad list of heterogeneous tools is witnessed with varied data formats and lack of support for data exchange among them [DR08]. Hence, AML was developed as a vendor-independent, neutral data format based on XML to support such an lossless exchange of engineering information [PA17]. This paper does not include introduction of the basic architecture of AML as it is already explained in IEC 62714-1[EN14]. Further, AML objects

(the core elements of AML) represent instances and can further consist of administration items, attributes, interfaces, relations and references [WH16].

3.1 Modelling of IACS system

To provide a better understandability of the approach, a simple example is used from [AU14]. This modeling methodology will be extended to model security relevant modelling in the next section. In this example, system consists of three logically connected control devices hosting the control application, a switch and three Ethernet wires establishing the physical network. The physical and logical topologies of the network example are depicted in Fig. 1 and its relevant modelling in AML is depicted in Fig. 2.

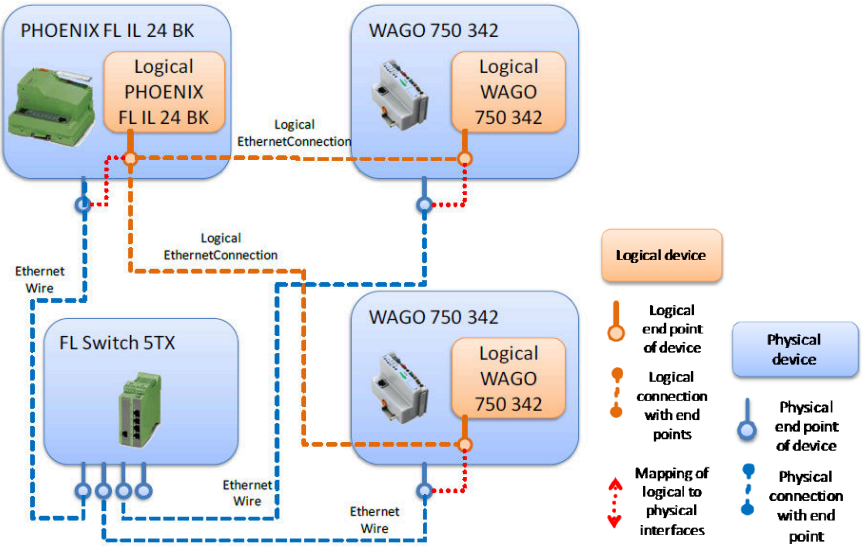


Fig. 1: Physical and logical topology of network example [AU14]

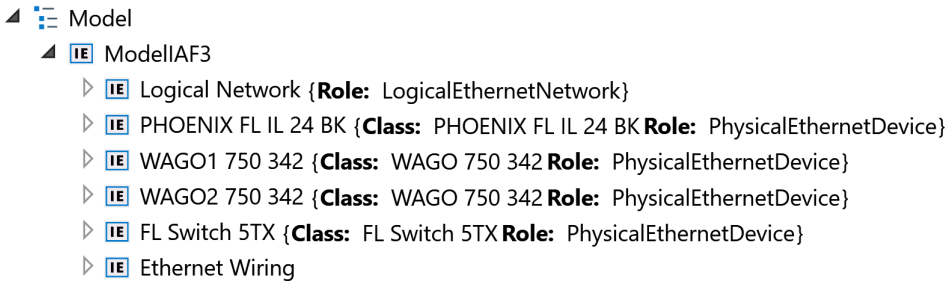


Fig. 2: Instance hierarchy of exemplary IACS

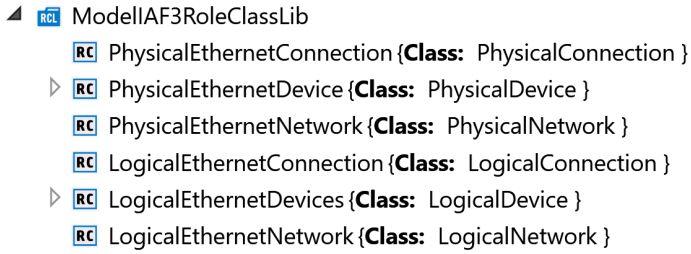


Fig. 3: RoleClass Library of exemplary IACS

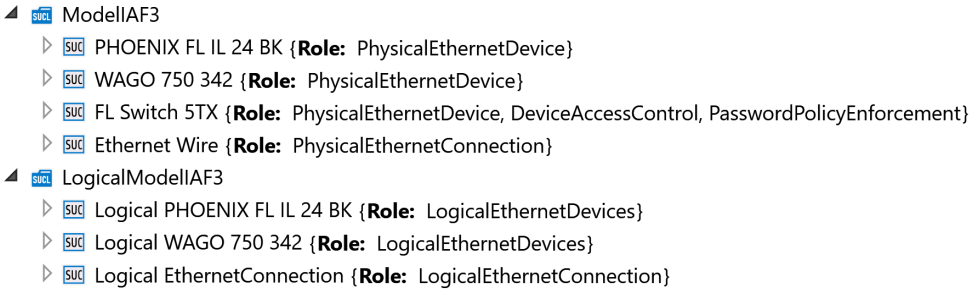


Fig. 4: SystemUnitClass Library of exemplary IACS

The above described modelling can be extended to any existing industrial plant to differentiate its logical and physical part as well as to describe requirements of relevant automation engineering. Already proposed automation model in [MP20] will be extended here for modelling its security relevant property in the following sections (see Fig. 6).

3.2 Modelling of high level security requirements

Before starting security relevant modelling, it is better to focus such a capabilities that comply with security requirements form standards. For an example, Foundational Requirements (FRs) form the basis for the technical requirements throughout the ISA/IEC 62443 series. All aspects associated with meeting a desired IACS security level (people, processes, and technology) are derived through meeting the requirements associated with the seven following Foundational Requirements:

- FR 1 – Identification and Authentication Control (IAC)
- FR 2 – Use Control (UC)
- FR 3 – System Integrity (SI)
- FR 4 – Data Confidentiality (DC)

- FR 5 – Restricted Data Flow (RDF)
- FR 6 – Timely Response to Events (TRE)
- FR 7 – Resource Availability (RA)

FRs include a series of Security Requirements (SRs) describing a number of layered security mechanisms as a baseline. To achieve a maturity for specific security control, a system may be required to demonstrate expected outcomes for specific SRs in their respective FRs. To narrow it down further, the following section targets mainly technical capabilities to fit in to the example, e.g. access control (The combination of FR 1 and FR 2 is sometimes called Access Control). Of course, proposed approach can also model non-technical capabilities but it is out of the scope of this paper.

Access control is the most commonly seen security requirement in any environment and is explained its applicability in manufacturing environment in the next section. Asset owners must develop and maintain a list of all users (humans, software processes and devices) and determine for each control system's component the required level of access control protection. The goal of access control is to protect the control system by verifying the identity of any user requesting access to the control system before activating the communication.

Mutual dependency of access control with other FRs (e.g. System Integrity and Data Confidentiality) are not specifically discussed here. Subsequently, this paper target explicitly security requirements from ISA/IEC 63443 and proposes relevant security modelling with most common SRs.

3.3 Access Control for IACS

To restrict physical and logical access to IACS systems and networks, users must be uniquely identified, authenticated, and authorized before gaining access. User authorization should follow the principle of least privilege that grants users with sufficient privileges to enable them to fulfil defined roles.

Authorization is the initial step in protecting an IACS system and its critical assets from unwanted breaches. It is the process of determining who and what should be allowed into or out of a system. Once this information is determined, defence-in-depth access control measures can be implemented to verify that only authorized people and devices can actually access an IACS system. The first measure is usually authentication of the person or device that is attempting access to an IACS system [IC13].

Authentication describes the process of positively identifying potential network users, hosts, applications, services and resources using a combination of identification factors or credentials. The result of this authentication process then becomes the basis for permitting

or denying further actions. Based on the response received, the system may or may not allow the potential user access to its resources [IC13].

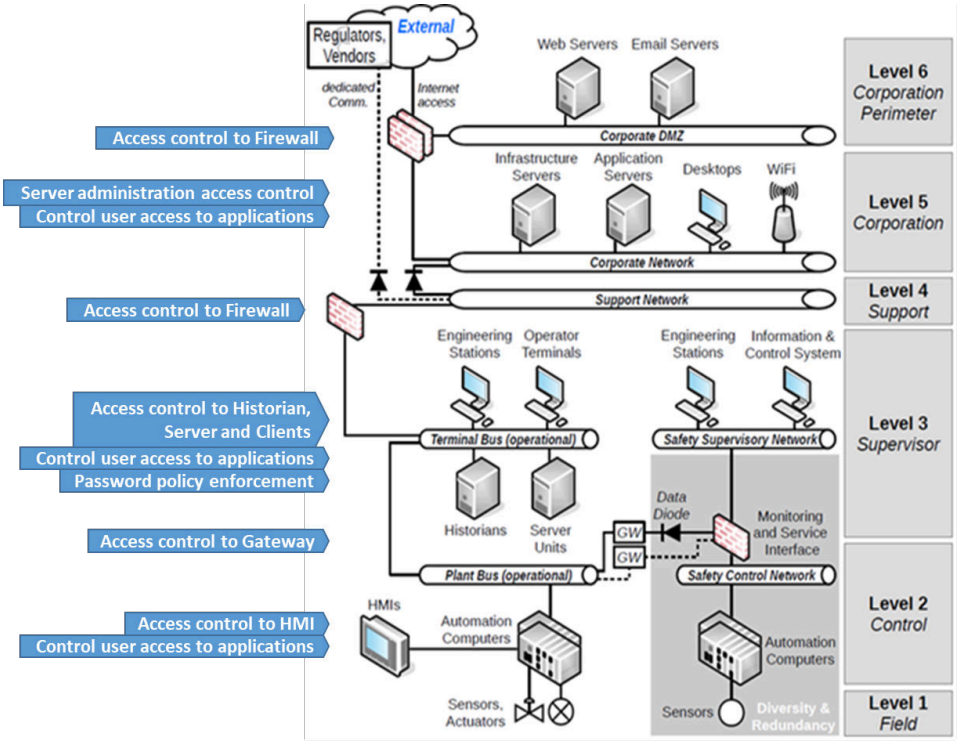


Fig. 5: Access Control for IACS

Fig. 5 illustrates the security patterns for the access control information security domain. The tags on the left side represent the access control security patterns that can be consistently applied across the IACS network [OL15].

Using existing automation system communication model from AutomationML whitepaper [AU14], we can further extend that modelling to include access control requirements form ISA/IEC 62443 standard as per Fig.6.

The resulting AutomationML file is a condensed version, also covering data of interest. A receiving target tool can automatically import those AutomationML files and can import the modification immediately.

4 Conclusion

This paper explains the need of security modelling and on how it can help in the earlier phase of engineering lifecycle with the example of modelling access control. Such a

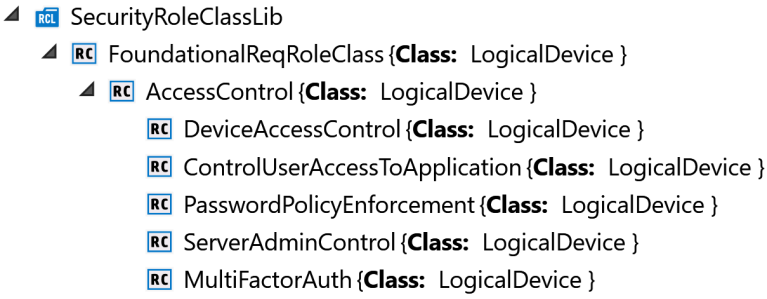


Fig. 6: SecurityRoleClass library (with AccessControl RoleClass) for IACS

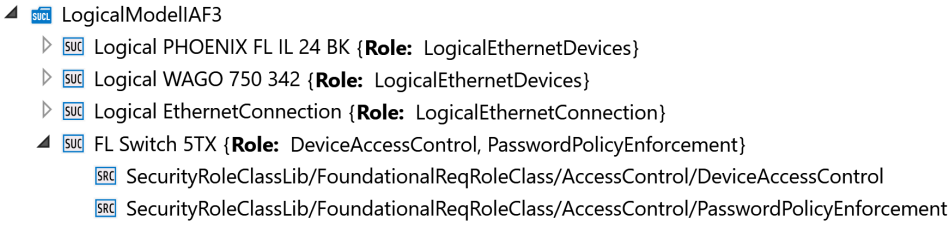


Fig. 7: SystemUnitClass library with assigned security RoleClass for the switch

model can avoid later all the possible effort that requires skills and resources to secure the existing IACS. The process of security abstraction and its modelling is based on ISA/IEC 62443 standard series that can be applied on real IACS and production lines, and closely linked to existing issues and problems. Further, it can also help engineers to comply with regulatory requirements and to educate on security issues in their daily operations. The main element of this solution is a continuous, heterogeneous integration of IACS components, process engineering and the relevant stakeholders by considering all aspects of an OT security requirement. This enables suppliers, asset owner and integrators of such automation components to receive support from this model e.g. when creating a new system, selecting appropriate security controls, defining incident response plan and for troubleshooting. Modelling based on AML is an acceptable data format among automation vendors and is currently used for the seamlessly data exchange during design process. Therefore, its main advantage can also be leveraged as justification during further development of OT security standards and to increase its effectiveness.

Bibliography

[PA17] F. Patzer, A. Sarkar, P. Birnstill, M. Schleipen and J. Beyerer, "Towards the modelling of complex communication networks in AutomationML," 2017 22nd IEEE International Conference on Emerging Technologies and Factory Automation (ETFA), Limassol, 2017, pp. 1-8, doi: 10.1109/ETFA.2017.8247571.

[Qu20] Quick Start Guide: An Overview of ISA/IEC 62443 Standards Security of Industrial Automation and Control Systems 2020.

[IC13] IEC 62443-3-3:2013 Industrial communication networks - Network and system security - Part 3-3: System security requirements and security levels

[MG19] Gartner: Market Guide for Operational Technology Security. Published 5 November 2019 - ID G00370177

[CF19] NISTIR 8183A Vol. 1: Cybersecurity Framework Manufacturing Profile Low Impact Level Example Implementations Guide: Volume 1 – General Implementation Guidance 2019.

[DR08] R. Drath, A. Lüder, J. Peschke, and L. Hundt, "Automationml-the glue for seamless automation engineering," in Emerging Technologies and Factory Automation, 2008. ETFA 2008. IEEE International Conference on. IEEE, 2008.

[EN14] IEC 62714-1:2014 Engineering data exchange format for use in industrial automation systems engineering - Automation markup language - Part 1: Architecture and general requirements.

[WH16] AutomationML Consortium, "Whitepaper AutomationML Part 1 – Architecture and general requirements," AutomationML - The Glue for Seamless Automation Engineering, 2016.

[AU14] AutomationML Consortium, "AutomationML Whitepaper Communication," AutomationML - The Glue for Seamless Automation Engineering, 2014.

[MP20] Parekh, M., Gao, Y., Jockenhoevel-Barttfeld, M., and Waedt, K., "Confluent Modeling of Heterogeneous Safety and Operational Instrumentation and Control Systems." ASME. ASME J of Nuclear Rad Sci. July 2020; 6(3): 031802. https://doi.org/10.1115/1.4046262

[OL15] Obregon. L, "Secure Architecture for Industrial Control Systems", SANS.edu Graduate Student Research, 2015. https://www.sans.org/reading-room/whitepapers/ICS/secure-architecture-industrial-control-systems-36327

R. Reussner, A. Koziolek, R. Heinrich (Hrsg.): INFORMATIK 2020,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2021 321

Operational Security Analysis and Challenge for IoT Solutions

Yuan Gao[1], Xinxin Lou[2]

Abstract: The marketing engagement of Internet of Things (IoT) shows a wide vista together with Industry 4.0 regarding modern manufacturing and services. However, the evolution of technologies and rising regulation concerns regarding security and privacy are bring challenges to IoT solutions. On one side, the security analysis of IoT solutions has to consider the security posture in a much wider scope including both edge and cloud sides even across global geo-locations. On the other side, new regulation requirements demand a full tracking of data access. In addition, authorizations should be evaluated explicitly and can be revoked any time for maximizing data protection. Both challenges can be solved by implementing a novel security model targeting those requirements while zero trust model is a good candidate. Thus in this paper, we compared the most commonly used perimeter security model and the zero trust model under the circumstance for modern IoT solutions. Furthermore, from the regulation perspective, the concepts of zero trust model are analyzed to show its compliance with regulation requirements. For easing the discussion of IoT solutions, a general IoT architecture is proposed and relevant zero trust model implementations are described. Especially, the zero trust model relevant security controls are highlighted as a guidance for the design of IoT solutions. As the conclusion, we propose a general implementation of zero trust model within the context of IoT solution to solve the challenges facing by the industry.

Keywords: Operational Security Model; Zero Trust Model; Cloud Security; Edge Computing; IEC 62443; Industry 4.0; GDPR; IoT; IIoT

1 Introduction

Along with the development of cloud computing and wireless technology, Internet of Things (IoT) is achieving the biggest ever market engagement. According to the GSMA Intelligence projects, the market cap of IoT will be more than 1 trillion until 2025 [GS18]. Apparently, the ubiquitous existence of IoT devices and the heavy back and forth data traffic are arising new security concerns and requirements. In addition, the deployment of cloud technology uses infrastructure sharing to enable the resource elasticity and to reduce the cost whenever applicable. Furthermore, the virtual factory concept in Industry 4.0 allows the temporary combination of services across different geographical locations [Rü15]. The above mentioned reasons are eliminating the security boundaries of IoT solutions especially the network perimeter. Meanwhile, the tremendous number of IoT devices and their limited security capacity require dynamic while robust security in architecture designs. At last, the

[1] Otto-von-Guericke University Magdeburg, Research Group Multimedia and Security, yuan.gao@ovgu.de
[2] Bielefeld University, xlou@techfak.uni-bielefeld.de

doi:10.18420/inf2020_30

new regulation requirements on personal identifiable information (PII) expose IoT solutions to the compliance risks: e.g. General Data Protection Rules (GDPR), Payment Card Industry Data Security Standard (PCI-DSS) and Health Insurance Portability and Accountability Act (HIPPA).

For handling the complex IoT solution architecture as well as fulfilling the compliance requirements, zero trust model is being introduced into modern system architectures. Zero trust model will assume no trust for any personnel, entities or services involved within a system process. Thus, every service request and reply should be under continuous monitoring and will trigger alarms for possible compromising. The operational 3-domains security model proposed in previous work considered a flexible representation of system security architecture as well as continuous security monitoring [GI2019]. Thus, it is suitable for further adaptions and extensions to handle the above mentioned IoT challenges. Within the 3-domain security model, we will address the required features for implementing an IoT solution in line with the zero trust model. In addition, analysis on chosen IoT scenarios implemented with zero trust model can show the compliance of this architecture regarding different security regulation requirements.

The rest of this paper is organized as follows: firstly, Section 2 describes a general IoT solution architecture especially defines the edge and cloud locations as two major focuses. Secondly, in Section 3 we discuss related works and relevant security regulation requirements. Then Section 4 discusses the challenges faced by the perimeter security model. Section 5 introduces the zero trust model within the IoT solution context. It is compared with perimeter security model and demonstrates how the regulation compliance can be met by implementing this model. In addition, Section 6 highlights the different security concerns between IoT and IIoT. Finally, the conclusion is summarized in Section 7 together with the discussion regarding future works.

2 IoT Solution Architecture

As a booming technology, innovative IoT solutions are being created every day. So there is no standard or typical IoT solution architecture. Here we will use Figure 1 as an example of IoT solutions for describing the common features and components. From a simplified data flow perspective, we can go through the picture from left to right. On the most left side, IoT devices collect data from their sensors (like the thermometer on the top) or receive voice commands from nearby users (e.g. the smart TV or the speaker in the middle). IoT devices can send data to the cloud either via a relaying/edging device as the green colored symbol indicated in the picture or communicate with cloud directly, like the IoT car module or the camera in the bottom left corner. In the middle within a public cloud, collected data will be processed and stored in object storage or databases for serving queries or historical tracing. Data stored in cloud will be exposed as services. They can be accessed and consumed by other (cloud-native) services for creating various business plausible results, e.g. a historical regional temperature record. Front-end services reside in the public cloud, such as web

servers and load balancers are omitted here. Finally, on the right side of this figure, useful information can be viewed on terminals, e.g. smart phones or laptops with effective UX designs like a dashboard or diagrams.

For easing the description, we may consider two IoT scenarios regarding Figure 1: smart home and smart car.

Smart Home: In this scenario, a user at home can talk to her/his smart speaker for controlling other devices at home. For example, the voice of "Open my TV" will be sent over to the cloud for recognition. The identified command for opening a TV registered in this home will be returned to the smart home router and after validation it will be further forwarded to the TV to react to this command.

Smart Car: Sensors are on a car keep sending values, e.g. location and speed to a central cloud service endpoint. At the same time, the cameras installed on top of traffic lights are transmitting pictures taken in a regular interval to the cloud for image processing. All information, collected from cars and cameras, are put together in a data processing application to build the real-time traffic loads within a specific city.

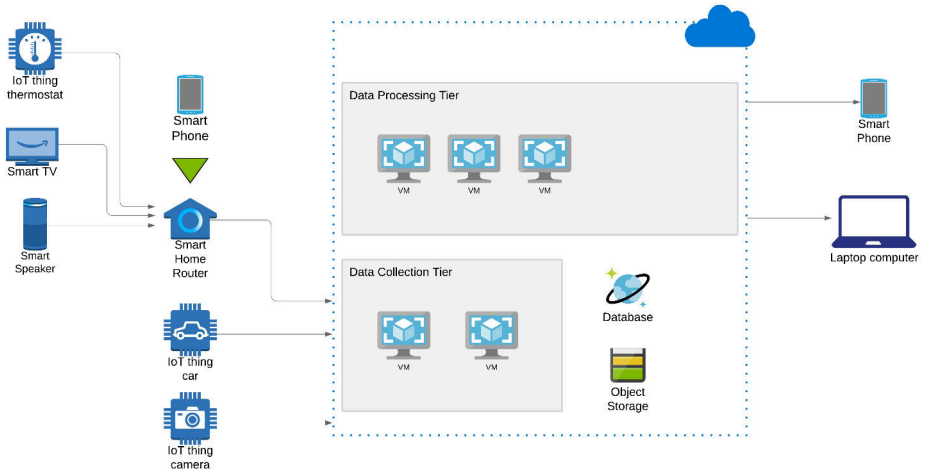


Fig. 1: A general IoT architecture.

Regarding the IoT solution architecture proposed in Figure 1, following declared important concepts can help the analysis in later chapters:

IoT Devices are devices that connect to network with specific cyber-physical functionality (e.g. sensors or actuators). Normal IoT devices can be massively produced with a low cost and have limited capacities on computation and storage. They can be equipped with only essential or little security features. However, a special kind of IoT device can have advanced network and security features, which we name as **edge device**. An edge device can relay the communication between other IoT devices and cloud services/terminals. Thus it is an

ideal host for placing security controls and temporary data processing functions. Figure 1 depicts the smart home router as an edge device (indicated by the green symbol). However, an edge device does not have to be a real device but can be a virtual one. In this case, its extra network processing and security capacities can be achieved by combine several IoT devices into a mesh.

Terminals are the ending devices by which users communicate with the IoT system. Two kinds of typical terminals are smart phones and PCs. By running applications on a terminal, users can view valuable results from the IoT solution and send commands to the system to perform specific actions or modify configurations. A terminal can connect to IoT services published on cloud or communicate with devices exposed an interface hosted on the cloud. Both connections are established over internet thus communication protection mechanism is needed. In addition, terminals can access IoT devices via WIFI or Bluetooth in a small area, e.g. within a home LAN.

IoT Applications can be any application running on IoT devices, cloud resources and terminals. According to their purposes, applications can be classified into two types: *Management Application* and *Data Application*. Management applications are used for controlling the IoT system and individual devices. Comparing to this, data applications collect, process and illustrate data as the functionality of an IoT solution. For example, in the smart scenarios, applications collect values and pictures from IoT devices and they are processed and build up together in the cloud services for illustrating the real-time traffic status.In this paper, data collection and processing will be further discussed due to their tight connections to security and regulations.

3 Related Work

The work in [KS18] reviewed the security issues and challenges in IoT. Some researchers consider combining the software-defined networking (SDN) into the IoT in order to bring security and the privacy with flexibility and scalability [KBL18]. DeCusatis et al. [De16] propose a network architecture which enables an explicit zero trust approach. This architecture is based on a steganographic overlay (with authentication tokens in the TCP packet request), and first-packet authentication. Vanickis et al. describe a policy enforcement framework to address many of open challenges for risk-based access control for zero trust networking [Va18]. Ahmed et al. [ANT2020] propose a model which provides the access control to sensitive data in zero trust model. In this model, an access control proxy is used to protect the sensitive data to implement the access control. The access control is realized by performing the analysis on access request, user type, device type, application type and data type.

To regulate the potential security threats, some privacy and security rules or standards are developed by specific organizations, e.g. the Guide to the General Data Protection Regulation (GDPR) [In18], Payment Card Industry Data Security Standard (PCI-DSS)

[PC04] and Health Insurance Portability and Accountability Act (HIPAA) [Of02]. They require PII data should be well protected. To achieve this, data access must be traceable, and protection action will be taken when the data is under a certain level of risks.

The proposed operational model here is an extension of our previous work of the 3-domains model regarding threat domain, system architecture domain and security domain [Ga19]. The work discussed in this paper focuses on the security domain.

4 Challenges of Perimeter Model

Before getting into detailed analysis regarding zero trust model, it is meaningful to declare the features of security perimeter model as a comparison. A perimeter model in security classifies a system into different areas or zones. Same security requirements are shared within one area/zone while the security level is going up for inner zones which have shorter perimeters to the system center, in other words, protection goals. The benefits of security perimeter model are quite clear. Firstly, the classification of zones with different security requirements reduces the analysis and management overhead thus increasing the security governance. Security postures can be improved by adding extra zones outside, such as demilitarized zone (DMZ). Secondly, by assuming same security requirements within one zone, traffic monitoring can focus on the communication between zones. This means investing network security appliance only on tunnels (conduits) to reduce costs. Finally, security zones are prioritized naturally by their perimeters to the center, which guides the continuous activities for security improvements. According to these reasons, the perimeter security model fits the traditional server-centric architecture well. The Zone-Conduit model introduced in IEC 62443-3-2 is an abstract implementation of perimeter model in industrial fields [IE15].

However, toward the evolution of IT technologies and new regulation requirements, the security perimeter model is not feasible anymore. On one side, the monolithic server-centered architecture is reaching its bottleneck while instead, distributed and decoupled architectures like micro-services are trending popular. Without a fixed combination of components, it is difficult to classify a stable list of security zones since the size and contained components can vary when time goes by. The virtual factory concept within Industry 4.0 is an example of this challenge. Against a virtual factory consists of dynamic available services over network on the global scale, it is almost impossible to implement a security perimeter model on it. On the other side, the authorization to access data is not checked explicitly in every access. In other words, the perimeter security model is a static model. Once an identity or a service is authorized to access the data, they obtain the access until the authorization is removed manually. For example, one employee can resign and leave the company at the same day. However, the users/passwords of the employee will stay in the company's systems for a long while until the IT-staff decide to do a cleaning. In addition, it is also a violation of the regulations of GDPR and PCI-DSS to have non-recorded data access without checking attached policies explicitly.

5 Zero Trust Model

5.1 Core Concepts

In Chapter 4 we discussed the features and challenges of the perimeter security model. In a perimeter model, a certain level of trust will be established when specific conditions are met. For example, if a server locates in the back-end network and hold the right access key, it will be authorized to access the database. However, after the authentication and authorization happened in the beginning, the database access (from the server) will not be tracked and the conditions for authorization will not be checked in a specific period. This mechanism can reduce authentication/authorization overhead effectively. However, it lacks the continuous monitoring on each access. In addition, servers will be vulnerable within a period of trust relations. For solving these challenges, zero trust model can be applied by implementing following rules:

- **Rule-1(Traced Access)**: all data access should be tracked and stored in logs.
- **Rule-2(Explicit Evaluation)**: data access authorization should be evaluated explicitly before deciding *Allow* or *Deny*.
- **Rule-3(Automatic Revocation)**: under certain conditions, the system is capable of revoking data access authorizations automatically.

First of all, all data access should be logged and this rule contains several layer of details. Since IoT solutions across edging and locations, the tracked data access includes both raw data collected on devices and internal processed data. From the security perspective, both the privacy of individual user and the trade secrets of IoT solution providers should be protected. In addition, logs need to be time-stamped correctly for later correlated analysis, e.g. using a security information and event management (SIEM) system. Based on this, it is possible to apply artificial intelligence (AI) for indicating possible compromises by learning from normal data access behaviors.

Secondly, for maximizing the protection of data, access request should be explicitly evaluated every time without exceptions. However, this will bring the system extra overhead while provide more secure access control. Later in Section 6 we will discuss relevant trade-offs especially for non-privacy scenarios. In general, a careful classification of data can help to design cost-effective evaluation mechanisms thus reducing the overhead.

Finally, the explicit evaluation discussed above enables the automatic authorization revocations to prevent malicious accesses. Revocation can be triggered by different ways. For example, when an abnormal access behavior is detected, the authorization can be revoked to interrupt possible on-going attacks. In the case of a false positive alarm, authorization can be requested by administrator or through user's self-services. Authorization can be automatically revoked by a timely manner as well. As an example, an administrative data

access authorization should be allowed only in a reasonable time windows and will be revoked automatically when expires. This will minimize the possible attack window and reduce the administrative efforts regarding the security governance.

In the following section, implementations of the three rules will be proposed to show how zero trust model can improve the security posture of IoT solutions.

5.2 Threats and Controls

Regarding an IoT solution described in Figure 1, it is meaningless to go through all attack vectors and associated mitigation. For example, uploading crafted malicious firmware to IoT device or virus affected PC/smartphones will not be considered. Instead, only the zero trust model relevant threats will be considered. There are threats for the edge location and the cloud location. It is also possible a threat with a long kill chain can go over both locations. Compared to this, security controls are bounded to their locations. A control or a mitigation will take effect either on the edge or on the cloud with limited exceptions. One such kind of exception is customized communication protocol which will affect both edge and cloud locations. However, considering the majority of IoT solutions involving public cloud, it is very rare that a customized protocol will be used. In summary, security controls which implement zero trust model on either edge or cloud location will be discussed. The threats identified here are not part of a formal risk assessment and they can be easily performed by *insiders*. Thus, in the following paragraphs, we will discuss threats without assessing their impact and likelihood explicitly.

Edge Location

Table 1 lists three identified threats towards a general IoT system on the edge location.

Tab. 1: Threats on Edge Location.

No.	Threat	Relevant Rule(s)
1	Traffic Sniffing	rule 1
2	Rogue Device	rule 2 and 3
3	Malicious Operation/Configuration	rule 2

The first considered threat is traffic sniffing. We can use the scenario of smart home to analyze this threat. An attacker might sniff data traffic within the home network, either via cable or wireless connections. It can be achieved by compromising the single router in the network or by hijacking wireless connections between IoT devices within the network. This kind of attacks are possible since violating rule 1: *Traced Access* described in Section 4.2. In the traditional perimeter security model, a home network after a router plus a firewall towards internet is considered secure. So communications within the home network are treated as harmless without protection. To fulfil the requirement of rule 1, communications within the home network must be monitored by recording at least basic information, such as

time, protocol and length of packets. In addition, communications within the home network should be encrypted for preventing manipulations which can break the integrity of traced records.

The second threat is to put a rogue device into the edge network. A rogue device can act as a normal IoT device. However, it can be controlled to perform malicious activities, such as eavesdropping its surroundings or joining DDoS attacks. This threat can be mitigated by implementing rule 2: *Explicit Evaluation* and rule 3: *Automatic Revocation*. For rule 2, in a zero trust model, a connection request must be checked whether its authorization exists and is still valid firstly. There should be no assumption that a device is *probably secure only* because it locates within the edge network. Request from an unknown device within the network should be rejected and an alarm will be triggered for attracting the administrator's attention. Furthermore, according to rule 3, regular review of existing devices should be performed. An administrator or a user has to provide a trace of the device how it is installed within the network in a given time window. If not, the device must be isolated automatically. Again in the smart home scenario, if the user only setup a speaker and a TV previously, then any packet from the unknow thermometer device should be dropped.

The last threat is performing malicious operations or configurations on IoT devices. Direct access to IoT devices, such as operating on device user interface is not protected due to the assumption the physical security protection is sufficient. As a security guideline for medical devices in Hospital, user interfaces should be protected by passwords and manufacturers do provide such functionality for compliance purposes. However, in reality, due to the high frequency of emergency situations, these passwords are normally disabled or set to an easily memorized value, such as *0000*. Apparently, this is a violation of rule 2. Considering our smart home example so far, it is also so annoying to input a password every time serve a cup of coffee from a coffee machine. As a plausible mitigation, innovative authentication methods like face recognition should be applied to reduce the impact while to fulfil the requirement of rule 2. In addition, implementing rule 1 to log user activities might provide a basement for detection of abnormal behaviors, which can contribute to mitigation too.

Cloud Location

On the cloud side, three threats are identified in Table 2.

Tab. 2: Threats on Cloud Location.

No.	Threat	Relevant Rule(s)
1	Insufficient Logging	rule 1
2	Hard-coded Access Key	rule 3
3	DoS on shared resources	rule 2 and 3

The first considered threat on the cloud location is insufficient logging configured for infrastructure as a service (IaaS), e.g. virtual machines (VMs). The access to data on VMs will be logged on the application level within individual VMs. Both Amazon Web

Services and Microsoft Azure provide the services to grab and backup logs from VMs for monitoring or regulation requirements. Without correct logging configuration, in case a VM is terminated due to a hardware failure under the virtualization layer, the application log content will be lost. This will cause missing data access records, which violates of rule 1 and regulation requirements, such as PCI-DSS. Thus, IoT solutions should follow the cloud best practices to enable IaaS logging and log collection functionality.

In the next we pay attention to the threat from hard-coded access keys. Access keys can be used for authenticating a user or an application for accessing data or cloud services. However, hard-coding access key in an application is a bad security practice. On one side, this violates rule 3, since it is very difficult to revoke the access key which is designed for a long retention time. Revoking an access key will invalid all encrypted data linked to this key. Furthermore, using access key cannot assign a user or an application authorization based on the least privilege principle. Holding an access key means the root access privilege. Instead, role-based access control (RBAC) should be used to provide granular access controls. It is also possible that the access key will be leaked to public not only to insiders. However, since normally cloud services are configured with firewall rules, here we consider only the access key might be misused within a trust network, e.g. a private network on cloud. Considering the smart car scenario, it is a best practice to segment service access controls regarding different cities. However, with a universal access key, no valid access control will be in place.

At last, the threat of DDoS attack on shared resources will be discussed. Dockerization is the state-of-the-art technology of computing virtualization. Applications will be packaged with dependencies and executed in individual light-weighted docker containers. Reflecting our smart car scenario, servers for different cities can be deployed in different containers thus to isolate them targeting a robust service architecture. However, as the design, several docker containers will be hosted on the same hardware. Thus, it is possible for one docker container to consume up all computing resources so that other containers will be in a status of DoS. To mitigate this threat, an implementation of rule 2 and 3 can be helpful. Within a computation windows, a container has to request computing resource before its execution (rule 2). Once allowed, the authorization to computing resource will be revoked after a given time period (rule 3). Thus, other containers can access computing resources without starving status.

6 IoT vs. IIoT

To summary the analysis of IoT solutions so far, implementations of zero trust model can effectively overcome the drawbacks of perimeter security model mentioned in Section 4. However, there is no single silver bullet for everything. Especially, according to the IEC 62443-3-2 [IE15], a perimeter-based zone-conduit model is proposed as the solution for Industrial Automation and Control System (IACS). In the context of this paper, this means towards an IACS enhanced with IIoT solution, perimeter security model is the major choice. This trade-off can be understood from two perspectives:

- **Availability vs. Confidentiality**: For IACS, the availability of system has higher priority.
- **Safety vs. Privacy**: Functional Safety is strictly regulated in industrial fields while regulation on privacy is being progressed.

On one side, availability has the highest priority in the availability, integrity and confidentiality (AIC) triad in the context of an IACS. Especially, when the availability of a critical infrastructure, e.g. the power grid is affected, the impact will be very high. In this case, security controls that affect availability should be avoided. For example, an automatic access deny on requests from safety-critical system is not allowed. In additional, the implementation of rule 2 (Explicit Evaluation) might affect the availability for time-critical functions, such as hard real-time tasks. On the other side, the zero trust model focuses on data access which is more important for protecting privacy. If the IACS or a partial system of it is running without processing PII, implementations of zero trust model is not mandatory without regulation requirements. There are more differences between a consumer-oriented solution and an industrial solution, by which trade-offs on system architecture will be decided.

To conclusion, zero trust model do not have to be implemented completely for an IACS. However, as the trend of Industry 4.0, the border between IACS and consumer systems is blurring. For example, medical device manufacturers are under the heavy regulation requirements from both safety and data protection aspects. Thus it is meaningful to discuss factors for deciding trade-offs between a perimeter security architecture and a zero trust architecture. The first trade-off is regarding the three rules of a zero trust model. The implementation of rule 2 and rule 3 can be flexible while rule 1 should be implemented whenever possible. Due to the highest priority of availability, security controls affect system performance should be avoided. The implementation of rule 1 will collect and store all data access and operation logs without put much impact to the system performance. Then the collected logs can used for analysis to detect abnormal behaviors which might indicate an attack.

The second trade-off is regarding the two architectures (perimeter and zero trust). In some cases, an IACS can be divided into sub-systems. The security architecture of individual sub-systems can be designed following different guidelines. Especially considering one example of IIoT scenario, a great number of sensors will be installed to monitor the water quality in a water supply area. In this example, the availability of sensors is robust it can be designed with redundancy (e.g. 2-3 times of the minimum required number). Then the security architecture can be designed following the zero trust model for maximizing security. However, due to a local disaster or a regular maintenance, when number of sensors are reduced to a certain level, the security system should be switched to a mode based on perimeter model, which will maximize the availability of the monitoring system for water supply.

7 Conclusion and Future Work

In this paper, we discussed the actual challenges over the security architecture of IoT solutions. For solving the problems, we compared the perimeter security model and the zero trust model. Three rules were abstracted and discussed within the IoT solution context. In the future, a more detailed IIoT solution can be selected and analyzed using the proposed methodology. It can be expected that more refinements about the trade-offs on security architecture will be discussed. Especially, together with a formal risk assessment, the priority and cost for security mitigation should be quantified for supporting decisions.

Bibliography

[De16] DeCusatis, C.; Liengtiraphan, P.; Sager, A.; Pinelli, M.: Implementing zero trust cloud networks with transport access control and first packet authentication. In: 2016 IEEE International Conference on Smart Cloud (SmartCloud). IEEE, pp. 5–10, 2016.

[Ga19] Gao, Y.; Zid, I. B.; Lou, X.; Parekh, M.: Operational Security Modeling and Analysis for IACS. Gesellschaft für Informatik, 2019.

[GS18] GSMA: Opportunities in the IoT: evolving roles for mobile operators, https://www.gsma.comot/wp-content/uploads/2018/09/New-Roles-for-Operators-in-the-IoT-k.pdf, [Online: accessed 02-August-2020], 2018.

[In18] intersoft consulting: General Data Protection Regulation, https://gdpr-info.eu/, [Online: accessed 02-August-2020], 2018.

[IE15] IEC: 62443 Security for Industrial automation and control systems, Part 3-2: Security risk assessment and system design, 2015.

[KBL18] Kouicem, D. E.; Bouabdallah, A.; Lakhlef, H.: Internet of things security: A top-down survey. Computer Networks 141, pp. 199–221, 2018.

[KS18] Khan, M. A.; Salah, K.: IoT security: Review, blockchain solutions, and open challenges. Future Generation Computer Systems 82, pp. 395–411, 2018.

[Of02] Office for Civil Rights, HHS: HIPAA Privacy Rules, tech. rep., [Online: accessed 02-August-2020], 2002.

[PC04] PCI Security Standards Council: Payment Card Industry Data Security Standard (PCI DSS), tech. rep., 2004.

[Rü15] Rüßmann, Michael; Lorenz, Markus; Gerbert, Philipp; Waldner, Manuela; Justus, Jan; Engel, Pascal; Harnisch, Michael: Industry 4.0: The future of productivity and growth in manufacturing industries. Boston Consulting Group 9, 2015.

[Va18] Vanickis, R.; Jacob, P.; Dehghanzadeh, S.; Lee, B.: Access Control Policy Enforcement for Zero-Trust-Networking. In: 2018 29th Irish Signals and Systems Conference (ISSC). IEEE, pp. 1–6, 2018.

Edge Computing Standardisation and Initiatives

Axel Rennoch [1], Alexander Willner [2]

Abstract: Since Edge Computing (EC) became more important in industry and research several standardisation groups and initiatives are considering related technologies in their strategies and future roadmaps. The work includes the definition of reference architecture models, access interfaces but also addresses edge node autonomy and security aspects. This contribution introduces some basic concepts and common understanding of EC within selected standardisation groups and industrial initiatives. Additionally, technical viewpoints and topics are discussed that are relevant for various communities.

Keywords: Edge Computing; Industry 4.0; Internet of Things; Industrial IoT

1 Introduction

Due to the Edge Computing (EC) paradigm several national and international organizations and interested communities had started to create e.g. initial terminologies, architecture models, reference technology stacks, recommendations and best practices. Currently the list of active groups is growing and it is essential to provide an actual overview in short periods. Therefore, this contribution provides a snapshot for discussion at the IACS workshop in September 2020 for discussions on technical topics and possible harmonisation strategies.

In the following sections dedicated groups and initiatives have been selected due to their actual work for EC. Section 2 address relevant standards developing organisations (SDOs) and section 3 provides a list of further interest groups. Of course, this list is not complete and should only reflect the scope and goals of different stakeholders.

2 Standardisation

The following list is a selection for further discussion.

[1] Fraunhofer FOKUS, Kaiserin-Augusta-Allee 31, 10589 Berlin, axel.rennoch@fokus.fraunhofer.de, https://orcid.org/0000-0003-3419-298X
[2] Fraunhofer FOKUS, Kaiserin-Augusta-Allee 31, 10589 Berlin | TU Berlin, Einsteinufer 25, 10587 Berlin, alexander.willner@fokus.fraunhofer.de, https://orcid.org/0000-0002-8615-4902

doi:10.18420/inf2020_31

2.1 ISO/IEC JTC1

Edge Computing is part of the work within the joint technical committee (JTC) of the International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC) (ISO/IEC JTC1). Since several subcommittees (SC) do address EC here just three important SCs will be emphasized.

First, the projects of the Cloud Computing Standardization Committee JTC1/SC38 have to be mentioned, in particular the technical report TR 23188 [Iso20a] about the EC landscape gives an overview on Cloud Edge Computing including e.g. architectural foundation, relationships to IoT, Cloud, Smart Infrastructures, Access and Service Networks, HW Capabilities, SW Platforms, Virtual Machines, Data Edge Processing, Management and Orchestration. Another SC38 project is related to the advisory group JTC1/JETI (JTC1 Emerging Technology and Innovation) presented by a JTC1/SWG7 white paper about a survey of Fog, Edge, Mist Computing and their relationships among themselves.

Secondly, JTC1/SC41 is working on Internet of Things and related technologies. Technical aspects of edge computing are also under discussion within this technical subcommittee [Iso20b]. And thirdly, in JTC1/SC27 experts are working on Information security, cybersecurity and privacy protection, with high relevance to EC, too.

The new technical report TR 30164 [Iso20c] describes the common concepts, terminologies, characteristics, use cases and technologies (including data management, coordination, processing, network functionality, heterogeneous computing, security, hardware/software optimization) of edge computing for IoT systems applications. This document is also meant to assist in the identification of potential areas for standardization in edge computing for IoT.

Further interesting work from IEC can be retrieved from documents IEC 61499, 62443 and 62541.

2.2 ETSI

At ETSI the Industry Specification Group (ISG) on Multi-access Edge Computing (MEC) is creating a standardized, open environment allowing the efficient and seamless integration of applications from vendors, service providers, and third-parties across multi-vendor Multi-access Edge Computing platforms. Recently the ISG MEC published a white paper [Mec20] about harmonizing standards for edge computing.

Additionally, the ETSI Management and Orchestration (MANO) group is aligning their activities with ETSI Network Function Virtualization (NFV) information models and the ETSI MEC interfaces.

2.3 GSMA

The GSM Association (GSMA) Foundation is not a formal standards body but global member-led organisation representing the mobile industry with impact on international standardisation work. Since operators want to make their assets and capabilities consistently available across networks and across national boundaries they also discuss a unified "operator platform" that will federate multiple Operators' edge computing infrastructure to give application providers access to a global edge cloud.

An Operator Platform Concept whitepaper published by the GSMA [Gsm20] addresses an edge cloud computing concept. Such platform comprises management functionalities and data model for e.g. service availability/roaming, onboarding and instantiation, session & mobility and federation.

2.4 Others

Further standardization groups include the 3rd Generation Partnership Project (3GPP): 5G Specifications, Deutsches Institut für Normung (DIN): 92222 / NA043-01-38AA, Distributed Management Task Force (DMTF): Open Virtualization Format (OVF), Institute of Electrical and Electronics Engineers (IEEE): 1934 / TSN, International Telecommunication Union (ITU): Q.5001 / SG11, Internet Engineering Task Force (IETF): IIoT-SFC-Edge-Computing, National Institute of Standards and Technology (NIST), oneM2M, Object Management Group (OMG).

3 Industrial Initiatives

The following list is a selection for further discussion.

3.1 EECC

In Europe the Edge Computing Consortium (EECC) has been announced and is in preparation [Eec18]. It aims at saving research and development efforts by providing technology stacks for Edge Nodes based on existing, matched components to small, medium and large enterprises for the rising Edge Computing market in smart manufacturing and other Industrial IoT domains.

Goals of this initiative include the specification of a Reference Architecture Model for Edge Computing (ECCE RAMEC), the development of reference technology stacks (ECCE Edge Nodes), the identification of gaps and recommendation of best practices by evaluating approaches within multiple scenarios (ECCE Pathfinders), and the synchronization with related initiatives/standardization organizations and the promotion of the results.

3.2 AIOTI

The aim of the Alliance for Internet of Things Innovation (AIOTI) is to contribute to the creation of a dynamic European IoT ecosystem. Within the AIOTI Strategic Research and Innovation Agenda (SRIA) the research and innovation priorities for the next decade will be stated. Its roadmap [Aio20] is aligned with the goals of the Next Generation Internet initiative proposed under the new Horizon Europe programme to achieve dependable IoT intelligent connectivity by jointly building on edge computing. Thus, SRIA proposes research that extend the industrial edge.

3.3 Others

Further EC working groups have also been initiated or planed within other organizations like e.g. Bitkom e.V. (the EC has been identified some years ago as an important technical trend [He18].), Standardization Council Industrie 4.0 (SCI), Verein Deutscher Ingenieure / Verband der Elektrotechnik Elektronik Informationstechnik (VDI/VDE), Verein deutscher Maschinenbau-Anstalten (VDMA), Zentralverband Elektrotechnik- und Elektronikindustrie (ZVEI), 5G Alliance for Connected Industries and Automation (5G ACIA) in Germany. Further, other activities Europe and its Member States that started discussing the EC topic include the European Processor Initiative (EPI) and GAIA-X in Europe, French Alliance Industrie du Futur, Industria 4.0 (Italy), Przemysł 4.0 (Poland) Industria Conectada 4.0 (Spain), or the Association of Industrial Automation of Ukraine.

Additionally, a number of software around the Linux Foundation have been established, such as Akraino, Baetyl, EdgeXFoundry, Edge Virtualization Engine (EVE), Fledge, Core Infrastructure Initiative (CII), or HomeEdge. Other activities include Kubernetes KubeEdge and the international Eclipse Foundation. The Eclipse IoT Edge working group already contributed a whitepaper on edge security challenges and concerns [Ho19]. There exists also a link to the Industrial Internet Consortium (IIC) that address EC in different task groups and provides several white papers, e.g. an introduction [Iic18] and a recent report that provides a framework, explores architectures, and specifies essential capabilities and interfaces for distributed computing at the edge [Iic20].

4 Conclusion

Currently multiple different aspects of EC are under discussion in various working groups of standardization bodies and industrial associations. The technical viewpoints differ due to the various stakeholders in the organisations and interest groups. Landscape documents already support the interested experts and public community. However, there is still a need for harmonization and common strategies.

Bibliography

[Aio20] AIOTI Strategic Research and Innovation Agenda. AIOTI, Brussels, Belgium, July 2020.

[Eec18] EECC press release, EECC, Berlin, Germany, December 2018.

[Gsm20] GSMA Operator Platform Concept – Phase 1: Edge Cloud Computing. GSMA, London, UK, January 2020.

[He18] Herzog, C.: Technologie Trends - Server, Speicher, Netzwerk. Bitkom e.V., Berlin, Germany, December 2018.

[Ho19] Hopkins, K. et.al.: Edge Security Challenges. Eclipse IoT Edge Working Group, Ottawa, Canada, July 2019.

[Iic18] Industrial Internet Consortium, Introduction to Edge Computing in IIoT, Version 1.0, Milford, MA, USA, June 2018.

[Iic20] Industrial Internet Consortium, Distributed Computing in the Edge, Draft Version (unpublished), Milford, MA, USA, September 2020.

[Iso20a] ISO/IEC TR 23188:2020, Cloud computing - Edge computing landscape, Edition 1.0 2020-02, ISO/IEC, Geneva, Switzerland, February 2020.

[Iso20b] ISO/IEC JTC1/SC41/WG3 WD 30141:2020, Internet of Things and related technologies – Architecture, Ed2 2020-07, ISO/IEC, Geneva, Switzerland, July 2020.

[Iso20c] ISO/IEC TR 30164:2020, Internet of things (IoT) - Edge computing, Edition 1.0 2020-04, ISO/IEC, Geneva, Switzerland, April 2020.

[Mec20] ETSI White Paper #36: Harmonizing standards for edge computing. ISBN No. 979-10-92620-35-5, Sophia Antipolis, France, July 2020.

MAC-layer Security for Time-Sensitive Switched Ethernet Networks

Venesa Watson[1] Prof. Dr. Christoph Ruland[2] Dr. Karl Waedt[3]

Abstract: Security remains a key discussion point for industrial networks within critical infrastructure and Industry 4.0 (I4.0)/Smart Manufacturing infrastructures. While availability remains the chief security requirement for highest safety, integrity protection has become somewhat equal to availability in industry. Common integrity protection mechanisms, however, are not practical for the time-sensitive networks (TSNs) characteristic of I4.0 and critical infrastructures, where the time-critical and mission-critical transmissions cannot be negatively affected by the security overhead. To sufficiently protect and support TSNs, it is necessary to design an integrity protection scheme that provides lightweight security particularly at the OSI MAC-layer where the TSN protocols are defined. The development and testing of lightweight cryptographic algorithms provide one mean by which to achieve such an integrity protection, however, additional steps are needed to design and prove a suitable scheme. TSN-MIC is proposed as a viable scheme for MAC-layer security for TSNs in critical infrastructure and I4.0/Smart Manufacturing.

Keywords: critical infrastructure; industry 4.0; integrity protection; time-sensitive networks; OSI MAC-layer; lightweight cryptography

1 Introduction

Time-Sensitive Networks (TSNs) are those specifications that deploy services such as time synchronization, traffic categorization and traffic shaping, to support time-critical and mission-critical transmissions. For future infrastructures (I4.0/Smart Manufacturing), the Time-Sensitive Networking (TSNg) IEEE 802.1 sub-standards are earmarked as the protocols to provide these listed services. Specifically, OPC UA/TSNg is proposed, where OPC UA is a communication protocol to support interoperability between the interconnected systems – OPC UA is defined in IEC 62451. Similar TSN specifications include ARINC 664 Part 7 - Avionics Full-Duplex Switched Ethernet (AFDX), the IEEE 1722-2016 defined Audio-Video Transport Protocol (AVTP), which like OPC UA/TSNg uses the IEEE 802.1 TSNg sub-standards, and SAE AS6802 Time-Triggered Ethernet (TTE) from TTTech. PROFINET, which is a popularly used industrial protocol, also supports time-sensitive services. These TSNs are compared in [WS18]. Even though these specifications implement

[1] Universität Siegen, Lehrstuhl für Digitale Kommunikationssysteme, Hölderlinstraße 3, 57076 Siegen, venesa.watson@uni-siegen.de
[2] Universität Siegen, Lehrstuhl für Digitale Kommunikationssysteme, Hölderlinstraße 3, 57076 Siegen, Christoph.Ruland@uni-siegen.de
[3] Framatome GmbH, Cybersecurity Concepts & Architecture, Paul-Gossen-Straße 100, 91052 Erlangen

doi:10.18420/inf2020_32

their time-sensitive services in different ways, they share one key similarity. That is, all their time-sensitive services are implemented at the OSI MAC-layer. As such, it is critical to implement dedicated security at the MAC-layer. Traditional security mechanisms implemented at higher OSI layers offer limited to no protection at the lower OSI layers – in that, the design of the OSI model means that protocols at one layer are unaware of issues at another layer, and security at higher layers will not benefit the MAC layer. Additionally, traditional security mechanisms are considered too resource intensive for time- and mission-critical transmissions, especially in safety-critical infrastructures, such as nuclear power plants. With the introduction of lightweight cryptography algorithms, however, the opportunity is presented to develop viable security schemes for TSNs.

Time-Sensitive Network – Message Integrity Code (TSN-MIC) is one such security scheme that offers lightweight integrity protection designed specifically for MAC layer TSN services. Where TSN-MIC further differs from other MAC-layer security schemes (for example, IEEE 802.1X Port-Based Network Access Control, IEEE 802.1AE MAC Security (MACsec) and IEEE 802.10 Standard for Interoperable LAN/MAN Security (SILS)) is in the additional mechanisms that are included for improved security while still observing the performance requirements of time-critical transmissions. TSN-MIC is designed with an online key management and key change-over mechanism, with feedback mechanisms to detect and restrict the propagation of error related to intention and unintentional actions. This scheme is described further in the subsequent sections as follows: Section 2 discusses the TSN-MIC parameters; Section 3 describes how the parameters come together in the overall TSN-MIC concept; Section 4 provides a demonstration of the TSN-MIC scheme; and Section 5 provides a summary.

2 Security Scheme Parameters

The key parameters of the Time-Sensitive Network – Message Integrity Code (TSN-MIC) scheme are the lightweight cryptography algorithms for the generation of the MIC (also Message Authentication Code) and the key management protocol for the key lifecycle processes (key generation, key deactivation, key update and key change-over). For these parameters, only ISO/IEC standards are considered to ensure that non-proprietary, peer-reviewed algorithms from a trusted source are used.

For the MIC generation, ISO/IEC 29192-6 Chaskey-12 [IS17] is used. The Chaskey algorithm that takes an arbitrary-length message (m) that it processes in 128-bit blocks using an n-bit (128-bit) key to create a MAC (τ) of 128 bits or less. The efficiency of the Chaskey algorithm is observed in [Mo15], where the results indicate that Chaskey-8 (Chaskey with 8 rounds) is between 7 to 15 times faster than AES-128-CMAC on the microcontrollers used, while Chaskey-12 (Chaskey with 12 rounds) is 15% slower than Chaskey-8 on the same microcontrollers. TSN-MIC can be implemented with any selected algorithm, however, Chaskey is highlighted here as a viable lightweight cryptography with provable efficiency and security. In [MM14], additional advantages of Chaskey are given:

- **Cross-Platform Versatility** – all microcontrollers do not support variable-length bit rotations and bit shifts, by selecting some rotation constants to be multiples of 8 (i.e. Chaskey-8, Chaskey-16), this limitation is overcome.
- **Dedicated Design** – Chaskey is designed for 32-bit microcontroller architectures.
- **Key Agility** – By generating subkeys into state through simple XOR operation, Chaskey is more efficient than if using a key schedule.
- **Nonces are optional** – an algorithm is susceptible to an attack if the nonce is reused, Chaskey does not require a nonce, therefore avoiding the security issue.
- **Patent-Free** – Chaskey has no known patents or patent applications.
- **Provably Secure** – Chaskey is designed based on an Even-Mansour block cipher based on the permutation operation. The minimum data complexity is $\boldsymbol{D = 2^{64}}$ for Chaskey. A data complexity $\boldsymbol{D}$ below $\boldsymbol{2^n/2}$ avoids chosen plaintext attacks (internal collisions). A time complexity $\boldsymbol{T}$ below $\boldsymbol{2^n/D}$ avoids attacks with a practical time complexity. By restricting the total number of blocks to be processed under one key to 2^{48} blocks, this makes the Chaskey implementation resistant to known plaintext attacks.
- **Resistance Against Timing Attacks** – Chaskey is inherently secure against timing attacks, as the message length determines the total number of cycles.
- **Tag Truncation** – the best attack on Chaskey with short tags is tag guessing, but the algorithm is otherwise robust under tag truncation. The recommended tag size is $|\tau| \geq \mathbf{64}$ for typical applications.

For the key management, the schemes in ISO/IEC 11770 are considered. The ISO/IEC 11770-2 mechanism 6 for key establishment is eventually selected, as it supports key authentication, replay detection, key confirmation, key compromise impersonation and entity authentication [IS16]. ISO/IEC 11770-2 mechanism 6 describes a key generation mechanism between an initiating system and a responding system, with parameters (keying material, identifier and random number) critical to generating matching key pairs and for assuring the integrity of the messages exchanged between these two systems. Analysis of this scheme suggests that the parameters could be vulnerable to manipulations that could disrupt the key generation process and go undetected. As such, the ISO/IEC 11770-2 mechanism 6 is expanded to design a more robust key lifecycle management protocol. Fig. 1 and Fig. 2 illustrate the changes between the original ISO/IEC 11770-2 mechanism 6 and the TSN-MIC mechanism based on the former. In the TSN-MIC mechanism, the major changes include:

- Use of the Random number parameters as message tags to assure message integrity (steps 2 and 3).

- Use of a challenge-response mechanism to ensure that the matching key pairs are generated (steps 4 to 7).
- Online key update and key change-over which monitor key usage and triggers the initiating system to first checks for (and, where necessary triggers the generation of) a successor session key (*next* session key to replace the *current* session key), and the swaps the session keys (steps 8 and 9).
- A threshold for monitoring successive mis-matched MIC in the MIC verification process (step 10).
- Key revocation for expired and/or unusable key pairs (step 11).
- Integrity check for messages transmitted between the initiating system and the responding system (step 12)

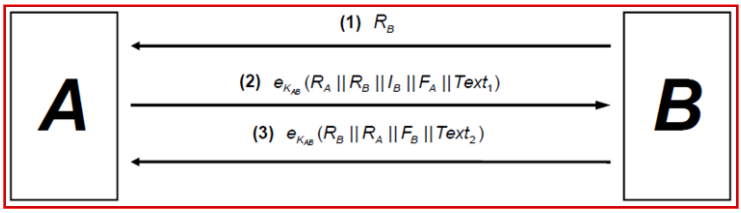


Fig. 1: ISO/IEC 11770-2 mechanism 6

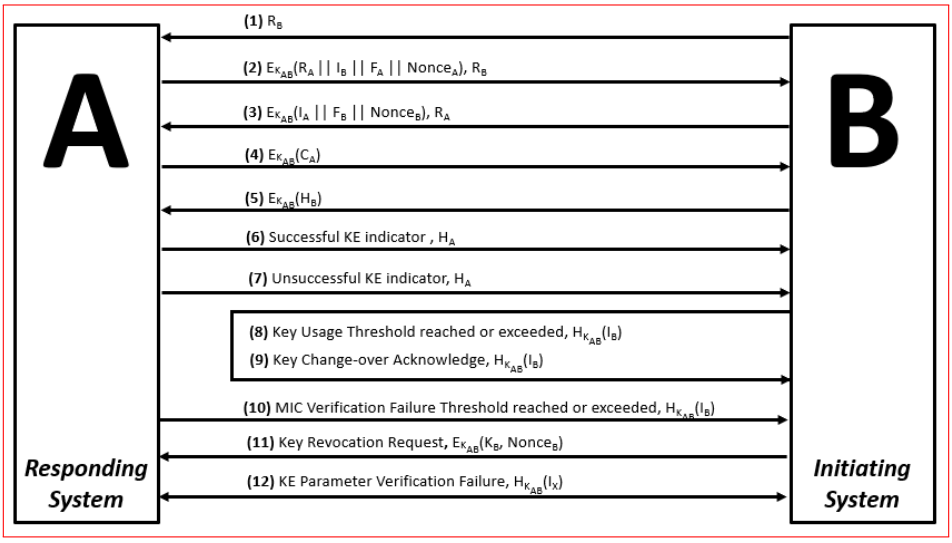


Fig. 2: TSN-MIC-based ISO/IEC 11770-2 mechanism 6

Both the use of Chaskey-12 and the TSN-MIC-based key management mechanism serve to ensure the efficiency and security of the overall TSN-MIC scheme. In the following section, the TSN-MIC is further described to illustrate how the given parameters are leveraged.

3 Security Scheme Concept

As noted, the TSN-MIC scheme is implemented at the OSI MAC-layer, just below the services/operations of the selected TSN protocol (e.g. AFDX, AVTP, TTE, etc.). Fig. 3 gives a depiction of this placement of the TSN-MIC operations (TSN-MIC_OP) below the TSN operations (TSN_OP). This design means that messages are first processed by the TSN-MIC operations before the messages are handled by the TSN operations- in that, for example, TSN-MIC operations for MIC verification is determined and non-compliant messages are dropped before the effort of the TSN operations are wasted. A more detailed description of the TSN-MIC operations is given below.

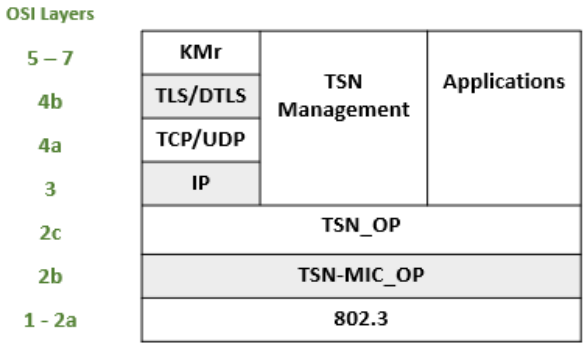


Fig. 3: TSN-MIC protocol stack

The TSN-MIC operations across a simple network (a Source End System, a TSN Switch and a Destination End System), there are seven (7) calculations. Namely, there are two main calculations that are repeated – that is three (3) Long Hash Calculations (LHCs) and four (4) Short Hash Calculations (SHCs). The LHCs are so-called as the hash (unkeyed) output of this calculation is always calculated over the payload of the frame, which is normally between 46 to 1500 bytes. The SHCs are so-called as the MIC (keyed) output is always calculated over the hash output of the LHCs, which is set at 16 bytes in this description of TSN-MIC but can be larger (recommended) or smaller. Across the *simple network*, these seven (7) calculations are as follows and illustrated in Fig. 4:

- **MIC Generation at Source End System**

  The ES generates frames that is processed by the TSN-MIC operation - one LHC followed by a SHC:

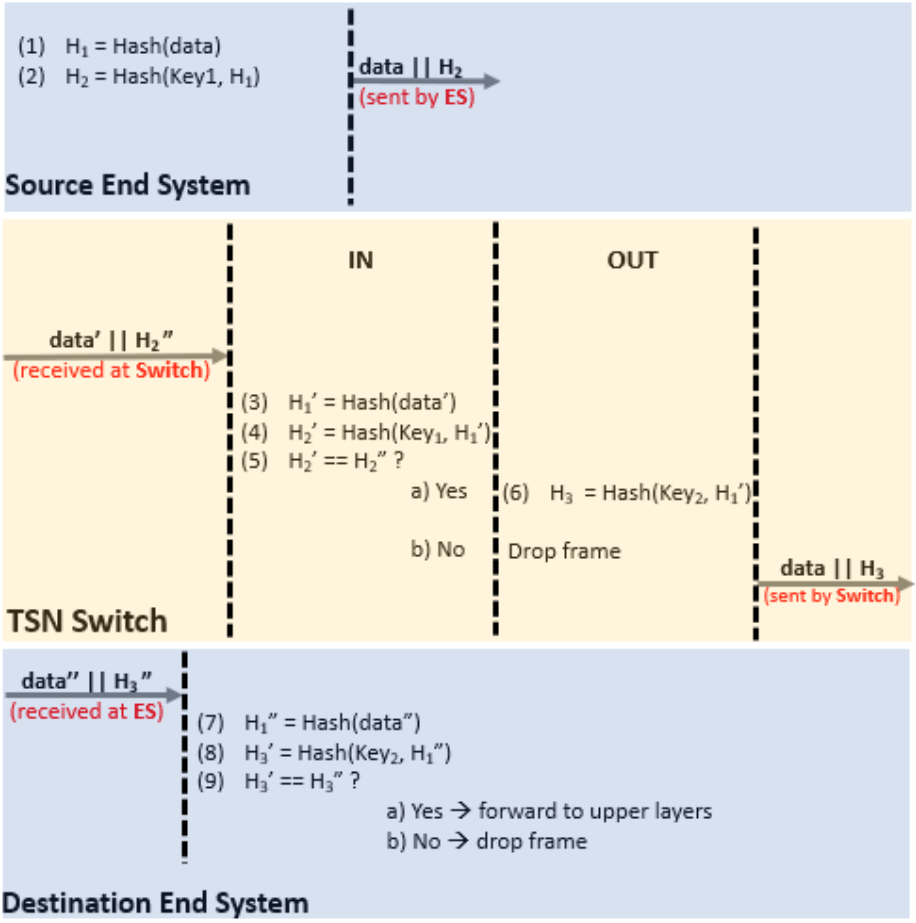


Fig. 4: TSN-MIC operations across a simple network

1. LHC: Generate an unkeyed hash over the data to output the hash H_1
2. SHC: Generate a keyed hash using the session key (Key_1) over the hash H_1 to output the MIC H_2

- **MIC Verification and Re-Generation at the Switch**

  At the TSN Switch, there is both a verification operation and a re-generation operation, involving one (1) LHC and two (2):

3. LHC: Generate an unkeyed hash over the data to output hash H_1'

4. SHC: Generate a keyed hash using the session key (Key1) over hash H_1' to output MIC H_2'
5. If MIC H_2' == MIC H_2", perform calculation (6), otherwise, drop the frame
6. SHC: Generate a keyed hash using the session key (Key_2) over hash H_1' to generate MIC H_3

- **MIC Verification at Destination End System**

  The TSN-MIC calculations at the Destination ES mirror that at the Source ES, but with the inclusion of a verification step:

7. LHC: Generate an unkeyed hash over the data to output hash H_1"
8. SHC: Generate a keyed hash using the session key (Key2) over hash H_1" to generate MIC H_3'
9. If MIC H_3' == MIC H_3", transmit data to upper protocol layers, otherwise, drop data

The initial session keys used to generate the MIC between the communicating systems are manually included, with subsequent keys being generated through the TSN-MIC key management mechanism. Each End System and TSN Switch is configured with two initial session keys (one for incoming messages and one for outgoing messages) that are used for calculating the MIC over the frame payload. Further, each End System and TSN Switch is configured with a master key that is used for the key establishment procedures to provide confidentiality for the keying material that is shared between the negotiating systems. The MIC calculations are done on a per link basis so that incoming messages and outgoing messages are respectively processed with the link-specific session keys. This means that the session keys and master keys can be distributed in a way to avoid the n^2 key distribution problem. For instance, a single session key or master key is not used for the entire end-to-end transmission of a message. At each traversing system, a different key is used. Therefore, each system does not need to have a copy of every key currently in use across the network, which eliminates the n^2 key distribution problem. Fig. 5 gives an illustration of how session keys and master keys are distributed in a TSN-MIC system.

With the parameters and concept determined, the next step is to implement and test the TSN-MIC scheme to assess its viability. This is discussed in the next section.

4 Security Scheme Implementation and Simulation

To test the concept, a software implementation was completed on a Banana Pi, while the efficiency of the TSN-MIC scheme was demonstrated using the OMNeT++ simulator. For the former, the Banana Pi was configured as a TSN Switch, modelled after the AFDX

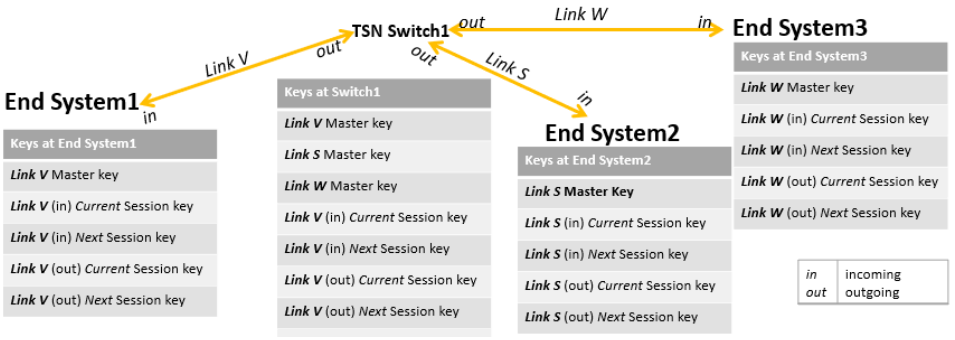


Fig. 5: Example TSN-MIC key distribution

specification. A laptop was then configured as an AFDX End System (ES), serving as both the source ES and the destination ES. Both devices were then configured to conduct the TSN-MIC operations as described earlier where the long hash calculations (LHCs) and short hash calculations (SHCs) are featured, as well as the key management protocol messages. This *simple network* is then activated to exchange the messages and to process them with the TSN-MIC operations (where appropriate, for example, key establishment messages are not processed by the TSN-MIC LHCs and SHCs). Outputs of this simple network in operation are given below. Fig. 6 gives a Wireshark® screenshot these messages being exchanged between the Banana Pi TSN Switch and the End Systems (laptop). Fig. 7 to Fig. 8 are example outputs of the key establishment process, while Fig. 9 shows messages that demonstrate the key update and key change-over process. In the former, the toggle bit is a reserved portion of the SHC MIC output that is flipped between “0” and “1” to indicate a session key update and change-over at the initiating system and to trigger the same at the responding system.

No.	Time	Source	Destination	Protocol	Length	Info
295	461.215612	192.168.178.92	192.168.178.90	UDP	502	2000 → 1045 Len=444
297	463.213079	192.168.178.92	192.168.178.90	UDP	502	2000 → 1045 Len=444
299	465.221103	192.168.178.92	192.168.178.90	UDP	502	2000 → 1045 Len=444
300	467.232230	192.168.178.92	192.168.178.90	UDP	502	2000 → 1045 Len=444
305	469.239657	192.168.178.92	192.168.178.90	UDP	502	2000 → 1045 Len=444
307	471.255473	192.168.178.92	192.168.178.90	UDP	502	2000 → 1045 Len=444

Fig. 6: TSN-MIC message transmission between TSN Switch and End Systems

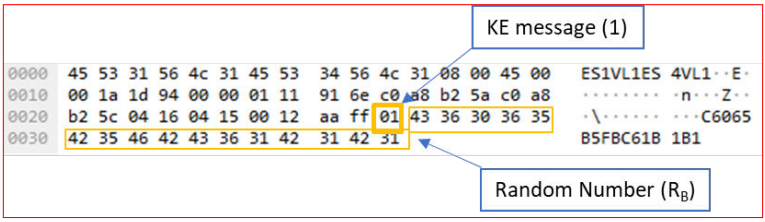


Fig. 7: TSN-MIC key establishment message (1)

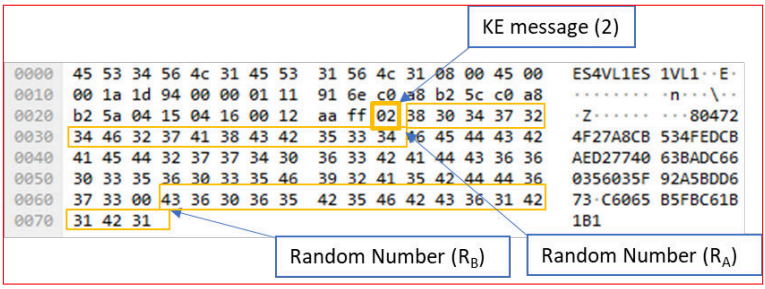


Fig. 8: TSN-MIC key establishment message (2) - unencrypted

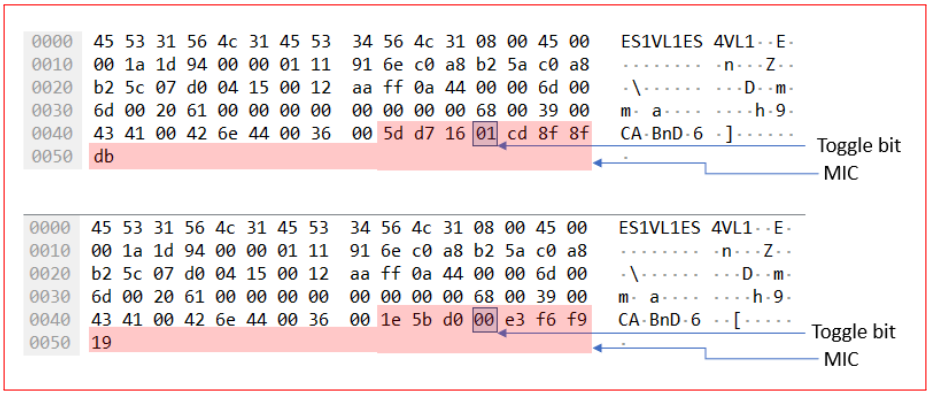


Fig. 9: TSN-MIC Key Update and Key Change-over with Toggle bit update

An OMNeT++ AFDX simulation was taken from [VH12] and modified to model the TSN-MIC design – in that, modules were added to represent the TSN-MIC LHCs and SHCs. Initially, theoretical values are derived to demonstrate the network efficiency and used as a benchmark to assess the efficiency of the TSN-MIC scheme in the OMNeT++ model. For this, the results for the performance of Chaskey-12 is taken from [MM14] and used as a reference point for the theoretical values. [MM14] demonstrates that the performance of Chaskey-12 is 10.5 cycles/byte (speed) and 84×10^6 cycles/second (processor speed) on the ARM Cortex-M3/M4; and at 25.4 cycles/byte (speed) and 48×10^6 cycles/second (processor speed) on the ARM Cortex-M0. In bytes per second, these performance values are respectively 8.0×10^6 on ARM Cortex-M3/M4 and 1.9×10^6 on ARM Cortex M0. In the TSN-MIC LHC, frames between 46 bytes and 1500 bytes are processed, while the TSN-MIC SHC will only ever process 16 bytes (maximum Chaskey MIC output). Based on [MM14], Chaskey-12 would then process 46 bytes in 5.98 μs to 24.38 μs, and between 195 μs to 795 μs for 1500 bytes. Tab. 1 gives the theoretical performance of TSN-MIC based on these values. Based on the theoretical calculations, on the ARM Cortex-M3/M4, the expected TSN-MIC delay is 8.06 μs to 197.08 μs for messages between 46 bytes to 1500 bytes, while on an ARM Cortext-M0, the TSN-MIC delay is 32.86 μs to 803.48 μs.

From this, it can be deduced that there is an overall expected increase of 1% to 35% in the transmission time when a message is processed by a combination of the TSN-MIC LHC and SHC as compared to a message of the same size being processed by a single Chaskey-12 calculation (Tab. 2).

Tab. 1: Theoretical performance of TSN-MIC calculations

Platform	Time per byte	TSN-MIC delay for 46-byte frame	TSN-MIC delay for 1500-byte frame
TSN-MIC Long Hash Calculations			
ARM Cortex-M3/M4	0.13 μs	5.98 μs	195 μs
ARM Cortex-M0	0.53 μs	24.38 μs	795 μs
TSN-MIC Short Hash Calculations			
ARM Cortex-M3/M4	0.13 μs	2.08 μs	2.08 μs
ARM Cortex-M0	0.53 μs	8.48 μs	8.48 μs

Tab. 2: Theoretical performance of TSN-MIC vs Single Chaskey-12 calculation

Platform	Chaskey-12 delay	TSN-MIC delay (1 LHC + 1 SHC)	Percentage Change
Ethernet frame of 46 Bytes			
ARM Cortex-M3/M4	5.98 μs	8.06 μs	+35%
ARM Cortex-M0	24.38 μs	24.38 μs	+35%
Ethernet frame of 1500 Bytes			
ARM Cortex-M3/M4	195 μs	197.08 μs	+1%
ARM Cortex-M0	795 μs	803.48 μs	+1%

The efficiency of the TSN-MIC security scheme is then assessed using the OMNeT++ AFDX model. First, the *QueryPerformanceFrequency* function is used to observe the average processing times TSN-MIC LHC and TSN-MIC SHC with the underlying Chaskey-12 algorithm. The average time taken for a TSN-MIC LHC is observed to be 26.6 ms of messages of 128 bytes, and 19.4 ms for a TSN-MIC SHC for messages of 16 bytes. Next, the OMNeT++ model was observed to determine the time taken for an end-to-end delivery, which is given as 10.88x10-6 simsec (1 simsec $\approx$ 4.867 seconds) or 53 ms in the real world. Using the output of the *QueryPerformanceFrequency* function and the OMNeT++ end-to-end delivery duration, a theoretical efficiency of the TSN-MIC security scheme on an AFDX network can be assumed. The theoretical performance considers the TSN-MIC operations at each AFDX component. The assumed impact of the TSN-MIC security scheme is given as 157.4 ms ($3 \cdot 26.6$ ms (LHC) + $4 \cdot 19.4$ ms (SHC)) over a simple network. This theoretical delay is 0.0339 simsec and represents a percentage in-crease of 312% above the above $10.88\text{x}10^{-6}$ simsec baseline.

The OMNeT++ model modified to create a TSN-MIC based model is then executed to determine the overhead of the TSN-MIC calculations. The outputs from the *QueryPerformanceFrequency* function are converted to simsecs and integrated into the *SendDelay()* function of the OMNeT++ model to assess the added delay. The delay caused by the verification step at the TSN Switch and destination ES is assumed to be negligible, and as such, is not considered here. The execution of the OMNeT++ TSN-MIC simulation shows that the actual end-to-end delay for a single message is $11.84 \text{x} 10^{-6}$ simsec, an increase of 8.82% above the baseline of $10.88 \text{x} 10^{-6}$ simsec. To determine if and how this delay overhead changes as the size of the messages also changes, additional averages were calculated for messages of sizes ranging from 40 bytes to 1500 bytes as shown in Tab. 3. It is observed that the overall increase in the simulation time (simsec) is much greater (2,032%) than the change in the TSN-MIC LHC processing time (111%). As the TSN-MIC SHC will always operate over the same size data, there is no change expected in the processing time, and as such, is not considered in this comparison.

Tab. 3: OMNeT++ efficiency versus TSN-MIC efficiency with increasing message size

Message Size (bytes)	TSN-MIC LHC average processing time (ms)	OMNeT++ processing time (simsec)
40	22.8	$3.84 \text{ x } 10^{-6}$
150	25.4	$12.64 \text{ x } 10^{-6}$
300	29.7	$26.64 \text{ x } 10^{-6}$
450	31.3	$36.64 \text{ x } 10^{-6}$
600	33.4	$48.64 \text{ x } 10^{-6}$
750	36.3	$60.64 \text{ x } 10^{-6}$
900	39.6	$72.64 \text{ x } 10^{-6}$
1150	44.3	$92.64 \text{ x } 10^{-6}$
1500	48.1	$119.36 \text{ x } 10^{-6}$
Percentage increase	+111%	+2,032%

Considering the actual 8.82% increase in transmission duration for an end-to-end transmission (3 LHCs and 4 SHCs) across a simple network, where it is assumed that the seven (7) calculations are equal, then per pair of TSN-MIC calculations (1 SHC and 1 LHC), the percentage delay incurred is 2.52% on average, which falls within the lower percentile of the theoretical range (1% to 35%) as given in Tab. 2. This indicates that the TSN-MIC scheme functions within the theoretical range. However, the differences in the testing environments (microcontroller versus laptop with OMNeT++) and implementation (hardware versus software) means that a true 1:1 comparison is not feasible. However, the OMNeT++ performance does indicate that the overall impact of the TSN-MIC security scheme on network performance is less significant than the impact of the frame payload size. Therefore, the integration of the TSN-MIC scheme should not negatively impact the time- and mission-critical services of the selected time-sensitive network (TSN).

5 Conclusion

The TSN-MIC efficiency based on the OMNeT++ shows an increase in the transmission time of 8.82% for each message from source End System to a destination End System, traversing a single TSN Switch. This indicates a 2.52% delay per pair of TSN-MIC calculations (1 LHC and 1 SHC). However, with the limitations of the testing environment, where more accurate solutions/environments are used, such as with comparable microcontrollers or with an FPGA, an accurate representation of the efficiency TSN-MIC security scheme can be obtained. Nevertheless, TSN-MIC demonstrates viability for I4.0/Smart manufacturing and critical infrastructure.

Bibliography

[IS16] ISO/IEC: ISO/IEC 11770-2:2016 Information Technology – Security techniques – Key management – Part 2: Mechanisms using symmetric techniques, 2016.

[IS17] ISO/IEC: ISO/IEC CD 29192-6 Information technology -- Security techniques -- Lightweight cryptography -- Part 6: Message authentication codes (MACs), 2018.

[MM14] Mouha, N., Mennink, B., Van Herrewege, A., Watanabe, D., et. al.: Chaskey: A Lightweight MAC Algorithm for Microcontrollers. Selected Areas in Cryptography -SAC 2014, Lecture Notes in Computer Science 8781, A. Joux and A. Youssef, Springer-Verlag, 2014.

[Mo15] Mouha, N.: Chaskey: a MAC Algorithm for Microcontrollers – Status Update and Proposal of Chaskey-12. https://hal.inria.fr/hal-01242648/document. [Research Report] Inria Paris Rocquencourt, 2015.

[VH12] Varga, A., and Hornig, R.: Avionics full-duplex switched Ethernet model for OMNeT++, https://github.com/omnetpp/afdx, 2017.

[WS18] Watson, V. and Sassmannshausen, J.: Time-sensitive Ethernet Technology for next Generation CPS/I4.0, GIACM Workshop I4.0, Berlin, September 24, 2018.

R. Reussner, A. Koziolek, R. Heinrich (Hrsg.): INFORMATIK 2020,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2021 351

Gossip protocol approach for a decentralized energy market with OPC UA client-server communication

Josef Schindler[1], Asmaa Tellabi[2], Karl Waedt[3]

Abstract: Gossiping is a well-researched protocol that enables decentralized information sharing. Being comparable to viruses spreading in a biological population, such concepts of data sharing are also called epidemic protocol. Without wanting to be impious with respect to recent pandemics, we propose its usage to facilitate a peer-to-peer (P2P) market for sharing energy between flexible loads or generation units, respectively. Gossip algorithms have been proposed several times in the context of power sharing in transmission grids. Our main contribution is the integration of such scenario with OPC UA. Comprising security by design, good interoperability attributes, several, well-maintained stack implementations and a widespread usage in automation, it reveals to be an outstanding framework for the proposed use case that will be explained in the first sections. After describing underlying physical models and the setup scenario, we will compare the results of the scenario that was conducted on non-OPC UA modules and an OPC UA implementation. Mostly, the performance is questioned at the comparison, still some beneficial concepts of OPC UA can be highlighted in the conclusion: Security controls can be added to the system at the Application Layer where Attribute Based Access Control (ABAC) can be performed, which allows a fine granularity of privileges expressed for subjects (agents in the gossiping algorithms) and objects (energy related assets) via semi-formal security policies. Additionally, UA Discovery service allows for plug and play availability. Concluding, a framework for a very efficient large-area algorithm is presented here to be researched in further work.

Keywords: gossip; peer-to-peer; decentralized energy market; OPC UA

1 Introduction

In this section, we provide basic information and a literature overview, first on the here implemented gossip protocols in general (section 1.1), second on the used communication protocol, namely OPC UA (section 1.2). and third on energy markets focusing on former use-cases of gossiping algorithms (section 1.3). In section 2 the consensus finding between two peers (section 2.1) and the formal description of the cost functions (section 2.2) for both – in section 3 described – implementations is given. Their test results are compared in section 4, before a short conclusion with an outlook is given in section 5.

[1] Friedrich-Alexander-Universität Erlangen-Nürnberg, Faculty of Engineering, Institute of Electrical Energy, Systems Cauerstr. 4, Erlangen, 91058, josef.s.schindler@fau.de
[2] University of Siegen, Faculty of Science and Engineering, Chair for Data Communication Systems, Hölderlinstraße 3, Siegen, 57068, asmaa.tellabi@uni-siegen.de
[3] Framatome GmbH, Erlangen, Germany, karl.waedt@framatome.com

doi:10.18420/inf2020_33

1.1 Gossip Protocol

Gossip algorithms are part of the wide class of distributed algorithms. They were created based on the concept of epidemic. The creation of such a class was a founded on the spreading of virus among populations, based on the idea that one source can infect more entities which subsequently become sources themselves [Kr11]. For computer systems, the notion of viruses is replaced by information that should be exchanged with other nodes in a network. During each communication round, nodes choose and exchange information with one or more of their surrounding nodes. Before the next round begins, the information is managed locally and modified according to measurements taken locally in order to modify the information value that has to be exchanged [KH15].

Fig. 1 depicts the basic concept of the information spreading inside a simple network. The nodes with the information are shown in orange. In each round they pick a random neighbour to hand the information over to him (compare vectors). The more the information is spread, the more it is likely to address a node that already has the information, as seen, when the vectors are pointing to an orange node.

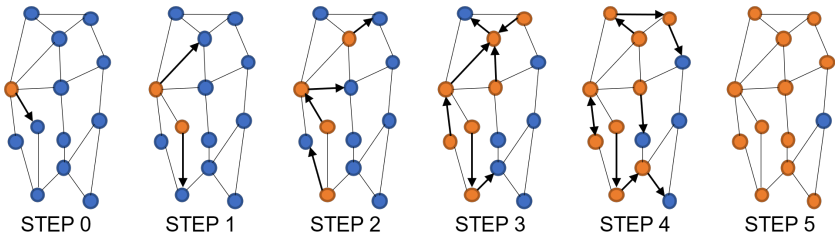


Fig. 1: Basic concept of gossiping algorithms with 5 rounds of information spreading

Gossip protocols are well-known in the research field especially in the computer science areas. They are a promising method that can be used to manage technical issues that occur in distributed systems since they are easy to implement and are able to detect and react to failures rapidly. In a basic gossip protocol implementation, each node chooses a partner randomly to send its current observed state [Al07]. Recently, more research on gossip protocols have been conducted especially for information processing in sensor networks [Di10]. In many distributed computer systems, such as cloud computing and peer-to-peer (P2P) computing systems, the rapidity and robustness characteristics that gossip protocols provide as well as their simple concepts and the absence of central management make them an interesting solution for future implementations. Gossip algorithms are useful for distributed systems for the following reasons [Je11]:

- Creating new protocols: gossip protocols are easy to implement, quick and robust. These characteristics are useful for sharing and collecting information as well as for their processing. These tasks are essential for distributed systems.

- Advanced research about security attacks: With the growing use of the Internet, cybersecurity attacks are becoming more and more sophisticated. Usually, advanced malicious attacks use infected computers that are grouped into networks known as botnets, which can perform coordinated cyber-attacks. Gossip protocols are used to understand and simulate such attacks to find more performing countermeasures.

At first, gossip algorithms were applied for the quick duplication of database updates. Conversely, they also tend to be seen as a potential application for data collection and the processing of information in sensor networks. After that, they were used for solving peer sampling issues, group and topology management, fault detection and other rising technical issues in communication. On the other hand, they can be implemented in other areas where information has to be diffused in distributed computation systems. This was the motivation behind new research dedicated to the implementation of gossip algorithms in future power systems and smart grids [KH15].

Three types of gossip algorithms – random, broadcast and geography – are presented in [Mi18] and their behaviour is sketched in Fig. 2. In random gossip protocols, a node chooses randomly a direct neighboured node to share information. Whereas geography gossiping is not limited to one hop. In the third type, the broadcast gossip, a node spreads its information to all its neighbours. Fig. 2 shows the node with the information to be spread in orange, all possible targets in green and the path to one (random, broadcast) or several (broadcast) targets with vectors. The approach in this paper uses geography gossip style, i.e. each node can request any other node of the network (compare section 3)

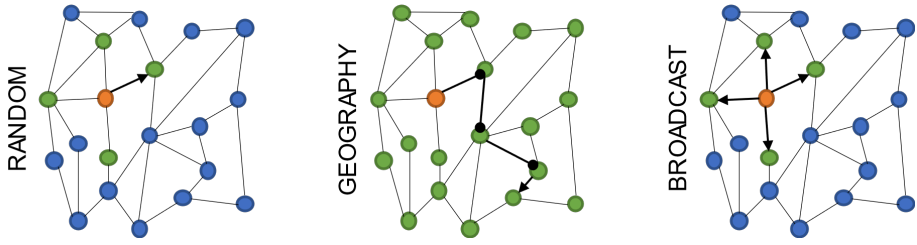


Fig. 2: Different types of gossiping protocols - random, geography and broadcast

1.2 OPC UA

OPC UA is a standard used for exchanging data between industrial systems. In 2011, OPC UA has been standardized by IEC as IEC 62541, which guarantees a better understanding and precision in the specification and an increasing acceptance and implementation within the industry. In addition, OPC UA takes into consideration requirements related to each industry's needs within its specifications and uses its extensible information modelling framework to create new industry specific information models. OPC UA provides information modelling that includes security features built in the design, it offers access rights controls, as well as the communication stack to deliver plug-and-play, machine-to-machine communication for plants and factories [SP20].

In [OP18], it is expected that the usage of OPC technologies will increase nearly by 45 % annually within the next five years. This increase of usage can be explained by the integration of OPC UA technology within smart automation devices and into new market segments like traffic signal systems in Germany. OPC UA can grow exponentially outside its main market segment, which is the automation market in different other markets, counting building automation, medical, telecom, data centres, warehouse, and transportation [OP18].

Message Queue Telemetry Transport (MQTT) is another promising communication protocol that is founded on a subscription-publishing concept, where publishers send messages to a server which is responsible of forwarding messages to subscribers in order to avoid point-to-point connections between subscribers and publishers. The subscribers do not necessarily need to know from where the information came, and which entity is subscribed. MQTT is created to be integrated in low-cost devices with limited resources, which have low-bandwidth networks and high latency [MU20]. The Constrained Application Protocol (CoAP) is a protocol dedicated to IoT architectures that is defined in RFC 7252. CoAP is a protocol with low overhead that is dedicated to resources constrained devices, such as microcontrollers, and networks. This protocol is used in M2M data exchange and has multiple similarities to HTTP [DZ20]. OPC UA is a wide architecture where the communication protocol is only a small part of it. An application that used OPC UA can see all network nodes, methods, and data structures. As stated before, MQTT and CoAP were designed to be lightweight and for smaller systems, which is not the case in industrial systems. Therefore, OPC UA was chosen for this implementation. OPC UA architecture already includes services such as UA Discovery, Pub/Sub, it also integrates security services in its specification and design.

A major challenge of Industry 4.0 and the industrial Internet of Things (IIoT) is to provide a secure, standardized data exchange between devices, machines, and services, coming from diverse industries [Mu20a]. In April 2015, the Reference Architecture Model for Industry 4.0 (RAMI 4.0) has recommended the IEC 62541 standard related to OPC UA [Fr13] as the only communication stack to implement the communication services. One of the requirements needed to integrate OPC UA inside the industrial 4.0 communication layers is a network based on the Internet Protocol (IP). The data exchange in OPC UA is based on 2 different mechanisms:

- A client-server model in which OPC UA clients use the services an OPC UA Server is providing.
- A publisher-subscriber model in which an OPC UA Server makes some information available to a certain number of receivers.

The OPC UA framework is suitable for small devices as well as to technologies used at the IT level such as the cloud computing. In 2012, the Fraunhofer Institute Lemgo reduced the size of the OPC UA Server to a 10 kB footprint, so it can be implemented into small size sensors. Framatome GmbH is one of the multiple companies who choose to integrate the

OPC UA Server into the sensor of monitoring devices for valves as well as their electric actuators. Framatome GmbH uses this solution in the nuclear industry to monitor distributed critical systems (without connection to the internet). For data reliability, built-in security features and interoperability aspects forming the essential parts behind the success of OPC UA technology see [SP20].

Since 2008, Beckhoff and Siemens supported the integration of the OPC UA standard and its implementation into their products. Currently, nearly all Programmable Logic Controller, visualization and Manufacturing Execution System-manufacturer support this standard. The OPC UA specifications were published as multipart standards since 2015 and the OPC communication stack was available as an open Source library for evaluation purposes in 2016. For the professional integration of OPC UA into specific industrial use-cases, some companies provide software toolkits and consulting services for platform and product specific implementations [SP20].

Besides commercial stacks, there are various open source implementations that are compatible with a different range of OS, written in different programming languages and offering a different level of functional coverage of OPC UA via services and features. In [Mu20a] and [Mu20b], authors provided a comparison of the most famous open-source OPC UA implementations, namely, open62541, node-opcua, UA-.NETStandard and FreeOPCUA. One of them, the FreeOPCUA library is used to implement OPC UA clients and servers in this paper.

A major basis for the Industry 4.0 model is interoperability between previous systems with newer technologies [Sc19]. Industrial communication protocols have been developed since the 1980s and currently OPC UA is considered as the main communication protocol that is used for monitoring and controlling purposed in industrial processes [OP10]. OPC UA can be implemented in order to connect industrial controllers in factories to the Internet.

OPC UA can be integrated on platforms that are running on multiple General Purposes Operating Systems (GPOS) such as Windows and Linux. The OPC UA protocol specification is based on the concepts of Address Space Model [OP18] and the services. The Address Spaces specify how objects between servers and clients are going to be represented. An OPC UA Object contains some variables and methods that are implemented as nodes which belong to different node classes inside the Address Model.

1.3 Energy markets

Energy production has developed throughout the last decades, especially in Germany, where windmills and photovoltaic power plants have a share of about 50% nowadays compared to 5% in the beginning of this millennium. The entry of such volatile energy sources makes the market more fluctuating and decentralized. To overcome the higher demand of control due

to the rising number of stakeholders in the market system, many hierarchical approaches have been proposed, e.g. smart-grids, smart-neighbourhoods, virtual power plants.

A concept of cross-commodity sharing such as electricity or heat, is proposed in the DECENT project. Herein, a centralized market on a neighbourhood level uses merit order on a 15 minute-base to facilitate the commodity-sharing. The consensus is then stored using Blockchain technology, so that the agreed trades are reliable and not negligible.

A notable number of scientists has also considered gossiping algorithms for the use in, and for the control and measure of smart grids. Gossiping algorithms were introduced first by [Kr11] for information dissemination in power systems. Furthermore, the authors state the requirements on ICT infrastructure to enable actual implementations. In [Cr17], a gossip-based scheme is used to control and to facilitate a distributed demand response in a virtual microgrid, i.e. the university campus. Checks on the applicability of flow updating for a large set of network nodes is done with a P2P network simulator. Also, [EAD16] proposes a P2P control and gossiping communication for the similar purpose of keeping the voltage of a smart grid within the boundaries. Surplus on that approach is seen in keeping all control local, i.e. leaving a central controller out, which increases the fault tolerance.

In [KH15] and [KH16], the authors propose a gossip-based approach to detect and resolve voltage violations, and to manage demand and thus power flows on radial distribution grids, respectively. Authors in [Le16] and [Mi18] have also used gossip algorithms for power sharing purposes. In two papers [Wa19] and [Wa20], R. Wang et al. first propose a gossip-based algorithm for the economic dispatch in a power grid with transmission losses and secondly research random communication link failures within the same setup. By modelling the system as a Bernoulli network in [Wa20], they try to overcome the communication link failures and make the algorithm able to solve the mathematical problem, i.e. the economic dispatch.

During the last years, even the car manufacturer Tesla launched its own so-called “Autobidder” software. Several forecasting mechanisms and a dispatch optimization are performed to autonomously monetizes battery assets. According the statement [TE20], the software is scalable from a single household prosumer up to the utility level. Such large company working in that domain shows a valid potential in that market section.

In this paper, we will present an approach where we have merged the DECENT idea of having a local market, with the discussed solutions for power sharing using gossip protocols. The result will be an algorithm that finds the market consensus on a neighbourhood-level. Furthermore, we compare an OPC UA stack with a basic implementation to measure their performances and to evaluate beneficial concepts of this communication standard.

2 Consensus finding

The system consists of several flexible prosumers, which are all connected electrically and communicate with each other. In this paper, only the communication network was implemented, with the assumption that the power network is a copper plate with no limitations and packet losses. Additionally, all participants have the same target – to assure the common welfare by producing power at the least overall- or community-cost.

Therefore, every flexible prosumer or entity has its own cost function, i.e. at which price can it produce which amount of energy. The entity runs a server and is continuously listening to requests from other entities to find their common optimum setpoint. If no request is upcoming at the server, the entity tries to request another, randomly chosen entity for a settlement from time to time. Therefore, it sets up an own client.

2.1 Entity-to-entity consensus

The overall process for consensus finding is depicted in Fig. 3 for entity A which sends a request to entity B. To avoid issues with parallel truths, the entity – which is represented by Client A – requests to block its server (see step 1 & 2). If Server A were blocked due to another request the process would stop immediately.

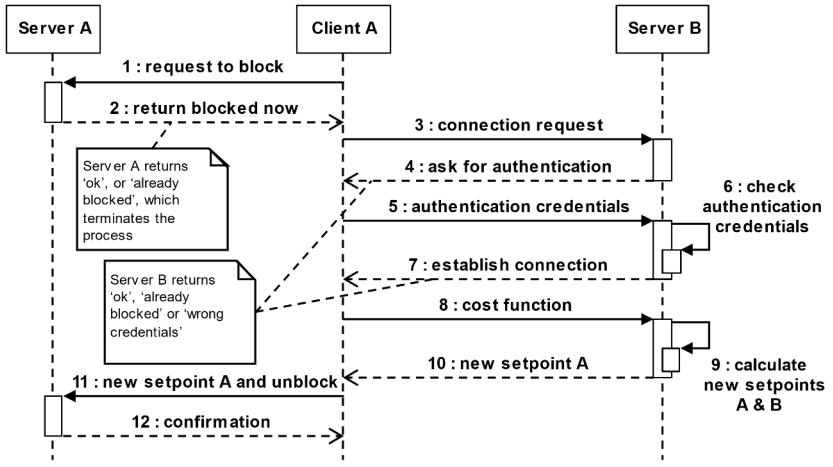


Fig. 3: Communication between OPC UA Server and Client based on the gossiping algorithm

Next, Client A reaches out to Server B to authenticate and establish a connection (see steps 3 to 7). Again, the process stops here if Server B were blocked or if the authentication failed. After a connection is established, Client A sends the cost function details to Server B (step 8). At step 9, Server B calculates the optimum setpoints and sets its own, new power production. The new setpoint A is resent to entity A, communicated to Server A and confirmation about that finalizes the process in steps 10, 11 and 12, respectively.

The optimization calculations in step 9 rely on polynomials, which can be noted in the standard form and a form showing the polynomial roots

$$c\,(p) = \sum_{i=0}^{n} a_i \cdot p^i = b_n \cdot \prod_{i=0}^{n-1} (p - b_i). \tag{1}$$

Hence, all necessary coefficients a_i, the setpoint p_0 and limitations $p_{\min}$ & $p_{\max}$ of entity A are sent in step 8 and will be named $c_A\,(p)$, whereas the cost function of entity B is named $c_B\,(p)$. To gather comparable functions, not the absolute power output p, but the relative change of the setpoint Δp is considered for both polynomials. Therefore, both functions are shifted by p_0:

$$b_i^{'} = \begin{cases} b_i - p_0, & i < n \\ b_i, & i = n, \end{cases} \tag{2}$$

Returning the shifted functions $c_{A,s}\,(p)$ and $c_{B,s}\,(p)$. If entity B raises its power output, entity A would need to lower its output to satisfy the system's demands. Therefore, entity A's polynomial is mirrored at the why axis

$$c_{A,s,m}\,(\Delta p) = c_{A,s}\,(-\Delta p)\,. \tag{3}$$

Afterwards, both functions can be added together:

$$c\,(\Delta x) = c_{A,sh,m}\,(\Delta p) + c_{B,sh}\,(\Delta p) \tag{4}$$

To receive an overall function whose optimum can be found. Differentiation of the polynomial from equation (4) returns all possible optima for Δp, from which all values within the boundaries $[p_{\min}, p_{\max}]$ and the boundaries themselves need to be considered.

2.2 Manifestation of the cost functions

The cost functions resemble the marginal cost to produce a certain amount of power and are specific to the type of the parent entity. To give a clearer vision, Fig. 4 depicts several schematic graphs, how they could look like for real applications. All graphs have been normalised, so the nominal power output equals 1.

The first graph belongs to a photovoltaic power plant. Obviously, the output is limited to 0 and a maximum power that depends on the current solar radiation and thus is usually

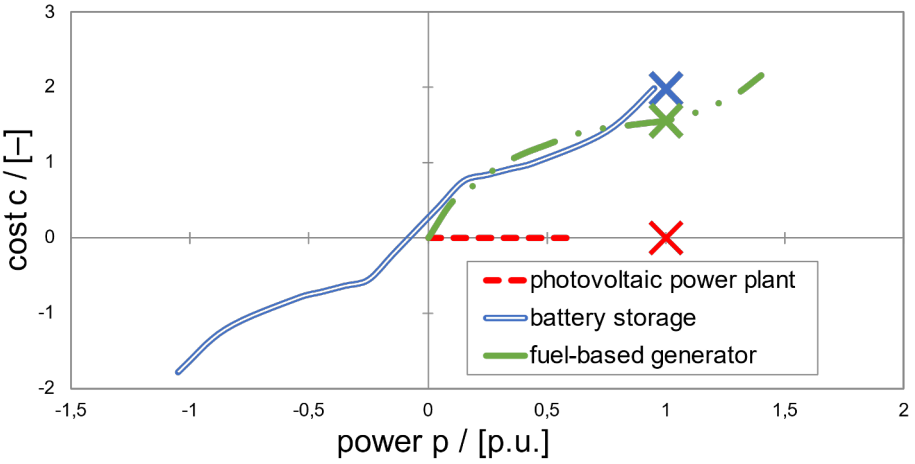


Fig. 4: Example schemes of cost functions of several possible entities

smaller than the nominal power. In between those boundaries, no significant price variation can be observed.

Secondly, a battery storage has a nearly rotatory symmetric shape, but is slightly shifted on both axes e.g. due to losses in power electronics and wear of the cells. Of course, a battery storage even wears out when it does not produce energy – compare calendric degradation. Again, this is a schematic overview. In a real application, a battery's function would be very complex and will depend on the state of charge or health, respectively.

The third graph in Fig. 4 belongs to a fuel-based generator, like a combined heat and power plant or a diesel generator. Considering only the usage of fuel would lead to a linear characteristic. Anyway, bad efficiency in low and greater wear on components in high power domains lead to higher gradients.

One entity can be a single specific flexible producer or consumer, but also a composition of several of them. So, in the virtual setup all parameters are randomly chosen at the beginning and then kept constant, to obtain a static system.

3 Implementation

In this paper two python-based setups with different approaches are compared. One uses a basic client and server (see section 3.1), whereas the other uses the library FreeOPCUA, i.e. OPC UA Server and Clients (section 3.2). Here is a list of the common specification

- Each entity is simulated on its own virtual machine inside one testing computer. Hence, the number of entities is limited to 24 according to the RAM of the computer
- Each entity/virtual machine has a specific IP address and a randomly and uniquely prepared cost function; the setup is given in Tab. 1
- Each entity is aware of the present entities from a locally stored setup file. In a real-life implementation, this could be facilitated via OPC UA Discovery Service or as a side effect of the gossiping protocol, i.e. during consensus finding, every entity shares the location of other known entities
- The delay for one entity to actively request another entity for consensus finding (compare Fig. 3) is chosen randomly within a range $d_{\min}$ and $d_{\max}$
- The topology of the gossiping network is not taken into account; in a real-life scenario, this might be an issue, especially when the power grid needs to be considered like in [KH16]

Tab. 1: setup overview; normalized parameters of all entities

	Entity with IPv4 Addresses: 192.168.56.1<ID>																							
<ID>	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
p_{init}	0.4	0.5	0.6	0.6	0.4	0.6	0.7	0.8	0.9	0.4	0.0	0.3	2.6	1.3	-1.0	0.8	4.3	0.3	0.2	3.3	4.3	-1.0	1.5	-1.0
c_{init}	2.58	1.88	4.12	4.12	3.38	1.82	4.04	4.28	2.03	0.31	3.0	0.2	4.52	-2.45	1.9	-2.21	-1.35	0.69	3.51	1.82	3.3	1.8	3.04	-1.1
p_{min}	-1.1	0.0	-2.0	-2.0	-1.5	0.6	-2.0	-2.8	0.1	-2.2	-1.0	-1.4	-1.4	-2.4	-2.2	-1.3	0.0	-2.2	-1.4	0.0	1.0	-2.0	-1.4	-1.4
p_{max}	1.9	3.0	3.0	3.0	1.5	6.0	1.7	1.8	2.2	1.3	1.0	1.2	8.4	1.4	1.1	0.0	5.0	1.1	1.4	5.0	7.0	3.1	1.6	0.3
a_0	1.0	2.0	4.0	4.0	1.0	2.0	4.0	4.0	2.0	0.5	3.0	0.2	4.0	-5.3	0.9	0.2	-3.1	0.6	3.2	0.3	-1.0	0.2	1.1	0.6
a_1	3.0	-0.5	-0.3	-0.3	5.0	-0.6	-0.7	-0.8	-1.2	-0.4	2.5	0.2	0.2	1.6	-0.8	-5.3	2.3	0.1	1.6	-0.2	1.0	-1.2	0.6	0.3
a_2	2.0	0.5	0.6	0.6	2.0	0.5	0.6	0.8	1.1	-0.3	-5.2	-0.8		0.38	1.0	1.1	-1.3	0.5	-0.3	0.2	0.0	0.6	0.8	-1.1
a_3	1.0		0.4	0.4	1.0		0.7	0.8	0.3	0.2	3.1	0.4		0.06	0.8	2.2	0.2	0.7	0.2			0.2	-1.2	0.5
a_4										0.1		0.2						0.1					0.2	0.2
a_5												-0.1											0.3	
a_6												0.1												

3.1 Basic implementation

The basic implementation uses `socket` from the standard python library. A server is created inside the entities code. While listening, it will block the execution of any other task and is therefore interrupted by `settimeout`. Then a client tries to connect to a randomly chosen server. Every connection is accepted, so no authentication (compare step 4 up to 6 in Fig. 3) is used. The sending buffer needs binary encoded data. So, a list containing the current setpoint, boundaries and cost function coefficients is packed and sent from client to server. Vice versa, the server returns a binary encoded new setpoint.

3.2 OPC UA implementation

The second implementation has a permanently running OPC UA Server at each entity. The python library `sched` is used for timing, i.e. to set the OPC UA Client up, whenever it is time to request a P2P consensus from a randomly chosen other entity.

The proposed authentication was not implemented so far but can be updated in future versions. OPC UA's specification enables such features, independent on the library used for the implementation. Another example on this, is the method call used for steps 8-10 in Fig. 3. UA Namespaces in OPC UA Servers provide all necessary information to OPC UA Clients, so they can browse through it and call certain methods, independent on the programming language or stack provider.

4 Simulation results

Both setups described in section 3 have been simulated several (in overall 72) times, varying the number of entities $N = \{8, 16, 24\}$ and the delay for the next active request for a consensus finding $d_{\min}$ and $d_{\max}$. Each run was finished after 90 seconds. An exemplary cost development of each entity and their sum is depicted in Fig. 5.

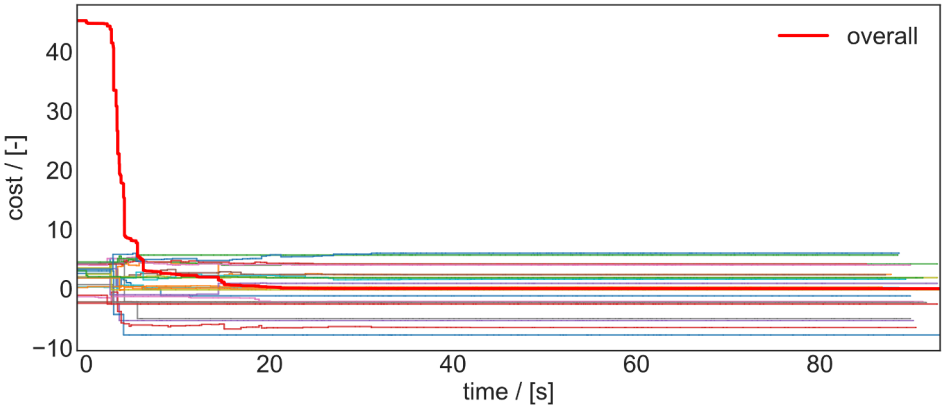


Fig. 5: cost development during one gossip simulation, with: $N = 24$, $d_{\min} = 0.01$sec, $d_{\max} = 0.3$sec, Run = 3

As a measurable result, the time span until the final overall cost are fitted up to 99% t_{99} and the final overall cost c_{end} itself are displayed for each run in Tab. 2. Each testing scenario is processed three times to cover for statistical deviations, the mean values are given in overview Tab. 2.

A pretty good assumption for an optimum can be calculated very efficiently using gossip protocols, but there is no clear statement, whether it is a local or the global optimum. In

Tab. 2: Simulation results from gossiping with varying number of entities, delay, and concept

d_{min}; d_{max}			0.01 sec; 0.3 sec				0.5 sec; 1.0 sec				1.0 sec; 1.5 sec				2.0 sec; 3.0 sec			
Type			basic		OPC UA		basic		OPC UA		basic		OPC UA		basic		OPC UA	
N	c_{init}	Run	t_{99}	c_{end}	t_{99}	c_{end}	t_{99}	c_{end}	t_{99}	c_{end}	t_{99}	c_{end}	t_{99}	c_{end}	t_{99}	c_{end}	t_{99}	c_{end}
8	26.23	1	3.881	16.340	1.121	16.343	4.275	16.343	4.406	13.953	3.496	16.343	6.163	16.343	8.128	16.343	6.032	16.343
		2	2.453	16.343	0.746	16.343	3.649	13.953	2.189	16.343	10.173	13.953	7.175	13.953	7.133	16.343	9.772	13.953
		3	2.342	13.953	1.756	16.343	2.341	16.343	3.312	16.343	7.985	13.953	11.925	16.343	7.782	16.343	19.370	13.953
		Ø	2.892	15.545	1.208	16.343	3.422	15.546	3.302	15.546	7.218	14.750	8.421	15.546	7.681	16.343	11.725	14.750
16	33.53	1	6.754	-2.919	7.225	-2.918	21.605	-2.918	10.940	-2.918	15.686	-2.918	22.971	-2.918	46.042	-2.918	32.889	-2.918
		2	7.081	-2.946	4.487	-2.918	16.779	-2.918	9.042	-2.918	18.532	-2.918	10.754	-2.918	29.510	-2.918	49.206	-2.918
		3	3.711	-2.918	6.047	-2.918	15.615	-2.922	14.719	-2.918	15.687	-2.918	18.104	-2.918	25.971	-2.918	23.767	-2.918
		Ø	5.849	-2.928	5.920	-2.918	17.999	-2.920	11.567	-2.918	16.635	-2.918	17.277	-2.918	33.841	-2.918	35.287	-2.918
24	45.25	1	35.316	-2.209	14.789	-0.061	51.898	-1.084	22.816	0.001	33.859	1.123	7.706	0.469	31.957	-0.042	13.756	0.001
		2	56.017	-1.003	8.152	0.001	5.125	-0.025	8.068	0.469	74.912	-2.479	17.406	0.469	23.103	0.020	28.636	0.016
		3	55.417	-2.503	18.829	0.001	63.858	-0.869	14.640	0.001	80.705	-16.86	16.864	0.001	50.545	0.602	54.278	0.008
		Ø	48.917	-1.905	13.923	-0.020	40.294	-0.659	15.175	0.157	63.159	-6.074	13.992	0.313	35.202	0.193	32.223	0.008

this simulation set, the three optima at $c_{end} = \{13.953, 15.546, 16.343\}$ for $N = 8$ entities prove this behaviour. Neither the basic nor the OPC UA implementation achieved the better optimum more preferably.

Still, at the basic implementation some errors or package losses might have occurred. The first run with $N = 8$ entities and the fastest delay beats both probable optima slightly. Also, when looking at the basic implementation's results for $N = 24$ entities, no clear optimum is visible. There are two possible reasons. First, the overall optimisation function is too complex, so it has that many optima. Second and more likely: the unreliable network connection in the basic implementation leads to unfinished consensus; Consequently, the overall initial power demand p_{init} is violated because requested Server B sets the new consensus setpoint, but entity A does not – no reliable check-up has been specified after step 10 (see Fig. 3). Tab. 3 proves that effect by plotting the summarized power output of all entities at the initial p_{init} and final state p_{end}, respectively.

Tab. 3: development of the power output; significant Standard Deviation (SD) is marked red

d_{min}; d_{max}			0.01 sec; 0.3 sec		0.5 sec; 1.0 sec		1.0 sec; 1.5 sec		2.0 sec; 3.0 sec	
Type			basic	OPC UA	basic	OPC UA	Type	basic	OPC UA	basic
N	p_{init}	Run	p_{end}	p_{end}	p_{end}	p_{end}	p_{end}	p_{end}	p_{end}	p_{end}
8	4.6	1	4.595	4.6	4.603	4.6	4.6	4.6	4.6	4.6
		2	4.601	4.6	4.6	4.6	4.6	4.6	4.6	4.6
		3	4.599	4.6	4.6	4.6	4.6	4.6	4.6	4.6
		SD	0.003	0.0	0.002	0.0	0.0	0.0	0.0	0.0
16	9.9	1	9.902	9.9	9.9	9.9	9.9	9.9	9.9	9.9
		2	10.055	9.9	9.9	9.9	9.9	9.9	9.9	9.9
		3	9.9	9.9	9.901	9.9	9.9	9.9	9.9	9.9
		SD	0.09	0.0	0.001	0.0	0.0	0.0	0.0	0.0
24	21.8	1	21.769	21.892	22.415	21.8	19.324	21.8	22.243	21.8
		2	23.004	21.8	22.198	21.8	26.662	21.8	21.786	21.8
		3	20.302	21.8	18.371	21.8	21.735	21.8	21.223	21.8
		SD	1.11	0.053	2.025	0.0	3.15	0.0	0.42	0.0

In comparison and still considering $N = 24$ entities, OPC UA shows local optima at $c_{\text{end}} = \{0.001, 0.469\}$. They occur at least three times during the simulations. Also, Tab. 3 shows the reliable behaviour in that context. Except one outlining run, no power deviations can be observed between initial and final state.

Despite OPC UA has a fully developed stack on the backend, which could slow down the application, it shows to be equally fast for $N = \{8, 16\}$ entities, and even faster for $N = 24$ entities compared to the basic implementation. Another expected phenomenon only partly occurred: even for the largest set of entities the single computer did not go very slow for the shortest delay at the OPC UA implementation, due to overload of the machine. In contrast, for the basic implementation it did. The time to get close to the consensus decreases when looking at "slower" parameters, i.e. higher delays.

5 Conclusion

The integration of OPC UA into an application can be an easy and a straightforward task. Compared to classic OPC, OPC UA is built on a simplified architecture. In case other features are needed, the same building blocks can be used to design new models without the need to start from the beginning again. The OPC UA protocol is an industrial protocol that is used to communicate between devices within plants and factories.

In the simulation application, the OPC UA stack shows advantages, compared to the basic implementation, although the functionality has not yet been fully exploited. As a future work, we want to research features OPC UA provides, such as UA Discovery or authentication or the more sophisticated ABAC, within this paper's context. Besides that, we intend to implement and do more experiments to study the effect a real time communication protocol such as Profinet can have on the communication.

6 Acknowledgement

Some of the addressed topics are being elaborated as part of Framatome GmbH's participation in the DECENT R&D (2018-2020) with Technical University of Munich (Germany), FORTISS (Germany), IBDM (Germany), FENECON (Germany), VTT Technical Research Centre of Finland Ltd (Finland), Empower IM Oy (Finland), Wirepas (Finland) and Fourdeg (Finland), partially funded by German Ministry BMWi as well as the ABAC R&D with University of Siegen.

Bibliography

[Al07] Alvisi, L. et.al.: How robust are gossip-based communication protocols?. ACM SIGOPS Operating Systems Review, vol. 41, no. 5, pp. 14-18, 2007.

[Cr17] Croce. D. et al.: Overgrid: A Fully Distributed Demand Response Architecture Based on Overlay Networks. IEEE Transactions on Automation Science and Engineering, vol. 14, no. 2, pp. 471-481, 2017.

[Di10] Dimakis, A. G. et.al.: Gossip Algorithms for Distributed Signal Processing. Proceedings of the IEEE, vol. 98, no. 11, pp. 1847-1864, 2010.

[DZ20] Dzone, CoAP Protocol: Step-by-Step Guide, https://dzone.com/articles/coap-protocol-step-by-step-guide, accessed: 23/05/2020.

[EAD16] Engels, J.; Almasalma, H.; Deconinck, G.: A Distributed Gossip-based Voltage Control Algorithm for Peer-to-Peer Microgrids. IEEE International Conference on Smart Grid Communications, Sydney, NSW, Australia, 2016.

[Fr13] Frejborg. A. et al.: OPC UA Connects your Systems – Top 10 reasons why to choose OPC UA over OPC. Finnish Society of Automation, Biannual Seminar, Finland, 2013.

[Je11] Jelasity, M.: Gossip. In (Di Marzo Serugendo G., Gleizes MP., Karageorgos A. ed.) Self-organising Software. Natural Computing Series. Springer, Berlin, Heidelberg, pp. 139-162, 2011.

[KH15] Koukoula, D. I.; Hatziargyriou, N. D.: Convergence Acceleration of Gossip Protocols Applied for Decentralized Distribution Grid Management. IEEE Eindhoven PowerTech, Einhoven, Netherlands, 2015.

[KH16] Koukoula, D. I.; Hatziargyriou, N. D.: Gossip Algorithms for Decentralized Congestion Management of Distribution Grids. IEEE Transactions on Sustainable Energy, vol. 7, no. 3, pp. 1071-1080, 2016.

[Kr11] Krkoleva, A. et.al.: Requirements for Implementing Gossip Based Schemes for Information Dissemination in Future Power Systems. IEEE PES International Conference and Exhibition on Innovative Smart Grid Technologies, Manchester, United Kindom, pp. 1-7, 2011.

[Le16] Lequay, V. et.al.: Flexible Load Shedding using Gossip Communication in a Multi-Agents System. IEEE International Conference on Self-Adaptive and Self-Organizing Systems, Augsburg, Germany, 2016.

[Mi18] Ming, Y. et al.: Distributed Energy Sharing in Energy Internet Through Distributed Averaging. Tsinghua Science and Technology, vol. 23, no. 3, pp. 233-242, 2018.

[Mu20a] Muehlbauer. N. et al.: Open-Source OPC UA Security and Scalability. 25th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA); Vienna, Austria, 2020.

[Mu20b] Muehlbauer, N., et al.: Feature-based Comparison of Open Source OPC-UA Implementations. GI conference – INFORMATIK 2020, Karlsruhe, Germany, 2020.

[MU20] Muutech, Comparison MQTT vs OPC-UA, https://www.muutech.com/en/comparison-mqtt-vs-opc-ua/, accessed: 25/05/2020.

[OP10] OPC Datahub, What is OPC?, https://opcdatahub.com/WhatIsOPC.html, accessed: 14/04/2020.

[OP18] OPC Foundation, OPC Technology Well-positioned for Further Growth in Tomorrow's Connected World, https://opcfoundation.org/wp-content/uploads/2017/11/OPC-UA-Interoperability-For-Industrie4-and-IoT-EN.pdf, accessed: 06/04/2020.

[SP20] SpotLightMetal, IoT Basics: What is OPC UA? https://www.spotlightmetal.com/iot-basics-what-is-opc-ua-a-842878/?cmp=go-aw-art-trf-SLM_DSA-20180820&gclid=Cj0KCQjw2PP1BRCiARIsAEqv-pQ7tSJYVdlJPyXvxcfH_e3k8a_WvCeIldc5iPXFkUZnjk0nnbaEZcIaArWBEALw_wcB, accessed: 14/05/2020.

[TE20] electrek: Tesla has a new product – Autobidder, a step toward becoming an electric utility, https://electrek.co/2020/05/03/tesla-autobidder-new-product-electric-utility/, accessed: 27/06/2020.

[Wa19] Wang, R.; et al.: A gossip-based asynchronous distributed algorithm for economic dispatch problem with transmission losses. IEEE PES Innovative Smart Grid Technologies Asia, Chengdu, China, 2019.

[Wa20] Wang, R.; et al.: A Gossip-Based Distributed Algorithm for Economic Dispatch in Smart Grids with Random Communication Link Failures. IEEE Transactions on Industrial Electronics, vol. 67, no. 6, pp. 4635-4645, 2020.

Feature-based Comparison of Open Source OPC-UA Implementations

Nikolas Mühlbauer[1] Erkin Kirdan[2] Marc-Oliver Pahl[3] Karl Waedt[4]

Abstract: OPC UA is an industry-standard architecture for automation, process controlling and monitoring. It is a detailed and complex machine-to-machine communication protocol which makes it challenging to implement. The complexity of the protocol leads to heterogeneity among implementations. Today, there are several open-source implementations without a compliance certificate accredited by the OPC Foundation. Certified implementations undergo various tests to fulfil interoperability. Every implementation fits different use-cases and requirements as each of them comes with its own features. In this paper, we make a feature-based comparison of the most common open-source OPC UA implementations. We investigate their support for the essential features and functionalities. Furthermore, we evaluate their interoperability. Overall, our study shows that open-source implementations have good coverage of features and functionalities, especially open62541 and UA-.NETStandard. Furthermore, our tests show that they do not have any significant interoperability issue.

Keywords: opc ua; open-source; interoperability

1 Introduction

OPC UA is a machine-to-machine communication protocol widely used in industrial automation [RC18]. It has use cases across the fields of transportation, oil and gas, energy and utilities and automation [OPb]. It is platform-independent and thus can run on various operating systems and hardware platforms. OPC UA can scale down to embedded controllers or mobile devices up to powerful servers controlling a collection of machines. Furthermore, it can also be integrated into cloud platforms [OPc].

OPC UA is used in critical infrastructures. It is developed above the already mature and widely used OPC Classic protocol. OPC UA is composed of detailed and complex series of specifications. The complexity of the protocol leads to variations between implementations. Consequently, OPC UA implementations are having a different level of compliance to the specification. The OPC foundation certificates implementations from various points such as compliance, interoperability, robustness usability and efficiency to control this heterogeneity. Implementations having the compliance certificates can communicate with each other, whereas this is not guaranteed for the non-certified implementations.

[1] Technical University of Munich, n.muehlbauer@tum.de
[2] Covalion, Framatome/Technical University of Munich, erkin.kirdan@framatome.com/erkin.kirdan@tum.de
[3] IMT Atlantique/Technical University of Munich, marc-oliver.pahl@imt-atlantique.fr/pahl@tum.de
[4] Framatome GmbH, karl.waedt@framatome.com

doi:10.18420/inf2020_34

OPC UA is deployed in use cases having different operating systems and hardware together. There is no single OPC UA implementation that fits every use-case. A typical OPC UA scenario is depicted in Fig. 1. Some machines run an OPC UA server that is developed by the machine manufacturer. The manufacturer can also deploy the machine with a third-party OPC UA implementation. Some IT companies provide cloud solutions for OPC UA that are integrated into the scenario. The main server collects data from the machines and makes them available for the office computer for organizational purposes. An implementation written in JavaScript can be preferred to develop a browser-based application for the offices. A C-implementation fits well to the mobile clients as well as embedded controllers due to its efficient resource usage. Implementations written in C# or Python can be preferred in the main server because of their languages. Principally, several OPC UA implementations should be able to run together in such a scenario. However, despite the commercial ones, open-source implementations mostly lack certificates.

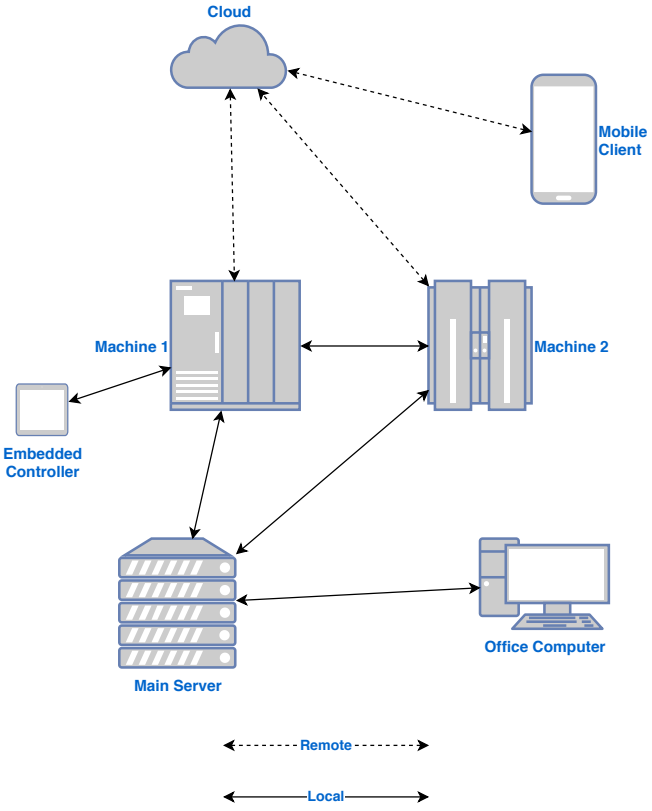


Fig. 1: OPC UA scenario

OPC UA is a client-server architecture [17a] and has been extended with the publish-subscribe pattern [18a] in 2018. However, most of the open-source implementations only support the client-server model, which relies on SecureChannels and Sessions [18b]. First, a client opens a SecureChannel which ensures integrity, confidentiality and application authenticity. Then a Session is created between the client and the server on top of a SecureChannel for user authorization and authentication. The client can access data or call methods on the server through Services [17c] such as the Browse or Read Services. In OPC UA, data and methods are implemented as nodes that are interconnected by references. Nodes and References form the hierarchical OPC UA address space [17b].

There are many open-source OPC UA implementations. For our consideration in this paper, an implementation must be

1. actively maintained,
2. full-stack server implementation and
3. completely open-source.

Among them, we selected the most popular ones based on the star counters in their GitHub pages and verified the ranking using trends in web search engines: open62541 [PP], node-opcua [Roa], UA-.NETStandard [OPd] and python-opcua [Frb]. In this group, open62541 has already the certificate and thus can be taken as reference in the interoperability tests. Furthermore, UA-.NETStandard is also certified and the reference implementation of the OPC foundation.

Since in 2015, the reference implementations of the OPC Foundation were not available for free and other open-source projects were not ready to be used in industrial automation, the authors of [Pa15] implemented open62541. The authors give a short feature comparison of open-source implementations back in 2015, including node-opcua and open62541. In [Ha17], both open-source and commercial OPC UA implementations are compared according to their features. However, some of the implementations have been discontinued, and others have been extended with new features. Also, the specification is extended with new services and security policies. The authors of [Ce19] test the open-source OPC UA implementations open62541, freeOpcUa C++ and python-opcua on embedded devices. This includes cross-wise RTT and service time measurements. A comparison in a broader sense is given in [Pr19], where the IoT protocols OPC UA, ROS, MQTT and DDS are investigated. Apart from a feature comparison of the protocols, the authors also analyse the protocol overhead and perform RTT measurements. These include tests with open62541 in client-server and in pub-sub mode. In [Mu20], the authors analysed the same open-source implementations covered in this paper from security and scalability aspects. The results show considerable differences in the scalabilities of the implementations depending on their languages. Moreover, the implementations have similar security models.

Our contributions in this paper are as follows. We compare the available features and functionalities of commonly used OPC UA implementations in Sect. 2. We evaluate their interoperabilities with various cross-wise connection tests in Sect. 3.

2 Features

OPC UA offers a set of many different functionalities which are continuously extended. Therefore, most implementations only support a subset of them. In this section, we give an overview of the supported Services, that are offered by OPC UA servers and can be used by clients. For most of the services, one needs to set up a SecureChannel that is secured by one of the standardized Security Policies, which are investigated in this section. Additionally, the different transport protocols of OPC UA are given. Finally, we provide an overview of implementations, offering clients a graphical user interface (GUI). To get this information, we inspect the source code and the readme-files in the repositories of the implementations.

2.1 Supported Services

The Specification part 4 [17c] defines the Services and aggregates them into Service Sets. Tab. 1 lists all currently standardized Services and whether they are implemented or not.

The Discovery Service Set is used by clients to find Servers and vice versa. Servers can publish their endpoints, i.e. the URIs including transport protocol and security settings. After the discovery, a client can connect to a Server by first establishing a Secure Channel and then a session using the respective services. Then, multiple services are available to the client: The NodeManagement Service Set can be used to add or delete nodes of the server's address space. For getting an overview of the address space, the Browse Service Set offers methods to show the sub-nodes of a node in the server. To view or alter the value of a variable, the View Service Set is used. When a server offers Methods, a client can call them via the Method Service Set. For clients, it is often of interest to get notified about data changes. This is achieved by creating a subscription using the Subscription Service Set, creating and binding Monitored Items to this Subscription and finally sending Publish requests to the server.

While UA-.NETStandard implements all Services, the other open-source projects miss some functionality. However, the major and commonly used services are available in all implementations. The Query Service Set is only supported by UA-.NETStandard. node-opcua does not support the management of nodes, i. e. adding and deleting nodes and references are not possible.

Tab. 1: Supported Services

Service	open62541	node-opcua	python-opcua	UA-.NETStandard
Discovery Service Set				
FindServers	✓	✓	✓	✓
FindServersOnNetwork	✓	✓		✓
GetEndpoints	✓	✓	✓	✓
RegisterServer	✓	✓	✓	✓
RegisterServer2	✓	✓	✓	✓
SecureChannel Service Set				
OpenSecureChannel	✓	✓	✓	✓
CloseSecureChannel	✓	✓	✓	✓
Session Service Set				
CreateSession	✓	✓	✓	✓
ActivateSession	✓	✓	✓	✓
CloseSession	✓	✓	✓	✓
Cancel		✓		✓
NodeManagement Service Set				
AddNodes	✓		✓	✓
AddReferences	✓		✓	✓
DeleteNodes	✓		✓	✓
DeleteReferences	✓		✓	✓
View Service Set				
Browse	✓	✓	✓	✓
BrowseNext	✓	✓		✓
TranslateBrowsePathsT.	✓	✓	✓	✓
RegisterNodes	✓	✓	✓	✓
UnregisterNodes	✓	✓	✓	✓
Query Service Set				
QueryFirst				✓
QueryNext				✓
Attribute Service Set				
Read	✓	✓	✓	✓
HistoryRead	✓		✓	✓
Write	✓	✓	✓	✓
HistoryUpdate	✓			✓
Method Service Set				
Call	✓	✓	✓	✓
MonitoredItem Service Set				
CreateMonitoredItems	✓	✓	✓	✓
ModifyMonitoredItems	✓	✓	✓	✓
SetMonitoringMode	✓	✓		✓
SetTriggering				✓
DeleteMonitoredItems	✓	✓	✓	✓
Subscription Service Set				
CreateSubscription	✓	✓	✓	✓
ModifySubscription	✓	✓	✓	✓
SetPublishingMode	✓	✓		✓
Publish	✓	✓	✓	✓
Republish	✓	✓	✓	✓
TransferSubscriptions				✓
DeleteSubscriptions	✓	✓	✓	✓

2.2 Security Policies

The security of the exchanged OPC UA messages heavily relies on the used Security Policy. They are defined in art 7 of the specification [17d] together with their security level as follows.

1. None (insecure)
2. Basic128Rsa15 (deprecated)
3. Basic256 (deprecated)
4. Basic256Sha256 (secure-high)
5. Aes128_Sha256_RsaOaep (secure-average)
6. Aes256_Sha256_RsaPss (secure-high)

None uses no cryptography at all, and thus it is insecure. *Basic128Rsa15* and *Basic256* rely on the broken SHA1 algorithm, and therefore they are deprecated. The remaining security policies are either targeting an average (128 bit) or high (256 bit) security level. The supported Policies are shown in Tab. 2.

Tab. 2: Supported Security Policies

Security Policy	open62541	node-opcua	python-opcua	UA-.NETStandard
None	✓	✓	✓	✓
Basic128Rsa15	✓	✓	✓	✓
Basic256	✓	✓	✓	✓
Basic256Sha256	✓	✓	✓	✓
Aes128_Sha256_RsaOaep	✓			✓
Aes256_Sha256_RsaPss				✓

None and the two deprecated Policies are provided by all implementations. The only secure security policy supported by all projects is *Basic256Sha256*. The support for *Aes128_Sha256_RsaOaep* and *Aes256_Sha256_RsaPss* is rather low. The open-source projects are continuously implementing new features, which can be seen in *open62541*, which added *Aes128_Sha256_RsaOaep* during our research.

2.3 Transport Protocols

OPC UA allows for different transport protocols. For client-server communication, HTTPS, WebSockets and TCP are available [17a]. For pub-sub communication, one can use UADP, a binary protocol on top of UDP, AMQP or MQTT [18a]. Tab. 3 shows the support for the specified protocols.

Tab. 3: Supported Transport Protocols. Some are experimental[e].

Transport	open62541	node-opcua	python-opcua	UA-.NETStandard
UA TCP	✓	✓	✓	✓
HTTPS				✓
Websockets	✓[e]			✓
UADP	✓[e]			
MQTT	✓[e]			
AMQP				

All implementations support the binary protocol over TCP, which is also the most efficient. UA-.NETStandard offers additional support for HTTPS and WebSockets. Open62541 has experimental support for Websockets and pub-sub with either OPC UA over UDP (UADP) or MQTT.

2.4 Client GUIs

For most of the users, the implementation of a custom client or the usage of a command-line interface (CLI) is inconvenient, and thus a GUI is favourable. Among the four investigated open-source implementations, three provide a GUI. UA-.NETStandard provides several example applications featuring a GUI in their repository [OPd]. However, these applications rely on the .Net Framework and thus only run on Microsoft Windows operating system. Additionally, a mobile client targeting Android, iOS and Windows UWP is available. The GUI of python-opcua is hosted in a separate repository [Fra] and is platform-independent. However, this project is not active since December 2019. An ncurses-based GUI, i. e. a text-based user interface (TUI), is offered by node-opcua. Again, this project is hosted separately from node-opcua and in [Rob]. Finally, open62541 does not provide any GUI.

3 Interoperability

Heterogenous OPC-UA-enabled machines are often interconnected, and users expect them to work together. While open62541 and UA-.NETStandard are officially certified by the OPC Foundation [OPd] [OPa], node-opcua and python-opcua do not have certificates. A certificate indicates that the product obeys the specification and can work with other certified implementations. In this section, we show that the open-source OPC UA solutions interoperate despite their missing certification. To do this, we first perform a simple test only consisting of a connection and retrieval of a standard value, namely the server time. In a second test, we include the usage of cryptography and further Services.

3.1 Cross-wise connection test

Our first test is performed by connecting clients to servers and requesting the server time using the Read Service with SecurityPolicy *None*. To connect to a server, a client sends a *HEL message* and further requests the following services: *OpenSecureChannel*, *OpenSession* and *ActivateSession*. To tear down the connection, the *CloseSession* and *CloseSecureChannel* services are called. All implementations are working with each other, which can be seen in Tab. 4. The results show that the connection mechanism and the package structure are compatible among the implementations.

Tab. 4: Connect and retrieve server time

server\client	**open62541**	**node-opcua**	**python-opcua**	**UA-.NETStandard**
open62541	✓	✓	✓	✓
node-opcua	✓	✓	✓	✓
python-opcua	✓	✓	✓	✓
UA-.NETStandard	✓	✓	✓	✓

3.2 Cross-wise functionality test

We perform pairwise connection tests further call various services. To limit the number of tests, we consider the following selection of services: *Browse*, *Read*, *CreateSubscription*, *CreateMonitoredItems*, *Publish* and *AddNodes*. These are some of the most useful services. Futhermore, we can get good coverage of the available Service Sets using these services. In the Services part of the specification [17c], the *Browse* service call is described as a method to retrieve the references of a node. One can start browsing the root node and recursively retrieve the whole tree of nodes. With the *Read* service, one can retrieve attributes of a node, for example, the server time in our basic connection test. To monitor changes of a value in the server, a client first creates a *Subscription*, then adds one or multiple *MonitoredItems* to the *Subscription* and finally sends *Publish* request. This service is used when supervising a production process, where a physical value such as temperature or pressure, should be monitored. For adding a node to the server, the *AddNodes* request is used. That is important when one wants to have a dynamic address space, like when a client desires to store additional information on the server. Since *Basic256Sha256* is the only secure and widely adopted Security Policy, we included it together with the security policy *None* in the test.

Since UA-.NETStandard is the reference implementation, we assume, that it obeys the standard and thus do not include it in this test. Further, open62541 can be seen as a benchmark in this test, as it is already certified.

As can be seen in Tab. 5, all implementations are working with each other. However, as node-opcua does not implement the NodeManagement Service Set, adding nodes does not

work whenever a node-opcua client or server is involved. All other implementations support all tested Services.

Tab. 5: Cross-wise functionality test. node-opcua does not support *Nodemanagement*^{n}.

server\client	open62541	node-opcua	python-opcua
open62541	✓	✓^{n}	✓
node-opcua	✓^{n}	✓^{n}	✓^{n}
python-opcua	✓	✓^{n}	✓

For the secure connection with Policy *Basic256Sha256*, the node-opcua server features the strictest certificate checking. It checks whether the client URI matches one of the certificates or not and whether the *KeyUsage* in the certificate is set correctly or not. Python-opcua and open62541 do not care about the *KeyUsage*. All three implementations come with scripts and small documentation on how to create the correct certificates.

4 Conclusion

In this paper, we investigated the four most popular open-source OPC UA implementations from the aspects of available features and interoperability. All implementations have essential functionalities and features. However, none of them implements the complete specification. The two implementations with most coverage are UA-.NETStandard and open62541. While the former implements all Service Sets of the client-server model, it does not support pub-sub communication. Whereas the latter features pub-sub and most but not all Service Sets. The newer Security Policies *Aes128_Sha256_RsaOaep* and *Aes256_Sha256_RsaPss* are only supported by UA-.NETStandard and partially by open62541. Except for open62541, all projects feature a GUI or TUI. Furthermore, we showed that the implementations work together in simple test scenarios. Our interoperability tests indicate that they obey the specification at least to a decent level. Overall, we conclude that open-source implementations can be utilized instead of or together with commercial solutions. Their features and programming languages make them suitable for different requirements. We can conclude that the implementations conform to the base standards since they support the basic features of the specification and they can interconnect.

References

[17a] OPC UA Specification Part 1: Overview and Concepts, 1.04, OPC Foundation, 2017.

[17b] OPC UA Specification Part 3: Address Space Model, 1.04, OPC Foundation, 2017.

[17c] OPC UA Specification Part 4: Services, 1.04, OPC Foundation, 2017.

[17d] OPC UA Specification Part 7: Profiles, 1.04, OPC Foundation, 2017.

[18a] OPC UA Specification Part 14: PubSub, 1.04, OPC Foundation, 2018.

[18b] OPC UA Specification Part 2: Security Model, 1.04, OPC Foundation, 2018.

[Ce19] Cenedese, A.; Frodella, M.; Tramarin, F.; Vitturi, S.: Comparative assessment of different OPC UA open–source stacks for embedded systems. In: 2019 24th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA). IEEE, pp. 1127–1134, 2019.

[Fra] Free OPC-UA Library: opcua-client-gui, https://github.com/FreeOpcUa/opcua-client-gui Accessed 28 June. 2020.

[Frb] Free OPC-UA Library: python-opcua, https://github.com/FreeOpcUa/python-opcua Accessed 08 May. 2020.

[Ha17] Haskamp, H.; Meyer, M.; Mollmann, R.; Orth, F.; Colombo, A. W.: Benchmarking of existing OPC UA implementations for Industrie 4.0-compliant digitalization solutions. In: 2017 IEEE 15th International Conference on Industrial Informatics (INDIN). Pp. 589–594, 2017.

[Mu20] Muehlbauer, N.; Kirdan, E.; Pahl, M.-O.; Carle, G.: Open-Source OPC UA Security and Scalability. In: 2020 25th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA). 2020.

[OPa] OPC Foundation: Certification of open62541 Server SDK, https://opcfoundation.org/products/view/open62541-server-sdk Accessed 28 June. 2020.

[OPb] OPC Foundation: OPC UA Case Studies, https://opcfoundation.org/resources/case-studies/ Accessed 28 June. 2020.

[OPc] OPC Foundation: OPC Unified Architecture, https://opcfoundation.org/about/opc-technologies/opc-ua/ Accessed 28 June. 2020.

[OPd] OPC Foundation: UA-.NETStandard, https://github.com/OPCFoundation/UA-.NETStandard Accessed 08 May. 2020.

[Pa15] Palm, F.; Grüner, S.; Pfrommer, J.; Graube, M.; Urbas, L.: Open source as enabler for OPC UA in industrial automation. In: 2015 IEEE 20th Conference on Emerging Technologies Factory Automation (ETFA). Pp. 1–6, 2015.

[PP] Pfrommer, J.; Profanter, S.: open62541, https://github.com/open62541/open62541 Accessed 08 May. 2020.

[Pr19] Profanter, S.; Tekat, A.; Dorofeev, K.; Rickert, M.; Knoll, A.: OPC UA versus ROS, DDS, and MQTT: performance evaluation of industry 4.0 protocols. In: Proceedings of the IEEE International Conference on Industrial Technology (ICIT). 2019.

[RC18] Resnick, C.; Clayton, D. A.: OPC Technology Well-positioned for Further Growth in Tomorrow ' s Connected World. In. 2018.

[Roa] Rossignon, E.: node-opcua, https://github.com/node-opcua/node-opcua Accessed 08 May. 2020.

[Rob] Rossignon, E.: opcua-commander, https://github.com/node-opcua/opcua-commander Accessed 29 June. 2020.

R. Reussner, A. Koziolek, R. Heinrich (Hrsg.): INFORMATIK 2020,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2021 379

Simulation Model for Threat and Impact Analysis on Modern Electrical Power Systems

Deeksha Gupta[1] Yongjian Ding[2] Dharini Govindaraj[3] Mathias Lange[2] Martin Szemkus[2] Karl Waedt[4]

Abstract: The increase in interconnected devices in electrical power systems raises the attack surface of a network and therefore the system connected within the network. Cyber-attacks can lead to power cut of an individual system or multiple systems required in the power generation stage and critical in normal operation of the plant. With the purpose to monitor and understand the impacts of cyber-attacks at the component level, system level and plant level, a testbed simulation environment for the Electrical Power Systems in a virtual power plant was modelled. This paper provides comprehensive information about the developed Simulink model for the electrical power system. The Simulink model was created to leverage freedom of customizing and testing of diverse cyber threat scenarios. The simulation model was set to communicate with physical controller to analyse the system level and plant level impacts of cyber-attacks on the physical devices.

Keywords: Cybersecurity; Matlab Simulink Model; Electrical Power System

1 Introduction

Cybersecurity issues in power systems have long been discussed. The simulated Aurora attack on an electrical power generator confirmed that vulnerabilities in protection systems could be exploited in order to cause severe damage to power system components [Ze11]. In order to keep the security of the Industrial Control Systems (ICS) and Electrical Power System (EPS) intact, it requires training of the individuals working on these systems. However, it is dangerous to conduct research and training directly on an operating commercial power plant, as minor disturbances can significantly lead to a negative impact on environment and economy. Therefore, a simulation model is necessary to leverage freedom of customizing and testing of diverse cyber threat scenarios [Gu20a].

The three stages of electric power supply are generation, transmission and distribution. After electrical power is generated at a power plant, it is transmitted over distances

[1] Technical University Dresden, Faculty of Electrical and Computer Engineering, 01069 Dresden, deekshagupta27@gmail.com
[2] Magdeburg-Stendal University of Applied Sciences, Institute of Electrical Engineering, 39114 Magdeburg, yongjian.ding@h2.de, mathias.lange@h2.de, mszemkus@icss.de
[3] Hochschule Darmstadt, Department of Electrical Engineering and Information Technology, 64295 Darmstadt, dharini.govindaraj@stud.h-da.de
[4] Framatome GmbH, 91052 Erlangen, Karl.Waedt@framatome.com

doi:10.18420/inf2020_35

using transmission line and the distribution system connects the transmission system to the consumers. Cybersecurity issues for electrical power systems in transmission or in distribution stages have long been discussed. However, there is very limited information available directing specifically cyber-attacks on EPS in generation stage. Therefore, the main focus of the simulation model was kept on electrical power systems in generation stage and located inside a nuclear power plant (NPP) [Gu20a].

A simulation environment was modelled to understand the physical process of EPS in a virtual nuclear power plant and also to analyse the impacts of cyber-attacks on EPS. Beyond the integration of the interface of the EPS model with the real digital devices, a key benefit for the plant personnel is the exercising and training of "what if" scenarios. These scenarios were simulated in the model and the model computed and showed the impacts at the component level, system level and plant level [Gu20b]. The analysed threat scenarios using simulation model include –data manipulation, availability attack, false data injection, Falsie trip command, etc. Additionally, the simulation model was set to communicate with physical devices (e.g. PLC) to analyse the effects of cyber-attacks on the devices. Open Platform Communications Unified Architecture (OPC UA) communication protocol [IE10] was used to setup the communication between the Simulink model and physical devices.

2 Simulation Model Design

Matlab Simulink tool was utilized to design the EPS simulation platform. The basic design of developed EPS Simulink model of the Electrical Power System was divided into the following 3 parts [Go19]: Power Generation, Grid Feed, and House Load (e.g. cooling system).

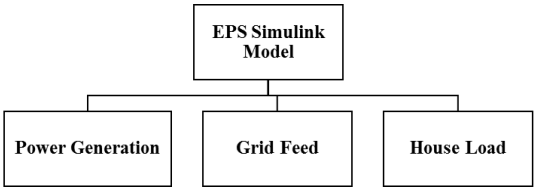


Fig. 1: Proposed EPS Model

3 Block Diagram of Simulink Model

The block diagram of the designed EPS Simulink model is represented in Fig. 1. The model focused mainly on start-up mode and normal operation mode, where the impact of the cyber-intrusion will be highest. Normal operation mode is the stage when the cooling loops require working at the demanding efficiency. Execution of attack in this state would lead the malfunctioning of the cooling system [Go19].

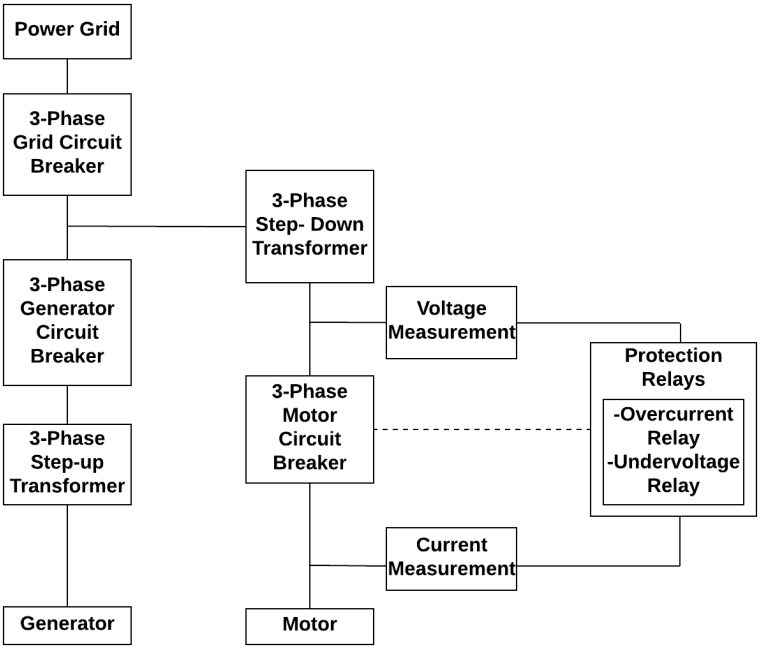


Fig. 2: Block diagram of EPS Simulink Model [Go19]

4 Implementation of Simulink Blocks

The picture of the Matlab Simulink Model is shown in Fig. 3. The Model is divided into 3 parts—first part includes Plant operations with specific attention on EPS, the second part focuses on control operation, finally, OPC communication module is the third part responsible to set up OPC communication between MATLAB Simulink mode and PLC [Go19]. This section provides the overview of the components in each part.

4.1 Power Generation

Power Generation module comprises components that are directly related to the power generation systems in an NPP. Fig. 4 demonstrates the Power Generation module that includes the following components: (1) Synchronous Generator; (2) Step-up Transformer; and (3) Circuit Breaker.

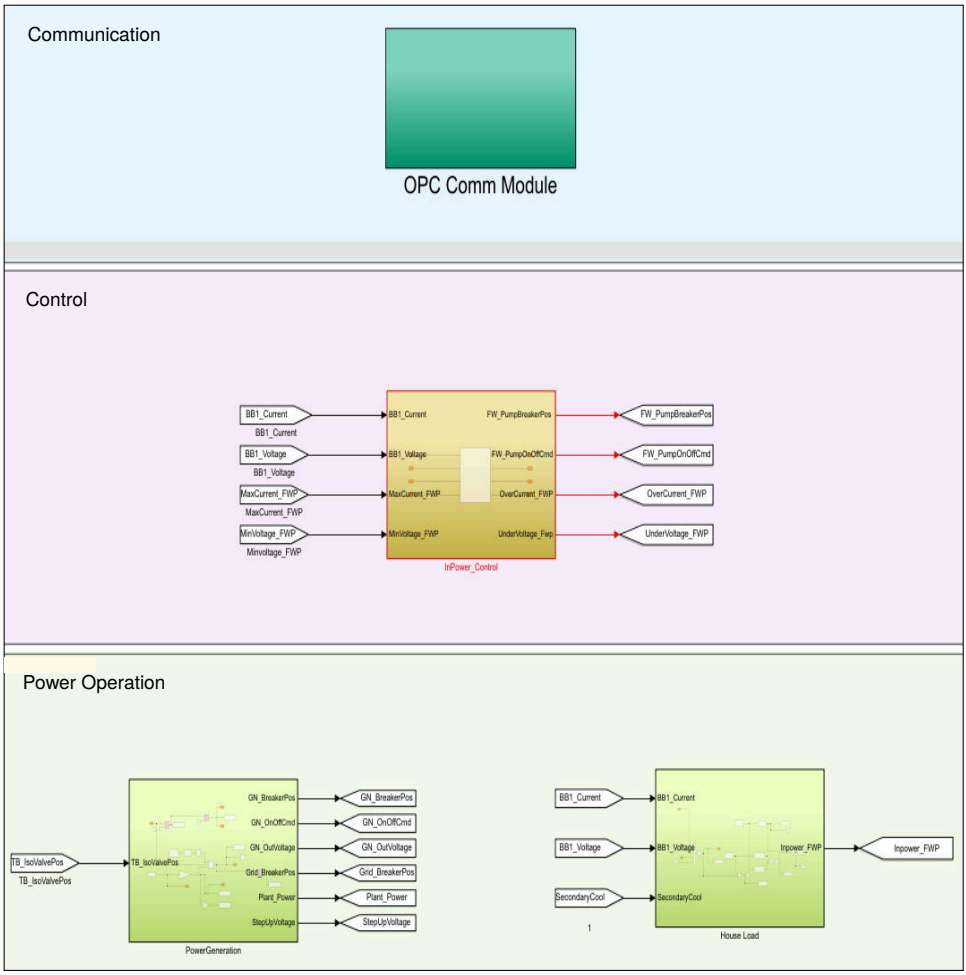


Fig. 3: Overview of MATLAB Simulink Model

4.2 Grid Feed

Fig. 5 shows the Grid Feed module that comprises the components of a simple power grid. This module encloses the following components: (1) Three-phase Source; (2) Transformer; (3) Three Phase; and (4) PI Section Line and Series RLC load.

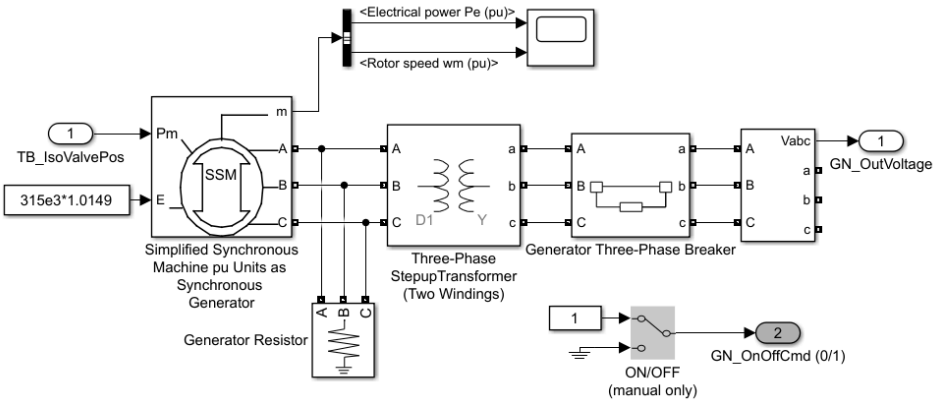


Fig. 4: Power Generation

4.3 House Load

Fig. 6 shows the modelled house load module in Simulink model. For the simplicity of the model a Feed Water Pump (FWP) is considered as house load. In a real power plant, house load encompasses multiple electro-mechanical machineries. House load module has the following electrical components: (1) Step-Down Transformer; (2) Circuit Breaker; (3) Asynchronous Motor; (4) Measurement Devices (current measurement and voltage measurement devices); and (5) Protection Devices for undervoltage and overcurrent protections.

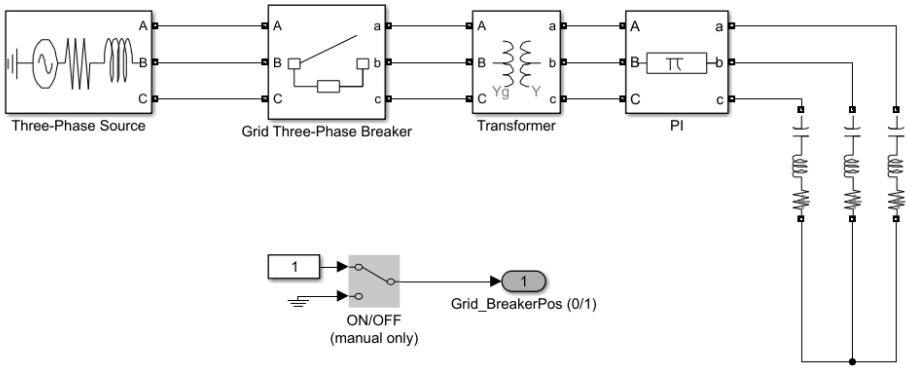


Fig. 5: Grid Feed Module

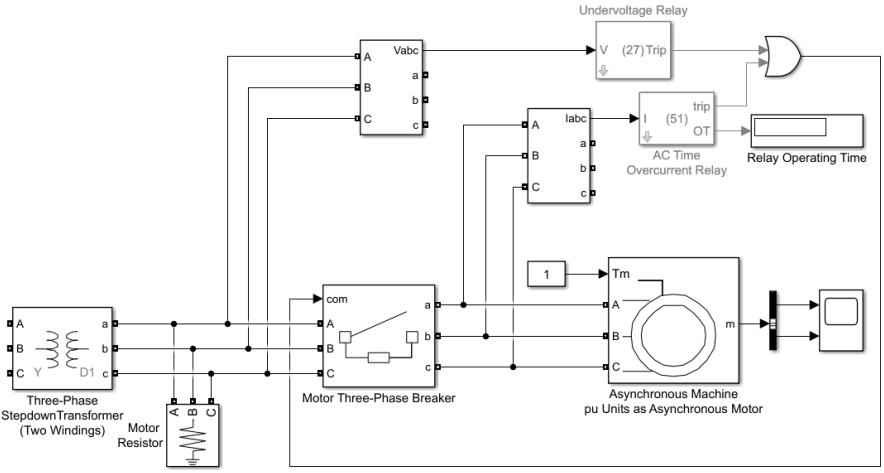


Fig. 6: House Load

5 OPC UA Communication between Matlab Model and PLC

An OPC UA communication protocol was established between the simulation model and the industrial controller by keeping in mind future scope implementation of communication between different hardware devices from different vendors. Fig. 7 shows details of the OPC Communication module. It can be observed from the figure that Simulink model includes four analog inputs and two digital inputs (total 6 inputs) and six analog outputs and five digital outputs (total 10 outputs). Tab. 1 elaborates the significance of each OPC tag that were used to transfer different parameter values from the Simulink model to the PLC and vice-versa.

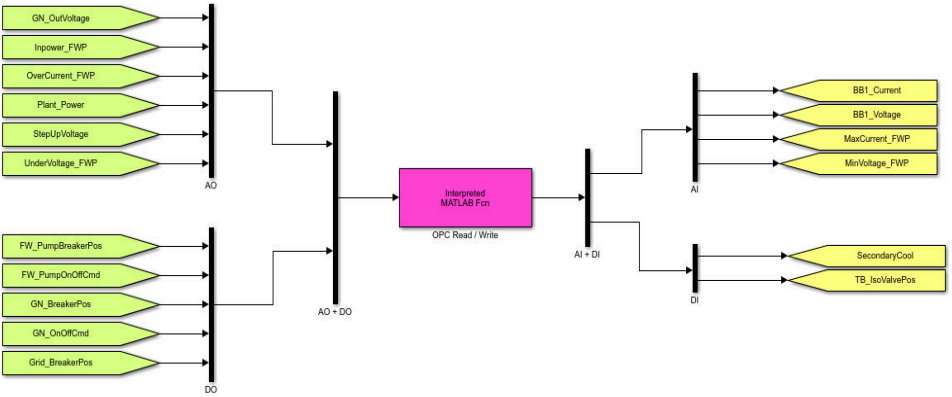


Fig. 7: OPC Tags for Communication

Tab. 1: OPC UA Tags Utilized in Simulink Model

Tag Type	*Tag Name*	*Significance*
Analog Input	BB1_Current	Value of current in Bus Bar 1
	BB1_Voltage	Value of voltage in Bus Bar 1
	MaxCurrent_FWP	Maximum permissible current to FWP
	MinVoltage_FWP	Minimum allowed voltage to FWP
Digital Input	SecondaryCool	State [On/Off (1/0)] of secondary cooling
	TB_IsoValvePos	Position [Open/Close (0/1)] of turbine isolation valve
Analog Output	GN_OutVoltage	Value of generator Voltage
	Inpower_FWP	Value of consumed power by FWP
	OverCurrent_FWP	Overcurrent set-point of FWP
	Plant_Power	Calculated power of a virtual NPP
	StepUpVoltage	Value of voltage of step-up transformer
	UnderVoltage_FWP	Undervoltage set-point of FWP
Digital Output	FW_PumpBreakerPos	Breaker position [Open/Close (0/1)] of FWP
	FW_PumpOnOffCmd	Status [On/Off (1/0)] of FWP
	GN_BreakerPos	Breaker position [Open/Close (0/1)] of generator
	GN_OnOffCmd	Status [On/Off (1/0)] of generator
	Grid_BreakerPos	Breaker position [Open/Close (0/1)] of main grid

6 Attack Implementation and Network Monitoring

The Matlab/ Simulink model, working on a Windows OS computer, was connected with other physical devices for attack implementation as shown in Fig. 8. Various Commercialized devices were used in this hardware-in-the loop (HIL) setup to perform the potential advanced threat scenarios. For our research purposes, an advanced industrial controller was selected. A medium sized (12 inches) PC-based commercial HMI-system was chosen as an Operator Panel. Attacker computer was a Linux based OS, including multiple open source tools for network protocol analysis and execution of cyber-attacks. A windows OS based network monitoring computer was connected to the network switch via a LAN tap to monitor the network traffic. Furthermore, in order to enhance the impact analysis capabilities and set up data transfer via OPC UA communication protocol between Simulink model and the physical device, PLC was programed with control logics and diverse electrical protection functions

of a motor protection relay. The main protection functions, considered in this research work cover – thermal overload, overcurrent, undervoltage, and start time supervision protections [Gu20a].

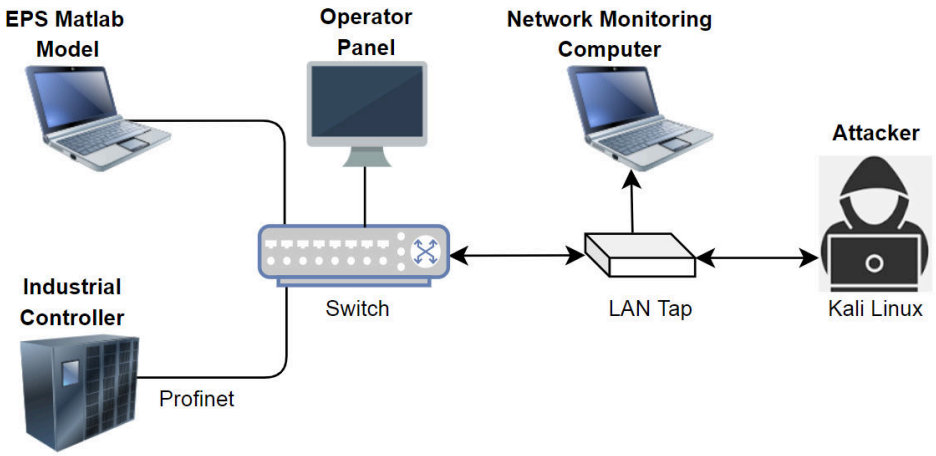


Fig. 8: Network Diagram for Attack Implementation and Network Monitoring

7 Experimental Analysis

An experimental result of an integrity attack using the hardware-in the loop setup is presented in this section. The target of this attack was the Circuit Breaker (CB) of a feed water pump, FWP1 of an NPP. An integrity attack is performed on the controller, controlling FWP1, by injecting false data configuration. The intention of this attack was to open the CB of FWP1 at an undesirable time. In Fig. 9, CB Status = 1 represents CB is closed and the FWP1 is running; CB Status = 0 represents CB is open and the FWP1 is disconnected from the power. Graph plotted in Fig. 9, is the representation of real time data received by Matlab/ Simulink model.

It can be noted from Fig. 9, at time t2 = 65 s, CB of FWP1 changed its status from 1 to 0. The alteration in the CB status was caused by the integrity attack that resulted in unintentional tripping of FWP1. The manipulated data configuration included overcurrent set-point parameter for the electric motor of a feed water pump in an NPP. This attack might also cause a healthy power line to become out-of-service even if there is no physical disturbance in power line.

Note: This threat scenarios was executed without considering the safety Instrumentation and Control (I&C) systems of an NPP. Therefore, only the operation I&C systems of the plant were targeted by leaving no impact on safety I&C systems.

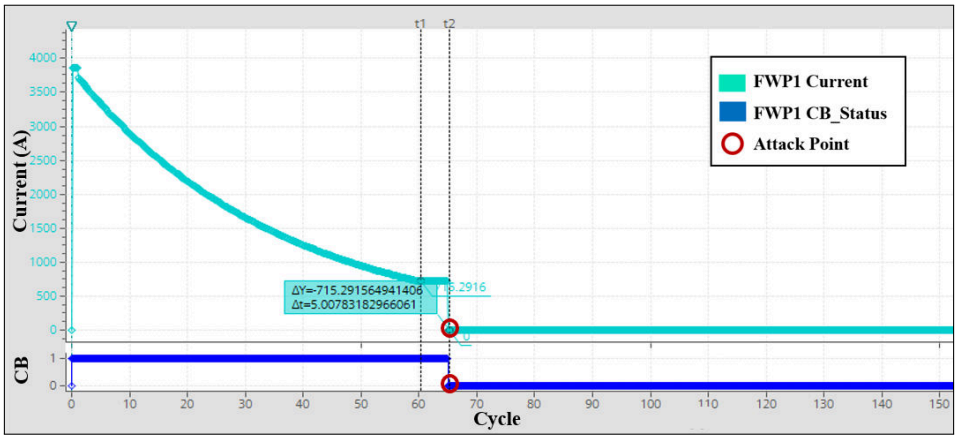


Fig. 9: Integrity Attack on FWP1 during Start-up Phase

8 Extension of the HIL Setup

The hardware in the loop setup presented in Section 6 can be further extended by an experimental model factory, operated by University of Applied Sciences Magdeburg Stendal [CB17] that has been modernized and expanded within the last years with regard to Industry 4.0. The model factory has been extended with common Siemens control (S7-1515F) and switches (Scalance SC615). Furthermore, a SIEM system was integrated. With this test setup, defined vulnerability tests are carried out and parameters are identified, which should enable an early detection of such attacks in the future.

With the structure described in Fig. 10 it is possible to model and experimentally evaluate attacks of different kinds [DI18]. The attacker system is a specially programmed framework or open source components. Thus, hidden channel attacks on different protocols or simple Denial of Service (DoS) attacks [SLD17] on different systems can be performed and analyzed.

As described in [SLD17] different attacks on PLCs could be evaluated and triggered. Thus disturbances, which are located on layer 2 and 3 of the OSI layer model, were tested by researchers. The disturbance caused by the test caused the blocking / isolation of the sensor signal, whereby the status change could no longer be transmitted to the control computer.

By integrating a manageable switch it is possible to implement virtual networks according to IEEE 802.1Q [IE14]. The segmentation resulting from the VLAN integration resulted in an increased protection against OSI Layer 2 and 3 attacks. However, this type of protection is only effective in conjunction with network segmentation.

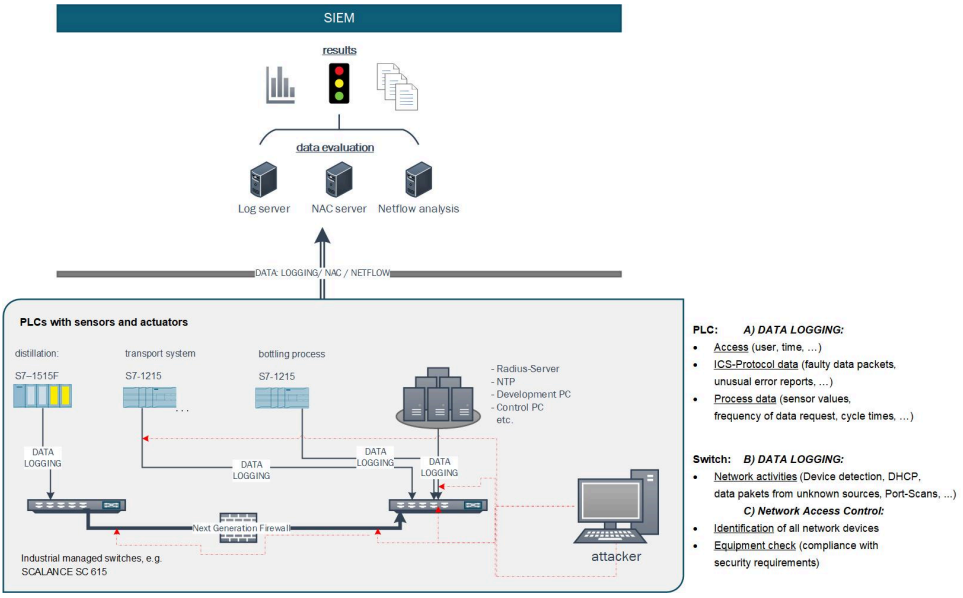


Fig. 10: Structure - Vulnerabilities Testing in an Experimental Environment

9 Summary and Outlook

The cyber-physical process model of targeted EPS inside a virtual NPP systems are described in this paper. The simulation model, divided in three parts (power generation, grid feed and house load) was designed and implemented using Matlab/ Simulink tool. The Simulink model was created to leverage freedom of customizing and testing of diverse threat scenarios. OPC UA communication protocol was used to setup the communication between Simulink model and physical devices. This simulation was used to analyze the impact of multiple cyber-threat scenarios by transferring the real time data from Simulink model to PLC and vice versa. A Network Diagram for cyber-attack implementation was also presented that was utilized to perform attack scenarios. An example integrity attack scenario was elaborated in this article. Furthermore, an extended version of the hardware in the loop setup – an experimental model factory was also presented in this paper that can be utilized for execution of more complex cyber-attacks and an early detection of such attacks in the future.

Acknowledgements

A part of our work in this domain is our contribution to IAEA project CRP J02008, entitled “Enhancing Computer Security Incident Response and Planning at Nuclear Facilities”. Some of the cybersecurity related topics are being elaborated as part of participation of Framatome

GmbH and University of Applied Sciences Magdeburg Stendal in the “SMARTEST” R&D (2015 - 2018) with German University partners, partially funded by German Ministry BMWi.

Bibliography

[CB17] Clausing, R,; Billowie, O.: Industrie 4.0 im Hochschul-Labor - Die Weiterentwicklung einer Modellfabrik. Tagungsband der Konferenz der Angewandten Automatisierungstechnik in Lehre und Entwicklung. Tagungsband - Angewandten Automatisierungstechnik in Lehre und Entwicklung (AALE), Wildau/Germany, 2017.

[DI20] DIN und DKE German Raodmap Industrie 4.0 Version 4, July 2020.

[DI18] Ding, Y.; Dittmann, J.; Szemkus, M; Lange, M.; Altschaffel, R.; Fischer, R.: Model-based vulnerability analysis of Complex infrastuctures, Berlin, 2018.

[IE14] IEEE Std 802.1, IEEE Standard for Local and Metropolitant Standard, Bridges and Bridged Network, Q. I. C. Society, 2014.

[IE10] IEC/TR 62541-2:2010 - OPC unified architecture - Part 2: Security model, 2010.

[Go19] Govindaraj, D.: Simulation of cybersecurity artefacts for Electrical Power Systems. Master Thesis, Darmstadt University of Applied Sciences, Germany, 2019.

[Gu20a] Gupta, D.: Nuclear Safety related Cybersecurity Impact Analysis and Security Posture Monitoring. PhD Thesis, Technical University Dresden, Germany, 2020.

[Gu20b] Gupta, D.; Govindaraj, D.; Altschaffel, R.; Waedt, K.: Blue Team Support for EPS Related Cybersecurity Readiness. In (ICONS 2020): IAEA International Conference on Nuclear Security, Vienna, 2020.

[Gu18] Gupta, D.; Bajramovic, E.; Parekh, M.; Waedt, K.: Threat Scenarios for Electrical Systems in Nuclear Power Plant. In (ICONE 2018): Proc. 26th International Conference on Nuclear Engineering, London, 2018.

[SLD17] Szemkus, M.; Lange, M.; Ding, Y.: IT-Security-Untersuchung an einer Modellfabrik unter Berücksichtigung der Industrie 4.0-Anforderungen. Tagungsband - Kommunikation in der Automation (KommA), Magdeburg, 2017.

[Ze11] Zeller, M.: Myth or reality does the aurora vulnerability pose a risk to my generator?, In (IEEE): Proc. 64th Annual Conference for Protective Relay Engineers, pp. 130–136, 2011.

Künstliche Intelligenz für kleine und mittlere Unternehmen

Künstliche Intelligenz für kleine und mittlere Unternehmen (KI-KMU 2020)

Natalja Kleiner[1] Alexander Dregger[2] Frauke Goll[3] York Sure-Vetter[4]

Abstract: Nach wie vor sind große Unternehmen Vorreiter der Digitalisierung. Doch der Mittelstand erschließt in immer stärkerem Maße seine digitalen Potentiale und holt auf. So planten im Jahr 2019 38,8% der in einer Studie der KfW[5] befragten kleinen Unternehmen mindestens zwei Digitalisierungsvorhaben in den kommenden zwei Jahren durchzuführen und weitere 28% dachten hierüber nach.

Vor diesem Hintergrund spielt auch Künstliche Intelligenz für Mittelständler eine immer wichtigere Rolle, denn laut einer Studie des Bundesministeriums für Wirtschaft (BMWi)[6] ist die Implementierung von KI im Mittelstand eine zwingende Voraussetzung, um auf dem internationalen Markt wettbewerbsfähig zu bleiben. Dabei liegen die Potentiale von KI im Mittelstand in sehr verschiedenen Bereichen wie z.B. Optimierung der Distribution und Logistik, gesteigerte Prozesseffizienz oder zielgenauere Werbung. Insbesondere Technologien wie intelligente Automatisierung, intelligente Sensorik oder intelligente Assistenzsysteme sind hierbei von großer Bedeutung.

Der Workshop „Künstliche Intelligenz für kleine und mittlere Unternehmen“ hat das Ziel, Forscher und Anwender von KI-Ansätzen zusammenzubringen, den Wissenstransfer zwischen den beiden Gruppen sowie den einzelnen Gruppen untereinander zu ermöglichen und KI in KMU greifbar zu machen. Dabei adressiert der Workshop zwei Schwerpunkte: Einerseits sollen Forschungsansätze präsentiert werden, die einen starken Bezug zu den Herausforderungen für kleine und mittlere Unternehmen im Bereich KI haben oder sich mit leichtgewichtigen KI-Anwendungen, die sich einfach und ohne großen Aufwand bei KMU integrieren lassen, beschäftigen. Andererseits können aber auch erfolgreich abgeschlossene oder laufende Pilotprojekte vorgestellt werden, die einen ersten Einblick in die Wirkung von KI in der Praxis geben und aufzeigen, welche Herausforderungen sich bei der Umsetzung von KI ergeben.

Die folgenden vier Beiträge des Workshops stellen aktuelle KI-Ansätze in unterschiedlichen Anwendungsfeldern und an unterschiedlichen Anwendungsbeispielen vor. Diese reichen vom Applikations- und Dosierprozess von Schmierfetten in der Produktion über die unternehmensübergreifende Nutzung von Predictive Manintenance Daten, die Duplikatenerkennung im CRM und CEM bis hin zu Empfehlungen von Radio Programmen. Dabei kommen verschiedene Algorithmen und Verfahren,

[1] FZI Forschungszentrum Informatik, Haid-und-Neu-Str. 10-14, 76131 Karlsruhe, natalja.kleiner@fzi.de
[2] FZI Forschungszentrum Informatik, Haid-und-Neu-Str. 10-14, 76131 Karlsruhe, dregger@fzi.de
[3] FZI Forschungszentrum Informatik, Haid-und-Neu-Str. 10-14, 76131 Karlsruhe, goll@fzi.de
[4] FZI Forschungszentrum Informatik, Haid-und-Neu-Str. 10-14, 76131 Karlsruhe, York.Sure-Vetter@fzi.de
[5] Zimmermann, V.: Unternehmensbefragung 2019, https://www.kfw.de/PDF/Download-Center/Konzernthemen/Research/PDF-Dokumente-Unternehmensbefragung/Unternehmensbefragung-2019-%E2%80%93-Digitalisierung.pdf, (2019)
[6] Lundborg, M., Märkel, C.: Künstliche Intelligenz im Mittelstand: Relevanz, Anwendungen, Transfer, https://www.mittelstand-digital.de/MD/Redaktion/DE/Publikationen/kuenstliche-intelligenz-im-mittelstand.html, (2019)

doi:10.18420/inf2020_36

wie z.B. K-Means Clustering, Federated Learning sowie Sequential Model Based Optimization bzw. Bayes'sche Optimierung zur Anwendung.

Programmkomitee: Dr. Andreas Abecker (Disy Informationssysteme GmbH, Karlsruhe), Dr. Dirk Achenbach (FZI Forschungszentrum Informatik, Karlsruhe), Jens Beyer (LAVRIO.solutions GmbH, Karlsruhe), Sergey Biniaminov (HS Analysis GmbH, Karlsruhe), Dr. Andreas Bildstein (Fraunhofer-Institut für Produktionstechnik und Automatisierung IPA, Stuttgart), Dr. Simone Braun (Uniserv GmbH, Pforzheim), Prof. Dr.-Ing. habil. Catherina Burghart (Hochschule Karlsruhe - Technik und Wirtschaft, Karlsruhe), Philipp Csernalabics (Neohelden GmbH, Karlsruhe), Darko Katic (ArtiMinds Robotics GmbH, Karlsruhe), Prof. Dr.-Ing. Jens Nimis (Hochschule Karlsruhe - Technik und Wirtschaft, Karlsruhe), Dr. Dominik Riemer (FZI Forschungszentrum Informatik, Karlsruhe), Hans-Peter Zorn (inovex GmbH, Karlsruhe), Thorsten Zylowski (CAS Software AG, Karlsruhe)

Keywords: Künstliche Intelligenz; KMU

KOBRA: Praxisfähige lernbasierte Verfahren zur automatischen Konfiguration von Business-Regeln in Duplikaterkennungssystemen

Simone Braun [1,2] Georges Alkhouri[3] Eric Peukert[4]

Abstract: Duplikaterkennung, -suche und -konsolidierung für Kunden- und Geschäftspartnerdaten, sog. „Identity Resolution", ist die Voraussetzung für erfolgreiches Customer Relationship Management und Customer Experience Management, aber auch für das Risikomanagement zur Minimierung von Betrugsrisiken und Einhaltung regulatorischer Vorschriften und viele weitere Anwendungsfälle. Diese Systeme sind jedoch hochkomplex und müssen individuell an die kundenspezifischen Anforderungen angepasst werden. Der Einsatz lernbasierter Verfahren bietet großes Potenzial zur automatisierten Anpassung. In diesem Beitrag präsentieren wir für ein KMU praxisfähige, lernbasierte Verfahren zur automatischen Konfiguration von Business-Regeln in Duplikaterkennungssystemen. Dabei wurden für Fachanwender Möglichkeiten entwickelt, um beispielgetrieben das Match-System an individuelle Business-Regeln (u.a. Umzugserkennung, Sperrlistenabgleich) anzupassen und zu konfigurieren. Die entwickelten Verfahren wurden evaluiert und in einer prototypischen Lösung integriert. Wir konnten zeigen, dass unser Machine-Learning-Verfahren, die von einem Domainexperten erstellten Business-Regeln für das Duplikaterkennungssystem „identity" verbessern konnte. Zudem konnte der hierzu erforderliche Zeitaufwand verkürzt werden.

Keywords: Sequential Model-Based Optimization; Nonlinear Regression; Reinforcement Learning; Entity Resolution; Identity Resolution

1 Einleitung

Ein wesentlicher Anteil des Aufwands in IT-Großprojekten fließt in die Migration von Daten von Alt- auf Neusysteme und die damit verbundene Integration verschiedener Datenquellen sowie deren Bereinigung. Fehler in der Migration sorgen für kritische Fehlfunktionen und Dateninkonsistenzen in neu entwickelten Systemen, die häufig erst zu einem späten Projektzeitpunkt entdeckt werden. Beispielsweise können falsche oder fehlende Kundendaten zu irrtümlich versendeter Werbung, falschen Rechnungen und Fehlbuchungen führen, die das Kundenverhältnis nachhaltig schädigen. Die Korrektur solcher Migrationsfehler sprengt meist den zeitlichen und finanziellen Rahmen des gesamten Großprojekts und führt

[1] UNISERV GmbH, Business Development, Rastatter Str. 13, Pforzheim, 75179, simone.braun@uniserv.com, https://orcid.org/0000-0002-4825-1648
[2] Hochschule Offenburg, Fakultät B+W, Klosterstraße 14, Gengenbach, 77723, simone.braun@hs-offenburg.de
[3] Institut für Angewandte Informatik e.V., Datenbanken, Goerdelerring 9, Leipzig, 04109, georges.alkhouri@gmail.com
[4] Institut für Angewandte Informatik e.V., Datenbanken, Goerdelerring 9, Leipzig, 04109, peukert@infai.org

doi:10.18420/inf2020_37

letztendlich zum Scheitern. Selbst wenn die Migration und das Gesamtprojekt erfolgreich waren, können auch in modernen IT-Systemen Fehl- oder Doppeleingaben oder veraltete Kundenstammdaten nicht ausgeschlossen werden.

Abhilfe schaffen hier hochleistungsfähige Datenmigrations- und Integrationswerkzeuge wie sie z.B. vom mittelständischen Unternehmen UNISERV zur Duplikaterkennung und -suche für Geschäftspartnerdaten (sog. „Identity Resolution“) angeboten werden. Identity Resolution ist die Voraussetzung für erfolgreiches Customer-Relationship-Management und Customer-Experience-Management, aber auch für das Risiko-management zur Minimierung von Betrugsrisiken und Einhaltung regulatorischer Vorschriften (z.B. Embargoverordnungen). Heute verstehen wir darunter nicht mehr allein Dublettenbereinigung im Rahmen von Datenqualitätsmanagement. Die Möglichkeit alle Informationen aus den unterschiedlichsten Kanälen, Systemen und Devices über Interessenten und Kunden zur 360°-Sicht in den Gesamtkontext zu stellen, bringt großen Mehrgewinn, z.B. besseren Einblick in Kundenbedürfnisse.

Der Bereich Duplikaterkennung ist an sich intensiv erforscht. Existierende Systeme sind in der Lage, Duplikate in Datenbeständen weitgehend automatisch und effizient zu identifizieren. Nur für kleine Mengen von Datensätzen sollen Nutzer noch einen manuellen Abgleich vornehmen müssen. Leider zeigt sich, dass Kunden sehr unterschiedliche und spezifische Anforderungen an die Duplikatidentifikation haben, so dass für jeden Anwendungsfall das Matching-System manuell neu konfiguriert werden muss. Typische Szenarien sind Haushaltsabgleiche, Umzugserkennung, Sperrlistenabgleiche etc., welche wiederum abhängig vom Geschäftsszenario in ihren Anforderungen variieren (z.B. Konsolidierung aller Finanz- und Versicherungsverträge in einem Haushalt, Vermeidung von Mehrfachzustellung einer Informationsbroschüre pro Haushalt oder Datenanreicherung und Personalisierung für Marketing-Kampagnen) und die jeweils ein unterschiedliches Verhalten der Duplikaterkennung erfordern. Bisher wird dies durch wenige Experten in einem zeitlich aufwendigen, iterativen Prozess gemeinsam mit den Fachanwendern beim Kunden umgesetzt (s.a. Kapitel 2). Dieser Anpassungsaufwand ist sehr kostenintensiv und benötigt lange und intensive Testphasen. In sehr großen Projekten mit komplexen Anwendungsfällen, wie bspw. im Finanz- und Versicherungsbereich, kann dieser Prozess gerne bis zu 50 Personentage in Anspruch nehmen; zeigt die Erfahrung von UNISERV. Diese Zeit steht jedoch oft nicht zur Verfügung, so dass neue Lösungen zur Automatisierung der manuellen Konfiguration von Duplikaterkennungssystemen gesucht sind, (a) unter Berücksichtigung kunden-individueller Anforderungen an die Duplikatidentifikation sowie (b) der praxisfähigen Anwendung und Einsatzes für ein mittelständisches Unternehmen wie UNISERV.

Herkömmliche Regel- und lernbasierte Duplikaterkennungen vergleichen sich mit einem Goldstandard von bekannten Ergebnissen um einen Qualitätswert zu erhalten. In der Arbeit von [KR08] wurde vor allem am Problem der Erstellung guter Trainingsdaten gearbeitet. Durch die gezielte Auswahl aussagekräftiger Paare konnte der Aufwand des manuellen Matchings für den Nutzer signifikant reduziert werden. Das Fever-System [KTR09, KR10] vergleicht verschiedene etablierte Lernverfahren im Objekt-Matching-Umfeld hinsicht-

lich des Parametrierungsaufwands, der erforderlichen Menge an Trainingsdaten und der erreichbaren Qualität anhand eines Webdaten-Matching-Problems. Hier konnte gezeigt werden, dass manche Lernverfahren bereits mit kleineren Mengen an Trainingsdaten gute Matching-Ergebnisse erzielen können. Zudem konnte für lernbasierte Ansätze insbesondere für schwierige Anwendungsfälle wie das Matching von Produktangeboten aus Web-Shops [Köp12] eine qualitativ bessere Dublettenerkennung als mit herkömmlichen manuell einzustellenden Ansätzen nachgewiesen werden. Jedoch ist auch hier die Menge an manuell zu labelnden Daten für den realen Einsatz mit ca. 500 Labeln sehr groß.

Großes Potenzial zur Adressierung des geschilderten Konfigurationsproblems und damit zur Reduktion des manuellen Labelaufwands versprechen lernbasierte Optimierungsansätze. Das algorithmische Framework der Sequential Model Based Optimization (SMBO), bietet einen globalen Optimierungsansatz für unterschiedlichste Lernmodelle, der sich für teure Blackbox-Funktionen als effektiv und dateneffizient erwiesen hat [Sha16]. Das Feld selbst erlebt eine Renaissance in der Anwendung zur automatischen Konfiguration von Algorithmen und zur Hyperparameteroptimierung im Bereich des maschinellen Lernens [Yao18, Sha16, HHL11, Feu15, Men16, ZL16]. Vermehrt Interesse an lernbasierten Optimierern gibt es auch im Bereich des Meta-Learning auf Basis von Recurrent Neuronal Networks (RNN). [TV20] entwickelten mit RNN-Opt einen Ansatz zur Blackbox Optimierung, wenn kein Gradient gebildet werden kann. Des Weiteren wurde Reinforcement Learning für die automatische Konfiguration virtueller Maschinen von [Rao09] genutzt. Bis zum jetzigen Zeitpunkt wurden nach unserem Wissensstand allerdings die Verfahren nicht in der realen Anwendung von KMU oder mit großen Datenmengen in der Domäne der Kunden- und Geschäftspartnerdaten evaluiert.

Im Folgenden gehen wir detailliert auf die Ausgangssituation und daraus abgeleitete Anforderungen und Zielstellung ein. In Kapitel 3 präsentieren wir die KOBRA-Lösung zur automatischen Konfiguration von Business-Regeln in Duplikaterkennungssystemen mit lernbasierter Unterstützung sowie in Kapitel 4 die Lernkomponente und implementierten Lern- und Optimierungsverfahren, bevor wir im Anschluss (Kapitel 5) auf Evaluation und Test der Verfahren und insbesondere die Praxisfähigkeit für ein KMU wie UNISERV eingehen. In Kapitel 6 diskutieren wir die Ergebnisse der Evaluation der Gesamtlösung am praktischen Anwendungsfall, bevor wir unseren Beitrag mit einer Zusammenfassung und einem Ausblick abschließen.

2 Ausgangssituation

Wann im Bereich der Kunden- und Geschäftspartnerdaten eine Dublette als eine Dublette erkannt werden soll, hängt stark vom kundenindividuellen Geschäftsszenario und ihren variierenden Anforderungen ab. Zum Beispiel kann es sich bei den Datensätzen „G. Mayer, Pforzheimer Str. 320, 70499 Stuttgart“ und „Gerhard Meier, Pforzheimer Str. 320, 70499 Stuttgart“ um dieselbe Person mit Tippfehler bei der Namenseingabe handeln oder um zwei getrennte Identitäten. Mag eine Zusammenführung der Datensätze bei

einer Datenanreicherung für eine Marketingkampagne weniger kritisch sein, so wäre eine Konsolidierung für die Zusammenführung von Bankkonten ggf. fatal.

Bei der Konfiguration der Business-Regeln (auch Matching-Regeln) in Duplikaterkennungssystemen wie dem UNISERV „identity“ ist nun kundenindividuell einzustellen, ob und an welcher Stelle ein strenger Abgleich, (zeichengenaue Übereinstimmung „Gerhard Mayer-Vorfelder“ | „Gerhard Mayer-Vorfelder“) oder toleranter Abgleich („Gerhard Mayer-Vorfelder“ | „Gerhart Meier-Forvelter“ | „Erhart Nayer-Vorfeider“ | „Gero MV“ | „Герхард Маыер-Ворфелдер“) oder auch Vergleiche auf Haushaltsebene (Erkennen der Zusammengehörigkeit von „Gerhard Mayer-Vorfelder“ und „Margit Mayer-Vorfelder“) u.v.m. erfolgen soll; wobei der Anteil an „false-positive“ und „false-negative“ Treffern gering zu halten ist um manuelle Nachbearbeitungen zu vermeiden.

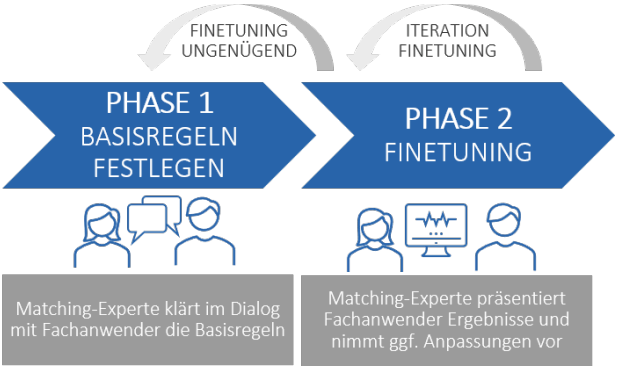


Abb. 1: Bisherige Vorgehensweise Erstellung der Matching-Regeln

Die aktuell übliche Vorgehensweise zur Regelkonfiguration im Matching-System „identity“ von UNISERV ist wie folgt (s. Abbildung 1): Der Kunde stellt einen repräsentativen Datenauszug bereit und erläutert UNISERV Consultants und Matching-Experten den Business Case, Ursprung und Inhalt der Daten (z.B. überwiegend natürliche oder juristische Personen) sowie weitere Kriterien (z.B. Existenz von Pflichteingabefeldern oder Toleranzen). Es erfolgt eine erste gemeinsame Analyse. Basierend auf Erkenntnissen aus dem Gespräch wählen die Matching-Experten aus einer Reihe von Standard-Konfigurationen Basisregeln für das Matching-System aus. Das Matching-System wird ausgeführt und das Ergebnis wird mit den Fachanwendern diskutiert und gemeinsam entschieden, ob das grundsätzliche Ziel erreicht wurde. Falls nicht, werden neue Basisregeln gewählt und der Vorgang wiederholt. Ansonsten erfolgt im nächsten Schritt die Feinjustierung anhand ausgewählter Beispiele (Markierung von false-positive und false-negative) bzw. neuer Beispiele seitens der Fachanwender. So können Fachanwender anhand ihres Anwendungsfalls z.B. folgende Situationen entscheiden: (1) Input: „G. Mayer, Pforzheimer Str. 320, 70499 Stuttgart“ und Possible Match: „Gerhard Meier, Pforzheimer Straße 320, 70499 Stuttgart“ → diese Art von Datensätze sollen als Dublette identifiziert werden, d.h. wären true-positive; (2) Input: „G. Mayer, Pforzheimer Str. 320, 70499 Stuttgart“ und Possible Match: „IT-Mayer GmbH, Rastatterstr. 320 70499 Stuttgart“ → diese Art von Datensätze sollen nicht als Dublette identifiziert werden, d.h.

wären false-positive. In diesem Prozess versuchen Consultants bzw. Matching-Experten insbesondere Widersprüche in den Entscheidungen der Fachanwender aufzulösen. Die Matching-Regeln werden auf Basis des Feedbacks der Fachanwender angepasst, das System ausgeführt und das Ergebnis erneut mit den Fachanwendern diskutiert. Dies wird so lange wiederholt bis ein zufriedenstellendes Ergebnis erzielt ist, was zeitlich sehr aufwendig sein kann.

Bisher ist die Erstellung der Matching-Regeln komplex und erfordert tiefes technisches Wissen und Erfahrung. Abbildung 7 zeigt einen Ausschnitt der Konfigurationsmatrix, die das Matching-Verhalten und die Bewertung sämtlicher Elemente steuert[5]. Über die Parameterwerte kann ausgedrückt werden, dass der Name tolerant sein kann, wenn die Adresse identisch ist und umgekehrt. Oder die Hausnummer darf verschieden sein, wenn Name und Adresse identisch sind. Aber auch wie eine phonetische Ähnlichkeit, zum selben Ort gehörige PLZ oder Hausnummernnachbarschaft zu bewerten sind sowie erforderliche Mindestübereinstimmung, Elementreihenfolge beim Abgleich, weitere Ähnlichkeitsgrade wie Größe von Wortabständen, Synonyme, Initialen oder Umgang mit leeren Feldern, Wortreihenfolge („Meyer-Vorfelder“ | „Vorfelder-Meyer“) oder Überhängen („Meyer“ | „Meyer-Vorfelder“), Berücksichtigung des Geschlechts beim Vornamen, Vertauschen von Tag und Monat im Geburtsdatum u.v.m.

So wird z.B. beim Abgleich der Adresse für jeden Adresselementtyp (wie PLZ, Ortsname, Straßenname, Hausnummer, Name, Vorname) und für jedes Vergleichsergebnis (identisch, phonetisch ähnlich, ähnlich, verschieden usw.) ein Parameterwert zwischen -100 und 100 festgelegt. Für jedes Wort der Eingabe wird der Grad der Übereinstimmung mit dem entsprechenden Begriff der Referenz ermittelt. Aus diesem Ergebnis ergibt sich im Zusammenhang mit dem Parameterwert der Wert, mit dem das Wort in den Gesamtwert eingeht. Die Gesamtbewertung der Adresse ist z.B. der Durchschnitt der für die Begriffe ermittelten Werte unter Einbeziehung der Gewichtung je Adresselementtyp. Nur Adressreferenzen, deren Gesamtbewertung über einem in der Abfrage angegeben Grenzwert liegen, werden als potenzielle Dubletten zurückgeliefert.

Die notwendige Genauigkeit einer Konfiguration ist hierbei immer vom konkreten Anwendungsfall abhängig. Werden bspw. bei Embargo-Prüfungen Datensätze (Personen) nicht erkannt, kann dies empfindliche Strafen zur Folge haben. Auf der anderen Seite verursacht jeder Verdachtsfall Kosten, da diese manuell überprüft werden müssen. Die Kosten nicht-erkannter Datensätze überwiegen hier jedoch, so dass false-negatives soweit wie möglich vermieden werden sollten, während die Kosten von false-positives weniger kritisch zu betrachten sind und im Zweifel lieber ein Treffer zu viel als zu wenig angezeigt werden sollte. Damit soll z.B. bei Prüfung gegen eine PEP-Liste („Politically Exposed Persons“) ein Treffer vorliegen, wenn eine der Personen im Namensfeld mit einem Eintrag in der PEP-Liste übereinstimmt, andere Attribute können dabei deutlich weniger Gewicht erhalten.

[5] Auf genaue Details der Matrix kann an dieser Stelle aus Gründen der KnowHow-Wahrung nicht weiter eingegangen werden.

Ein anderes Beispiel stellt die Pflege einer Adressdatenbank dar: Soll eine Datenbank von Dubletten bereinigt werden, um bspw. alle Verträge einer Kundin in einem Online-Portal zusammenführen zu können, wird man die gegenteilige Strategie wählen. Falls im Fehlerfall eine Endkundin die Daten einer dritten Person einsehen könnte, wäre der Schaden ungleich höher, als wenn diese fälschlicherweise zwei Logins bekäme. False-positives sind hier kritischer zu sehen und die Konfiguration sollte sich in Zweifelsfällen gegen eine automatische Dubletten-Zusammenführung entscheiden.

Bei der Erstellung der Business-Regeln für eine Matching-Verarbeitung geht es in einigen Fällen nicht nur um doppelte oder mehrfach vorhandene Stammdaten, sondern oft auch um zugehörige Transaktions- oder Stammdaten aus anderer Quellen, die über "weiche" Identifikationskriterien wie Namen + Adressen, Namen + Geburtsdatum, Telefon-Nr., E-Mailadresse oder Social Login zugeordnet werden müssen. Somit kann es sein, dass die Abweichungstoleranz für einige Attribute sehr klein sein muss und für andere wiederum größer sein darf. Das heißt, das System ist so zu parametrieren, dass Datensätze auch dann als Treffer ausgewiesen werden, wenn diese sich in bestimmten Feldern unterscheiden.

Zusammenfassend lassen sich folgende Kernanforderungen ableiten:

- Die Entwicklung optimaler Regeln soll beschleunigt werden.
- Fachanwender sollen den Konfigurationsprozess selbstständig und an ihren individuellen Business Case anpassbar durchführen können.
- Fachanwender sollen keine zusätzlichen Kompetenzen im Bereich Konfiguration von Matching-Regeln aufbauen müssen.
- Anhand von Beispielen sollen Fachanwender eine passende Konfiguration automatisch ableiten lassen können.

3 Lösung

Ziel ist es, die Entwicklung von optimalen Matching-Regeln gegenüber dem bisherigen Vorgehen deutlich zu beschleunigen und zu vereinfachen. Fachendanwender des UNISERV Matchingsystems „identity“ sollen in die Lage versetzt werden möglichst optimale Matching-Regeln ohne die Unterstützung von Machting-Experten mit möglichst geringem Zeitaufwand und Know-How-Aufbau zu erstellen (vgl. Abbildung 2).

Die Grundidee ist, dass Fachanwender das System für ihren individuellen Business Case trainieren, indem sie manuell gelabelte Trainingsdaten zur Anreicherung eines vortrainierten Modells zur Verfügung stellen. Ein dadurch verbessertes Vorhersagemodell tritt an die Stelle des Consultants bzw. Matching-Experten und optimiert mithilfe der gelabelten Trainingsdaten die Konfiguration des Matching-Systems.

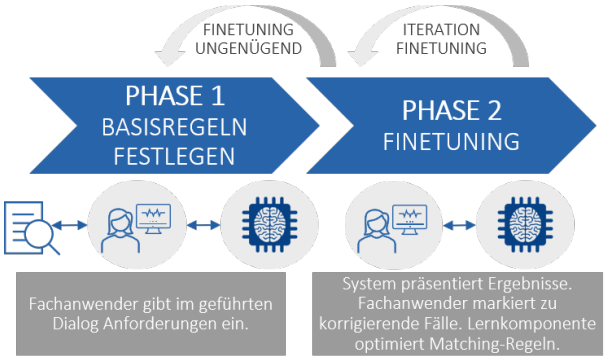


Abb. 2: Erstellung optimaler Matching-Regeln mit lernbasierter Unterstützung

Damit ein Projekt wirtschaftlich ist, muss die Zeit für das manuelle Labeling der Daten deutlich geringer sein als eine direkte Konfiguration zu erstellen. Hierzu ist die Extraktion einer repräsentativen Trainingsmenge notwendig. In der Regel unterscheidet sich eine gute Konfiguration von einer schlechten nicht im Bereich von sicheren Dubletten und auch nicht im Bereich von sicheren Nicht-Dubletten, sondern anhand der unsicheren Dubletten.

Um die Anforderungen und das gesteckte Ziel zu erfüllen, wurde die KOBRA-Lösung entwickelt. Sie besteht im Wesentlichen aus vier Komponenten: (1) Webapplikation für den Anwenderdialog, (2) Backend für die Datenverwaltung, (3) Matching-System auf Basis des UNISERV-Produkts „identity“ zur Durchführung der Duplikaterkennung und (4) Lernbasierte Komponente, die mittels sequentieller Optimierung und auf Grundlage der manuell gelabelten Daten ein Vorhersagemodell trainiert, um eine optimale Konfiguration der Matching-Regeln zu erstellen.

In einem geführten Prozess werden zunächst die Anforderungen ermittelt. Über die Webapplikation erstellen Fachanwender ein neues Projekt und beantworten im Dialog Fragen bzgl. Anwendungsfall oder spezieller Kriterien (z.B. phonetische Toleranz, Verwerfen eines Matches bei Differenz etc.). Parallel dazu erfolgt die Analyse der bereitgestellten Daten bzgl. Herkunftsland, Sprachraum, Befüllungsgrad, Qualität etc. Anhand der Eingaben und Datenanalyse werden die Basiskonfigurationen generiert.

Darauf erfolgt auf Grundlage der ermittelten Basiskonfigurationen die Bildung und Selektion von Dublettengruppen (es werden die Teilergebnisse selektiert, welche nicht identisch zu den Teilergebnissen anderer Konfigurationen sind). Die Gruppen werden den Fachanwendern zur Qualifizierung in (a) true-positive – gewünschter Treffer und (b) false-positive – unerwünschter Treffer angezeigt (s. Abbildung 3). Die Fachanwender bewerten die Dublettengruppen, indem sie eine Kopfdublette[6] auswählen und jeden weiteren Datensatz

[6]Bei einer Konsolidierung werden verschiedene Datensätze, die sich auf dieselbe Real-Welt-Entität beziehen, in einer Dublettengruppe mit einem Kopfdatensatz (der mit den meisten Informationen) zusammengefasst. Dieser kann mit weiteren Informationen aus den gefundenen Dubletten-Datensätzen ergänzt werden.

entsprechend qualifizieren, ob dieser richtig oder falsch als Mitglied dieser Dublettengruppe erkannt wurde. Das Ergebnis ist eine qualifiziert bewertete Trefferliste mit gewünschten und unerwünschten Dubletten.

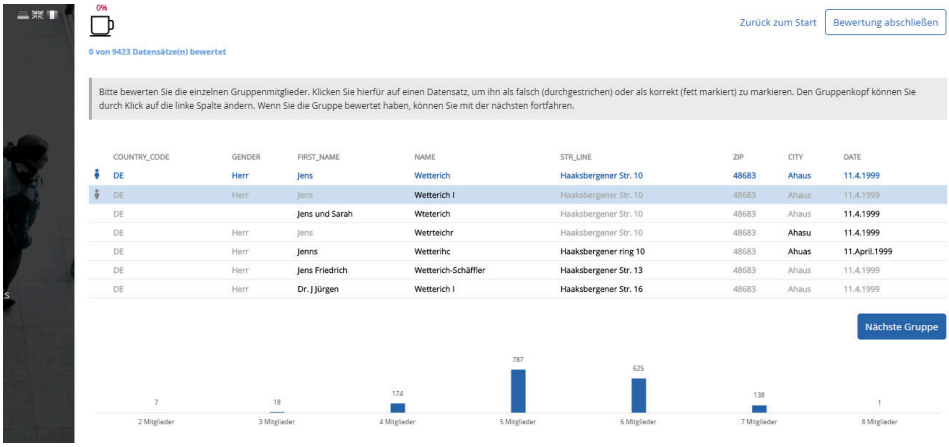


Abb. 3: Bewertung der Dublettengruppen

Mit dem entwickelten Goldstandard erfolgt zuletzt die Ermittlung einer optimalen Konfiguration mit Hilfe der Lernkomponente. Ziel ist es eine möglichst gute Konfiguration mit den vorhandenen Ressourcen zu ermitteln. Dabei werden wiederholt verschiedene Konfigurationen getestet und durch das Vorhersagemodell ausgewählt und mögliche Verbesserungen anhand des Goldstandards bewertet.

Im Allgemeinen können nicht alle Anforderungen des Goldstandards erfüllt werden und somit ist dieses Problem nicht fehlerfrei lösbar. Um eine Bewertung zu ermöglichen, wird ein Abstandsmaß zwischen dem Ergebnis der automatisch erstellten Konfiguration und dem Goldstandard des Fachanwenders berechnet. Die verwendete Metrik liefert verschiedene Kenngrößen zur Qualitätsbewertung sowie einen einzelnen Qualitäts-Score, welcher im Idealfall den Wert 100 oder den normierten Wert 1.0 annimmt.

4 Lernkomponente

Das Finden einer geeigneten Konfiguration ist gleich zu setzen mit einem hohen Qualitäts-Score, wie dem F-Score. Somit lässt sich ein sinnvolles Lernproblem aus der obigen Problemstellung als ein Regressionsproblem formulieren. Hierbei fungieren die Konfigurationsparameter als unabhängige Variablen mit Hilfe derer das Regressionsmodell den Qualitäts-Score ableitet. Intuitiv soll das Modell Wertebereiche im Konfigurationsraum finden, welche eine hohe Qualität besitzen und somit eine gute Duplikaterkennung ermöglichen sollen. Anschließend dienen die gefundenen Werte als Konfigurationsparameter für das Matching-System. Um erfolgreich ein Regressionsmodell zu trainieren, werden gelabelte

Daten in Form von Konfigurationen $\boldsymbol{x}$, mit resultierendem Qualitätswert $\boldsymbol{f(x)}$ benötigt, da es sich um einen Supervised-Lernalgorithmus handelt. Wie in Kapitel 1 erwähnt, sind solche Datenpaare in einem nicht ausreichenden Maße vorhanden. Wir lösen dieses Problem, indem wir Trainingspaare durch die Umgebung selbst generieren und als algorithmisches Framework die Sequential Model Based Optimization [HHL11] nutzen. Im Folgenden gehen wir näher auf das SMBO Interface ein.

Abbildung 4 zeigt vereinfacht die entwickelte Lernkomponente als SMBO Interface. Zu Beginn der Iteration wählt ein Modell $\boldsymbol{M}$ eine zu testende Konfiguration $\boldsymbol{x}$ aus (3). Dies geschieht mithilfe einer Acquisition-Funktion, welche bestimmt, wie der zugrunde liegende Konfigurationsraum exploriert werden soll und welche Parameter der Konfiguration geeignete Kandidaten sind. Anschließend evaluiert das Matching-System $\boldsymbol{f}$ die Konfiguration $\boldsymbol{x}$ (1). Dabei ist die Evaluation der Konfiguration ein kritischer Punkt, da diese die Laufzeit und somit auch die Performanz des Optimierungsprozesses beeinflusst. Daher sollten nur gut geeignete Konfigurationen für diesen Schritt ausgewählt werden. Im letzten Schritt der Iteration wird der Qualitäts-Score $\boldsymbol{f(x)}$ und die Konfiguration $\boldsymbol{x}$ genutzt, um das Regressionsmodell $\boldsymbol{M}$ neu zu trainieren (2). Die Hoffnung ist, dass somit wichtige Korrelationen von Konfigurations- und Qualitätswerten erlernt werden. Idealerweise erkennt das Modell die Eigenheiten der Konfigurationsparameter und kann bei neuen Optimierungsläufen schneller eine gute Konfiguration finden.

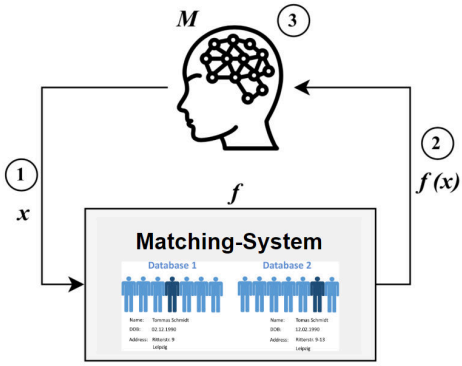


Abb. 4: SMBO Interface für die automatische Konfiguration eines Matching-Systems

Für die Vorhersage der Qualitätswerte haben wir zwei Regressionsmodelle ausgewählt, um diese im SMBO Interface zu nutzen: zum einen Extremely Randomized Trees und zum anderen Gaußsche Prozesse. Um zu zeigen, dass solch ein Konfigurationsproblem nicht trivial durch einen Direct Search Algorithmus gelöst werden kann, haben wir Random Sampling implementiert. Weiterhin haben wir auch einen, zum SMBO Interface q^1artverwandten Reinforcement Learning DQN-Ansatz [Osb16] für die Auswahl und Evaluierung von Konfigurationen verwendet.

5 Evaluation der Lern- und Optimierungsverfahren

Zu Evaluations- und Testzwecken wurden mittels eines Datengenerators synthetische Datensätze mit Namens- und Adressdaten erzeugt, die reale Datensätze mit einer Datenqualität repräsentieren, wie sie in Unternehmensanwendungen (z.B. CRM- oder ERP-Systeme)

vorzufinden sind. Der Rückgriff auf synthetische Daten erfolgt aus Rücksicht auf Datenschutzgesetze. Bei der Generierung werden gezielt Dubletten und False-Positives erzeugt. Als weiterer synthetischer Datensatz wurde ein realer Zensus-Datensatz der Mainzelliste[7] mit ≈50k Einträgen genutzt, der deutsche Adressdaten (Nachname, Vorname, Adresse, Telefon, PLZ, etc.) mit ca. 1200 bekannten Matches enthält.

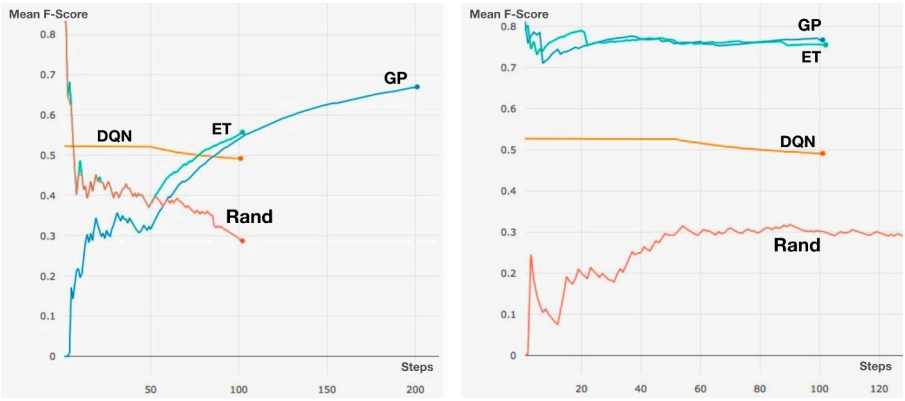


Abb. 5: Vergleich Regressions- und Optimierungsmodelle im Training- (li) und Test (re).

Auf Basis dieser erfolgte eine Einteilung in Trainings- und Evaluierungsdatensätze für einen ersten Vergleich von Random Sampling (Rand), Extremely Randomised Trees (ET), Gaussian Process (GP) und Deep Q-Learning (DQN) (s. Abbildung 5) zur Autokonfiguration des Matching-Systems von UNISERV. Die Modelle wurden mit den Trainingsdatensätzen vortrainiert und anschließend mit dem Evaluierungsdatensatz erneut trainiert (im 2. Training sollen Verbesserungen beobachtet werden). Idealerweise sollte der Agent mit zunehmender Anzahl von Versuchen lernen, das Matching-System optimal zu konfigurieren.

Im Laufe des Trainings nimmt der durchschnittliche F_1-Wert von ET und GP zu und wächst für GP logarithmisch und stabil. Rand und DQN zeigen bereits im Training kein gutes Verhalten. Die unterschiedlichen Startpunkte in der Qualität auf der Y-Achse sind zufällig gut gewählte Konfigurationen. Auch in der Evaluation zeigten Rand und DQN keinerlei Fähigkeit, eine Konfiguration zu verbessern. Dies ist für Random Sampling (Rand) nachvollziehbar, da hier nur zufällige Konfigurationswerte ausgewählt werden. DQN ist durch die Auswahl und Manipulation der Konfigurationswerte sehr begrenzt, da nur ein Wert pro Iteration verändert werden kann. Zusätzlich ist Q-Learning ein Off-Policy RL-Algorithmus, welcher erheblich mehr Beobachtungen benötigt, um sinnvolle Schlussfolgerungen für das jeweilige Optimierungsproblem zu ziehen. Somit kann DQN mit lediglich 50 bis 150 Iterationen bzw. Beobachtungen nicht erfolgreich trainiert werden. Durch die lange Laufzeit einer Deduplizierung ist der Ansatz in der Praxis nicht anwendbar. GP produziert hingegen im 2. Training mit Evaluationsdaten sofort hohe und stabile F_1-Werte, was die zu erwartende Dateneffizienz dieses Algorithmus belegt.

[7]https://www.unimedizin-mainz.de/imbei/informatik/ag-verbundforschung/mainzelliste.html

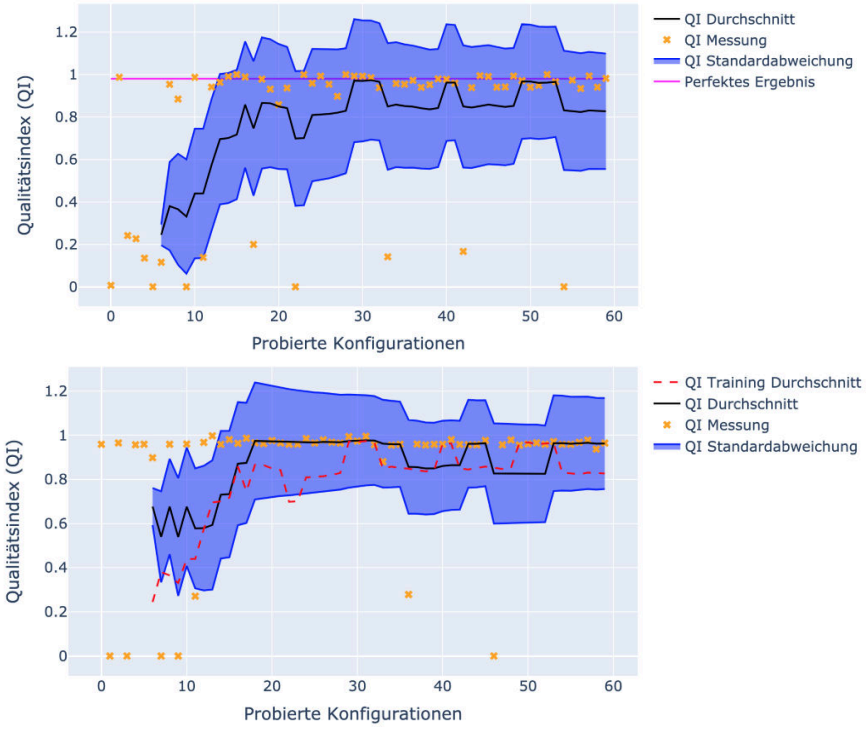


Abb. 6: Training (oben) und Test (unten) mit GP

Anhand der o.g. Datensätze wurden Gaußsche Prozesse als Regressionsmodell ausgewählt und die Evaluation intensiviert. Abbildung 6 zeigt Ergebnisse aus der Trainings- und Testphase. Visualisiert werden einzelne Messungen, Standardabweichung sowie die durchschnittliche Qualität. In der Trainingsphase benötigt GP ca. 30 probierte Konfigurationen, um ein gutes Ergebnis zu erreichen. Der Vergleich erfolgt zu einer manuell definierten bekannten Konfiguration. In der Testphase wird mit einem neuen Datensatz evaluiert und es zeigt sich, dass GP sehr viel früher gute Konfigurationen findet und schon nach 20 Schritten eine gute Konfiguration erreicht. D.h. GP hat ein Modell des Parameterraums des Matchsystems gelernt.

6 Evaluation der Gesamtlösung

Zur abschließenden Bewertung wurde die Gesamtlösung im praktischen Anwendungsfall evaluiert und automatisch erstellte Konfigurationen manuell von erfahrenen UNISERV-Matching-Experten erstellten Konfigurationen gegenübergestellt. Im Folgenden dargestellt ein spezifischer Use Case über die Bereinigung eines Datenbestands aus natürlichen

Personen für den Import in ein CRM-System, bestehend aus 1.262.769 Anschriften aus Deutschland. Der Ursprung der Daten (z.B. Call Center, Werbekampagne o.ä.) ist nicht bekannt. Pflichtfelder sind Ortsname und Nachname. Normale Fehlertoleranz ist gewünscht (Zeichenfehler, Synonyme, Initiale etc.) und mindestens gefordert bei Ortsname, Nachname und PLZ.

Abbildung 7 zeigt die durch einen Matching-Experten manuell erstellte Konfiguration zur gegebenen Problemstellung, mit einer Ergebnisqualität von 92. Die Parameter von Ortsnamen und Nachname sind identisch. Ebenso ähnlich ist die Konfiguration von Straßennamen. Die Parameterwerte werden bei manueller Einstellung (meist) auf volle 10er Schritte gerundet, da eine Mikro-Optimierung enorm aufwendig ist und Kunden meist nicht bereit sind diesen Mehraufwand zu bezahlen.

Abbildung 8 zeigt die durch die Lernkomponente automatisch erstellte Konfiguration zur gegebenen Problemstellung. Für die Gruppenbewertung nach true-positive / false-positive wurden von Fachanwender 250 Dublettengruppen qualifiziert. Nach ca. 35 Minuten konnte die Lernkomponente bereits eine Konfiguration ermitteln, die eine Ergebnisqualität von 98 erzielte. Im Vergleich dazu wurde für die manuell erstellte Konfiguration ein Wert von 92 erreicht.

```
[customer]
elements = zip,street_name,hno_num,last_name,first_name,city_name,[…],phone
#========================+======+======+======+======+======+======+======+======+======+
element = last_name      , 4   , 100 , 90  ,  0  ,  !  , 70  ,  0  ,  !  , 90 ,[…]
element = first_name     , 3   , 100 , 90  ,  0  , -100, 70  ,  0  , 90  , 90 ,[…]
element = name_rest      , 1   , 100 , 90  ,  0  , 0   , 70  , 90  , 90  ,  * ,[…]
element = zip            , 4   , 100 , 90  ,  0  ,  !  , 70  ,  0  , 90  , 90 ,[…]
element = city_name      , 4   , 100 , 90  ,  0  ,  !  , 70  ,  0  , !   , 90 ,[…]
[⋮]
element = phone          , 4   , 100 , 70  ,  0  ,  0  , 75  ,  *  ,  *  ,  * ,[…]
#========================+======+======+======+======+======+======+======+======+======+
min_mval = 70
```

Abb. 7: Konfiguration erstellt von Matching-Experte

Die Ergebnisse der automatisch erstellten Konfigurationen wurden umfassend ausgewertet. Zu bemerken ist:

- Die automatisch ermittelten Parameterwerte sind feingranularer mit 71, 62, 64, etc. Punkten und im Vergleich zur manuellen Konfiguration nicht auf 10er-Schritte gerundet. In der Praxis würde eine Rundung auf „volle 5er-Schritte“ eine ähnliche Qualität liefern bei zeitlich deutlich höherem Aufwand und entsprechender Erfahrung des Experten.

```
[optim_AI]
elements = zip,street_name,hno_num,last_name,first_name,city_name,[…],phone
#======================+======+======+======+======+======+======+======+======+======+
element = last_name      , 4   , 100 , 64  , 62  , -100, 55  , 35  , !   , 99 ,[…]
element = first_name     , 4   , 100 , 71  , 62  , 50  , 70  , 80  , 60  , 85 ,[…]
element = name_rest      , 2   , 100 , 73  , 55  , -50 , 60  , 5   , 65  , 15 ,[…]
element = zip            , 1   , 100 , 76  , 10  , !   , 99  , 30  , 90  , -50,[…]
element = city_name      , 3   , 100 , 67  , 42  , !   , 55  , 0   , !   , 10 ,[…]
[:]
element = phone          , 4   , 100 , 54  , 71  , 50  , 75  , 10  , 25  , 99 ,[…]
#========================+======+======+======+======+======+======+======+======+======+
min_mval = 55
```

Abb. 8: Mittels Lernkomponente automatisch erstellte Konfiguration

- Die Lernkomponente ermittelt einen deutlich tieferen Gesamtschwellenwert min_mval mit 55 Punkten (Mindestübereinstimmung, die erreicht werden muss, damit eine Dublette erkannt wird). Der Matching-Experte hat einen UNISERV-Default von 70 gewählt. Damit nutzt die Konfiguration der Lernkomponente einen breiten Punktebereich aus.

Bei den verschiedenen Tests wurde deutlich, dass die Qualität der automatisch erstellten Konfiguration stark abhängig von der Qualität der Bewertung in true-positive / false-positive (Gold-Standard) durch die Anwender ist. Für eine gute Qualität müssen Anwender etwa 100 Dublettengruppen bewerten. Bei dieser Menge ist in der Praxis immer mit widersprüchlichen Angaben (im Sinne der Matching-Software) durch die Anwender zu rechnen. Hier ist auf ein sorgfältiges Vorgehen zu achten.

7 Zusammenfassung und Ausblick

Deduplizierungssysteme, wie die von UNISERV angebotene „identity“-Lösung für Kunden- und Geschäftspartnerdaten, sind hochkomplex und müssen individuell an die kundenspezifischen Anforderungen angepasst werden. Die spezifische Konfiguration und Erstellung der Matching-Regeln erfordert tiefes technisches und großes Erfahrungswissen. Daher war Ziel eine Lösung mittels lernbasierter Verfahren zu entwickeln, welche die Erstellung von optimalen Matching-Regeln gegenüber dem bisherigen Vorgehen deutlich beschleunigt. Diese Regeln sollen nicht nur von Matching-Experten erstellbar sein, sondern es sollen auch Fachanwender des UNISERV Matching-Systems „identity“ in die Lage versetzt werden, diese Matching-Regeln zu erstellen.

Hierzu wurden verschiedene für ein KMU praxisfähige, lernbasierte Verfahren zur automatischen Konfiguration implementiert und evaluiert. Besonders geeignet erschienen hier Reinforcement Learning Ansätze. Es zeigte sich jedoch, dass aktuelle Reinforcement Learning Techniken nicht den gewünschten Effekt erreichen. Das Problem liegt dabei in der Übersetzung der vielseitigen Aktionen in der Konfiguration eines Matching-Systems auf die möglichen Aktionen eines Agenten. Deutlich bessere Ergebnisse wurden mit Sequential Model-based Optimization (SMBO)-Techniken erreicht. Tests in praktischen Anwendungsfällen mit der lernbasierten Gesamtlösung im Vergleich zu menschlichen Experten zeigen, dass optimale Matching-Regeln von sehr guter Ergebnisqualität auf Basis von Anwenderfeedback in relativ kurzer Zeit erstellt werden können. Mit dieser Lösung können bereits jetzt Junior Consultants von UNISERV in die Lage versetzt werden, in einem Kundenprojekt in kurzer Zeit eine produktionsfähige Konfiguration zu finden. Darüber hinaus kann auch eine Mikrooptimierung erzielt werden, die ansonsten zu teuer und zeitaufwendig ist. Die lernbasierte Komponente wird zukünftig auch Fachanwender (UNISERV-Kunden) befähigen, eine maßgeschneiderte Konfiguration in kurzer Zeit zu erstellen.

Danksagung: Dank an das ZIM-FuE Projekt „KOBRA – Konfiguration von Business-Regeln für Anwender von Duplikaterkennungssystemen“ (Ref. Nr. 16KN061125, https://infai.org/kobra/)

Literaturverzeichnis

[Feu15] Feurer, M.; Klein, A.; Eggensperger, K.; Springenberg, J.; Blum, M.; Hutter, F.: Efficient and robust automated machine learning, In: Adv Neur In, 2962-2970, 2015

[HHL11] Hutter, F.; Hoos, H.H.; Leyton-Brown, K.: Sequential Model-Based Optimization for General Algorithm Configuration. In: Proc. of Learning and Intelligent Optimization, LNCS 6683:507–23. Springer Berlin Heidelberg, 2011

[KR08] Köpcke, H.; Rahm, E.: Training Selection for Tuning Entity Matching. In: 6th Int. Workshop on Quality in Databases and Management of Uncertain Data (QDB/MUD), 2008

[KTR09] Köpcke, H.; Thor, A.; Rahm, E.: Comparative evaluation of entity resolution approaches with FEVER. In: Proc. 35th Int. Conf. on Very Large Databases (VLDB), 2009

[KR10] Köpcke, H.; Rahm, E.: Frameworks for entity matching: A comparison; Data & Knowledge Engineering, 69, 2, Elsevier Science Publishers, 197-210, 2010

[Köp12] Köpcke, H.; Thor, A.; Thomas, S.; Rahm, E.: Tailoring entity resolution for matching product offers. In: Proc. 15th Intl. Conf. on Extending Database Technology (EDBT), 545-550, 2012

[Men16] Mendoza, H.; Klein, A.; Feurer, M.; Springenberg, J.T.; Hutter, F.: Towards Automatically-Tuned Neural Networks. In: Proc. of the Workshop on Automatic Machine Learning, in PMLR 64:58-65, 2016

[Osb16] Osband, I.; Blundell, C.; Pritzel, A.; Van Roy, B.: Deep exploration via bootstrapped DQN. In: Proc. of the 30th Int. Conf. on Neural Information Processing Systems (NIPS'16). Curran Associates Inc., Red Hook, NY, USA, 4033–4041, 2016

[Rao09] Rao, J.; Bu, X.; Xu, C.-Z.; Wang, L.; Yin, G: VCONF: a Reinforcement Learning Approach to Virtual Machines Auto-Configuration. In: Proc. of the 6th Int. Conf. on Autonomic Computing (ICAC '09). ACM, New York, NY, USA, 137–146, 2009

[Sha16] Shahriari, B.; Swersky, K.; Wang, Z.; Adams, R.P.; Freitas, N.D.; Taking the Human Out of the Loop: A Review of Bayesian Optimization. Proc. of the IEEE, 104, 148-175, 2016

[TV20] TV, V.; Malhotra, P.; Narwariya, J.; Vig, L.; Shroff, G.: Meta-Learning for Black-Box Optimization. In: Proc. Of Machine Learning and Knowledge Discovery in Databases (ECML PKDD 2019), LNCS, vol 11907, Springer, Cham, 2020

[Yao18] Yao, Q.; Wang, M.; Escalante, H.J.; Guyon, I.; Hu, Y.-Q.; Li, Y.-F.; Tu, W.-W.; Yang, Q.; Yu, Y.: Taking Human out of Learning Applications: A Survey on Automated Machine Learning. CoRR abs/1810.13306, 2018, https://arxiv.org/abs/1810.13306

[ZL16] Zoph, B.; Le, Q.V.: Neural Architecture Search with Reinforcement Learning. CoRR abs/1611.01578, 2017, https://arxiv.org/abs/1611.01578

Content-based Recommendations for Radio Stations with Deep Learned Audio Fingerprints

Stefan Langer[1] Liza Obermeier[2] André Ebert[3] Markus Friedrich[4] Emma Munisamy[5]
Claudia Linnhoff-Popien[6]

Abstract: The world of linear radio broadcasting is characterized by a wide variety of stations and played content. That is why finding stations playing the preferred content is a tough task for a potential listener, especially due to the overwhelming number of offered choices. Here, recommender systems usually step in but existing content-based approaches rely on metadata and thus are constrained by the available data quality. Other approaches leverage user behavior data and thus do not exploit any domain-specific knowledge and are furthermore disadvantageous regarding privacy concerns. Therefore, we propose a new pipeline for the generation of audio-based radio station fingerprints relying on audio stream crawling and a *Deep Autoencoder*. We show that the proposed fingerprints are especially useful for characterizing radio stations by their audio content and thus are an excellent representation for meaningful and reliable radio station recommendations. Furthermore, the proposed modules are part of the *HRADIO Communication Platform*, which enables hybrid radio features to radio stations. It is released with a flexible open source license and enables especially small- and medium-sized businesses, to provide customized and high quality radio services to potential listeners.

Keywords: Hybrid Radio; Multimedia Services; Recommender Systems; Unsupervised Learning; Deep Audio Fingerprints; Deep Learning

1 Introduction

Despite emerging competition from on-demand content services, linear radio broadcasting still remains one of the most popular entertainment and information media in Europe. Its advantage lies in its technical simplicity, its topicality, and its personal approach conveyed by professional moderators. However, with services like Spotify, Deezer, or Google Play, strong competitors have recently appeared. Those on-demand media services have the

[1] Ludwig-Maximilians-Universität München, Institut für Informatik, Oettingenstrasse 67, 80538 Munich, Germany, stefan.langer@ifi.lmu.de
[2] inovex GmbH, Data Management and Analytics, Lindberghstrasse 3, 80939 Munich, Germany, lobermeier@inovex.de
[3] inovex GmbH, Data Management and Analytics, Lindberghstrasse 3, 80939 Munich, Germany, aebert@inovex.de
[4] Ludwig-Maximilians-Universität München, Institut für Informatik, Oettingenstrasse 67, 80538 Munich, Germany, markus.friedrich@ifi.lmu.de
[5] Ludwig-Maximilians-Universität München, Institut für Informatik, Oettingenstrasse 67, 80538 Munich, Germany, emma.munisamy@ifi.lmu.de
[6] Ludwig-Maximilians-Universität München, Institut für Informatik, Oettingenstrasse 67, 80538 Munich, Germany, linnhoff@ifi.lmu.de

doi:10.18420/inf2020_38

advantage of providing a personalized listening experience and a wide range of contents combined with precise recommendation systems.

The challenge for radio broadcasters is now to enrich their classic, linear radio programme with online-based, personalized technologies in order to improve the listening experience and to bridge the gap between linear and on-demand content providers. So-called hybrid radio technologies comprise of techniques for privacy-preserving user data collection that are capable of providing individual and secure feedback channels. These channels enable customers to actively take part in the composition of their radio programme. Another important enhancement compared to the established linear programme design is the on-the-fly substitution of content, e.g., replacing ads with songs from a pre-selected music playlist or the skipping of disliked content from one radio station with preferred content from another radio station. In combination with other hybrid radio technologies, such features help mainstream stations as well as small radio businesses to catch up with online streaming services and to even surpass them in specific disciplines, e.g., interactivity and feedback channel communication. To realize such rich features, a precise and comprehensive recommender engine is necessary. Therefore, on the one hand a large amount of well-structured and rich data is needed, on the other hand different concepts for the analysis of meta and audio data have to be selected, implemented and carefully evaluated.

This paper focuses on the latter and thus on the question of how meaningful recommendations for radio stations can be generated on the basis of rich metadata provided by the *HRADIO Communication Platform*, which was implemented by the authors within the scope of previous work [Fr19]. The findings and implemented concepts presented in this work are part of the open accessible HRADIO project[7]. In the following, two main approach categories are distinguished for analyzing and recommending radio stations and programmes: Recommendations based on *Collaborative Filtering* (CF) and *Content-based Filtering* (CBF) approaches. CF focuses on user opinions and behavior, which can lead to privacy issues as well as the *Cold Start Problem* (see Section 2). Thus, a CBF-based recommender system working with characteristics derived from available station data is preferred in context of this work. In [Fr19], it could be shown that station recommendation based on available metadata is possible, but only if the metadata reaches a certain quality level. This is provably often not the case. Other meaningful characteristics can be directly derived from the station's audio signal which significantly reduces metadata quality requirements. Therefore, we propose a *Deep Learning*-based audio crawling and fingerprint extraction pipeline for the characterization of radio stations and show visual results for numerous stations. Furthermore, we detail on how a recommender system based on the developed station fingerprints can be implemented.

Due to these unique technical circumstances as well as the fact that the *HRADIO Communication Platform* is released under a permissive open source license and available at

[7] `https://www.hradio.eu/` - The HRADIO project and thus this work was funded by H2020, the EU Framework Programme for Research and Innovation. is driven by several radio stations, small development companies, technology experts, research institutes and the Ludwig-Maximilians-University Munich

no cost, new opportunities are offered especially for small- and medium-sized businesses within the radio landscape. It significantly increases the visibility of radio stations and makes them searchable via a search service that can be used by websites and mobile applications regardless of the size of the company. For this purpose, the daily programme and the audio stream of a station are analyzed. This is done automatically, without additional effort and without costs. This creates enormous added value for both radio stations and potential listeners. In particular, small or regional stations that were previously unknown to non-regional listeners can now be found on the basis of a listener's programme taste. In addition, previously untapped potentials are opened up with regard to findability, routing and provision of niche topics as well as the allocation of tailor-made advertising relevant only to certain listener groups.

To present the underlying concept, its implementation, and its evaluation of these ideas, this paper is structured as follows: Section 2 provides a brief overview of recommendation concepts and related work. Section 3 explains the basic principles of the proposed fingerprinting pipeline and its implementation. Its recommendation capabilities are evaluated in Section 4 while Section 5 summarizes this work.

2 Related Work

Next to recommendation approaches such as *You May Like* (YML), *Knowledge-based Filtering* (KF), and *Demographic Filtering* (DF), there are two mainly recognized concepts: *Content-based Filtering* (CBF) and *Collaborative Filtering* (CF) [Be07, Bu02, RRS11]. CF-, DF-, and KF-based systems show good performance if enough user data is available. But an open issue is the so-called *Cold Start Problem*, which occurs in the initial phase of the system where not enough user data is available to create meaningful recommendations [Ch15, La08, CAS18]. Another issue for CF and DF systems are so-called *Filter Bubbles*, describing the creation of closed-off, synthetic environments, in which always the same items are recommended, disregarding the existence of contrary or different items beyond the bubble [Pa11, GLH19]. In contrast to that, an initial issue of CBF-based systems is the need for high-quality metadata precisely describing the items to recommend [Fr19]. The concept presented in this paper is part of the *HRADIO Communication Platform*, which provides a vast amount of metadata and audio information within a hybrid radio context [Fr19] and is completely open source. For this reason, and to avoid issues like *Filter Bubbles* or the *Cold Start Problem*, it takes a CBF-based approach.

In order to compare radio services on basis of their audio features, existing approaches which utilize *Deep Learning* for music genre recognition (MGR) can be utilized [DBS11, Or18]. Gwardys et al. propose a concept using transfer learning in combination with a convolutional neural network for MGR [GG14]. Logan et al. and Siddiquee et al. present methods for measuring the similarity of music on basis of audio signals [LS01, Si16]. After clustering raw audio features, Logan et al. compare entities using the *Earth Movers Distance* (EMD) [RTG00]. Çataltepe et al. use adaptive features and user grouping to take note to the aspect

that different traits within music are of different importance for each user by including historical information about the users' listening behaviour [ÇA07]. Together with others, these works provide valuable input for the proposed concept.

3 Concept

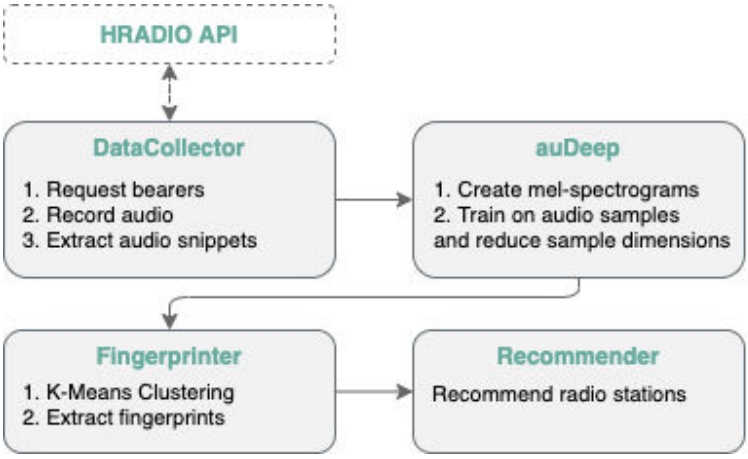


Fig. 1: The proposed pipeline consists of four modules: the *Data Collector*, the *auDeep* autoencoder, the *Fingerprinter*, and the *Recommender*. The *Data Collector* ensures the binding to the *HRADIO API*.

This chapter details the steps of the proposed analysis pipeline consisting of a *Data Collector*, the *auDeep* autoencoder, the *Fingerprinter*, and a *Recommender*, as depicted in Figure 1. In this context, the *Data Collector* unit (see Chapter 3.1) ensures the binding of the pipeline to the *HRadio API* (which is used for requesting a list of radio services from the *HRADIO Communication Platform* [Fr19]), records their audio streams and extracts audio bytes in predefined intervals. The raw audio snippets are transformed into mel-scaled spectrograms. These spectrograms express human perceptible image representations of audio signals and serve as training input for the *Deep Neural Autoencoder* provided by the *auDeep* toolkit. It is trained to reduce the dimensions of the audio representations [Fr17]. By applying the trained encoder component in our concept, the input data is compressed and the samples are reduced to vectors with 1024 dimensions. These vectors are the input of the *Fingerprinter*, which trains a *K-Means* clustering model [Pe11]. The distribution of samples in each cluster per radio station is regarded as the station's fingerprint. The last component shown in Figure 1 is the *Recommender*, which recommends radio stations similar to a particular input station on basis of the Euclidean distance of their fingerprints. A small distance implies a comparably high similarity (see Section 3.3 for more details).

3.1 Data Collector

The *DataCollector* unit (see Figure 1) consists of several components that collect and process radio data. One of them records radio stations by requesting a list of radio services and their HTTP *bearers*[8] from the *HRADIO Communication Platform* [Fr19] via its REST API. Currently, a list of 461 valid and unique streams is received. The HTTP stream address of a sender is queried in order to receive the audio bytes of the corresponding audio stream using ICY ('I Can Yell') tag technology. The ICY tag protocol defines the transmission of textual content within audio streams and is commonly used due to its simple integration. Using *RabbitMQ*[9] the received audio bytes are transferred to a separate component which is responsible for extracting audio snippets out of the radio streams. During 24 hours, each station is recorded for five seconds within intervals of two minutes. In order to reduce the amount of news sequences included in the recorded snippets, the five minutes before and after full hours are not recorded. This leads to a total number of 576 samples per radio station. Thereby, a full day of samples for 431 radio stations could be recorded, while 30 stations could not be recorded entirely due to server-side connection problems. In total, we collected 266,239 audio snippets, whereas 17,983 belong to incomplete radio station recordings.

3.2 Deep Fingerprints

The audio signals extracted in the previous step are transformed into a time-frequency representation, called mel spectrogram (see Figure 2), using the *auDeep* toolkit. Mel spectrograms are close to the human perception of audio signals with the mel scale being an assignment of actual to perceived frequencies. In the example in Figure 2, time is depicted on the x-axis and f frequencies are shown on the y-axis. In addition, the colour represents frequency amplitudes (or loudness): dark areas indicate low amplitudes, bright areas indicate high (or loud) amplitudes. For mel spectrogram generation we use the hyperparameters suggested by *auDeep* (a mel scale of magnitude 128, a window width of 0.08 with an overlap of 0.04, a fixed length of 5 seconds per input snippet, and clipping of values below $-60db$) [Fr17].

Subsequently, the *auDeep* autoencoder was trained on all 266,239 audio files, whereas the fingerprinting is only applied to complete sets of audio snippets (for 431 radio stations). The network is trained across 64 epochs with a batch size of 64 on 2 layers with each having 256 gated recurrent units (GRU), a learning rate of 0.001 and a dropout of 0.2. Training the network for 7 days resulted in a loss of 0.237. After the training is finished the *Fingerprinter* creates fingerprints of all complete stations, using all 266,239 vectors generated by the *auDeep* component. All these data points are divided into n clusters by applying the *K-Means* clustering algorithm. The parameter n is determined by using the

[8] Connection information, how a radio station can be received by a device (broadcast or streaming)
[9] An open source message broker (`https://www.rabbitmq.com/`)

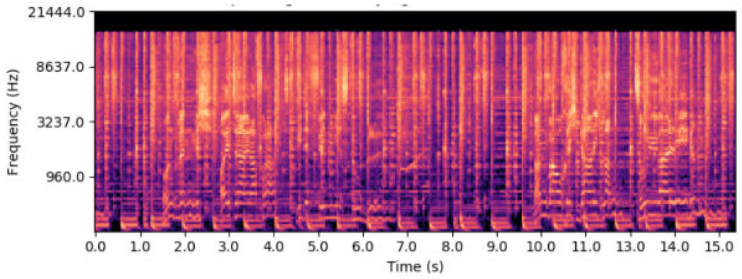


Fig. 2: An example of a mel spectrogram with 128 mel features. The x-axis represents the time scale and the y-axis the frequency scale.

silhouette coefficient [AT07], resulting in a range from 9 to 16 and showing a peak value at 11 as can be seen in Figure 3. Each data point is assigned to exactly one cluster and the fingerprint is derived from a histogram across all clusters for each station. This fingerprint vector now serves as input for the *Recommender*.

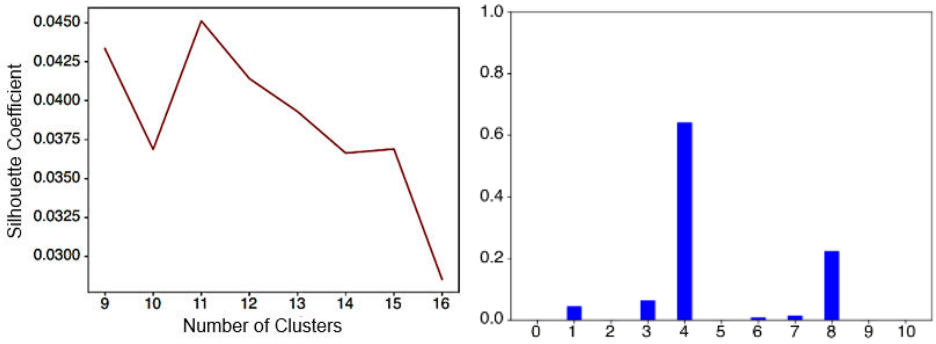


Fig. 3: Silhouette coefficents for different numbers of clusters (left) with an optimum cluster count of 11. On the right side a sample histogram of *BR KLASSIK* is depicted after the feature extraction process. The cluster index is depicted on the x-axis whereas the per-cluster frequency of occurrence is on the y-axis.

3.3 Recommender System

Based on learned per-station fingerprints (see Section 3.2) it is possible to define a similarity metric for radio stations which is essentially the Euclidean distance between fingerprint vectors, considering the space of all 11-dimensional fingerprint vectors as an Euclidean

vector space. This similarity metric can be used to establish a content-based comparison of radio stations (the fingerprint represents the content of a station): A small distance between two station fingerprints implies high similarity. Using this mechanism for recommendations, there are essentially two possibilities:

1. The k nearest radio stations are suggested to the user as similar.
2. Only radio stations within a certain Euclidean distance are listed.

The first possibility carries the risk that the returned station list may also contain distant radio stations. In contrast to that, for the second advance, the distances of provided radio stations are small in any case, while it carries the risk that the result set may be empty.

Because of the subjectivity of user taste and ratings, providing high-quality recommendations is not a trivial task and cannot only be based on station fingerprints. The *HRADIO Communication Platform* offers therefore a multitude of different recommendation strategies with the fingerprint-based recommender being one of it. The different modules are highly configurable and can be combined to a customized radio station recommendation experience. Examples for additional modules are trend-, metadata- and location-based recommenders.

4 Evaluation

In this section, the recommender system is evaluated quantitatively and its results are discussed by example (Section 4.1). Furthermore, we were interested in identifying station archetypes that represent certain categories (or genres) of stations. For that purpose, an *Archetypal Analysis* of the station fingerprint dataset was conducted with its results being presented in Section 4.2. Finally, fingerprint deltas between different times of day are analyzed and discussed while highlighting some significant examples in Section 4.2.1.

4.1 Recommender System

The evaluation of the recommender system is based on the station fingerprint dataset and corresponding genre labels that stem from existing metadata as visualized in Figure 5. In the following, we show how the distance measures map to these genre labels by choosing the most meaningful examples.

Table 1 shows the three closest radio stations (according to their fingerprint distance) with their distance to a requested service. The station *BR Klassik* is assigned to the genres *Classical Music* and *Special Music*. The closest 3 stations are *NDR Kultur* with the genres *Classical Music* and *Cultural* with an Euclidean distance of 59.58, *HR 2* with the genres *Classical Musik*, *Cultural*, and *Special Music* and an Euclidean distance of 60.81, and *Classic FM* with the genres *Classical Music* and *News* and an Euclidean distance of 60.93.

Requested station	1st closest station	2nd closest station	3rd closest station
BR Klassik ***Classical Music*** ***Special Music***	NDR Kultur *Classical Music* *Cultural* distance: 59.58	HR2 *Classical Music* *Cultural* distance: 60.81	Classic FM *Classical Music* *News* distance: 60.93
Heart UK ***Classic/Dance/Pop-rock*** ***Disco***	Capital XTRA Reloaded *Rap/HipHop/Raggae* distance: 41.09	3FM Isle of Man *Hit-Chart* distance: 86.34	FFH Rock *Rock* *Soft Rock* distance: 88.29
LBC UK ***Local/Regional*** ***News***	Bayern 5 Plus *Information* distance: 117.97	WDR 3 *Classical Music* *Cultural* distance: 118.22	SWR 2 Archiv Radio *Documentary* distance: 118.25

Tab. 1: Three examples of recommendation requests and the three closest results including their Euclidean distances.

All 3 stations near to *BR Klassik* play similar content within the genres *Classical Music* and *Cultural*. The second radio station *Heart UK* is assigned to the genres *Classic/Dance/Pop-rock*, *Disco*, *Local/Regional*, *Dance/Dance-pop*, and *Showbiz*. The closest three stations are *Capital XTRA Reloaded* with the genres *Rap/Hip Hop/Reggae* with an Euclidean distance of 41.09, *3FM Isle of Man* with the genre *Hit-Chart* and an Euclidean distance of 86.34, and *FFH Rock* with the genres *Rock*, *Soft Rock*, *Grunge*, *Heavy Rock*, and *Rock & Roll* and an Euclidean distance of 88.29.

The only station close to *Heart UK* by genre is *3FM Isle of Man*. The other two stations can be considered not similar, being assigned to *Rap/HipHop/Raggae* versus *Rock, Soft Rock, Grunge, Heavy Rock, Rock & Roll*. The third radio station *LBC UK* is assigned the genres *Non-fiction*, *Local/Regional*, and *News*. The closest 3 non-variant stations are *Bayern 5 Plus* with the genre *Information* with an Euclidean distance of 117.97, *WDR 3* with the genres *Classical Musik* and *Cultural* with an Euclidean distance of 118.22, and *SWR 2 Archiv Radio* with the genre *Documentary* and an euclidean distance of 118.25.

All three stations near to *LBC UK* publish a lot of spoken content, as suggested by their genres *Local/Regional*, *News*, or *Documentary*.

As we could show and explain by example, the proposed station fingerprints in combination with the discussed similarity metric is a well-working approach for content-based station recommendations which outputs similar stations for a given station.

4.2 Archetypal Analysis

In order to distill station categories, we conducted an *Archetypal Analysis* on the fingerprints which is an unsupervised learning method used to extract representative individuals in a dataset. A data point is defined by its affiliation to k archetypes [CB94]. In soccer, for example, a player could be described by 10% defender, 50% midfielder, and 40% striker. We use this concept to find representative stations on basis of their fingerprints. In our case, the optimal amount of four archetypes was determined using the so-called *elbow criterion* [EL09]. Therfore, the residual sum of squares (RSS) for different numbers of archetypes is visualized in a scree plot (see Figure 4). The optimal number of archetypes is the one where the curve has its strongest bend.

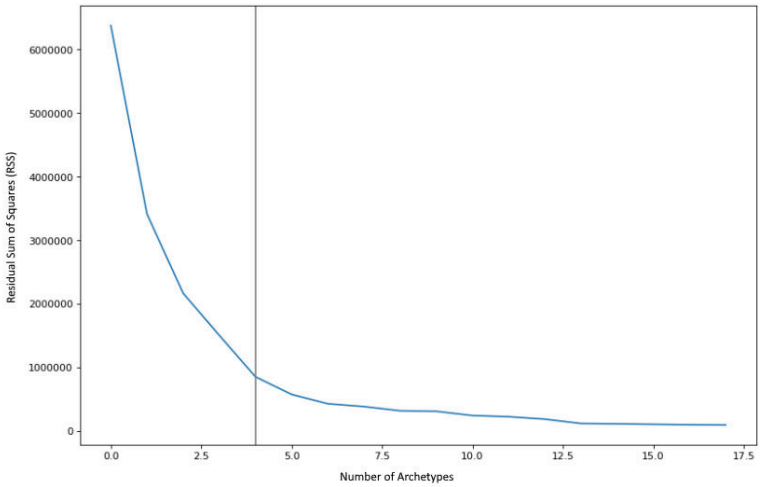


Fig. 4: Number of archetypes and corresponding RSS values. This so-called scree plot is used for the selection of the best number of archetypes.

Figure 5 shows a plot of all fingerprints, reduced to 2 dimensions by using a *Principal Component Analysis (PCA)* [Jo11], where stations are points coloured according to their genre. Archetypes are represented as black triangles. The first archetype is at [115.77, −164.82]. *Antenne P* defined as genre *Oldies* is the closest station to it with a distance of 19.86. Within a radius of an Euclidean distance of 150, 71 stations could be found in total.

SWR1 RP is nearby next with a distance of 51.90, encompassing the genres *Pop Music* and *Regional*. The second archetype is at [292.96, 385.21]. *Noods Radio* is the closest station with a distance of 19.23 not listing any genres. Within a radius of an Euclidean distance of 150, 7 stations could be found in total. *95.3 KGY Olympia* is the next nearest station with a distance of 41.56, classified as *AOR / Slow Rock / Soft Rock*.

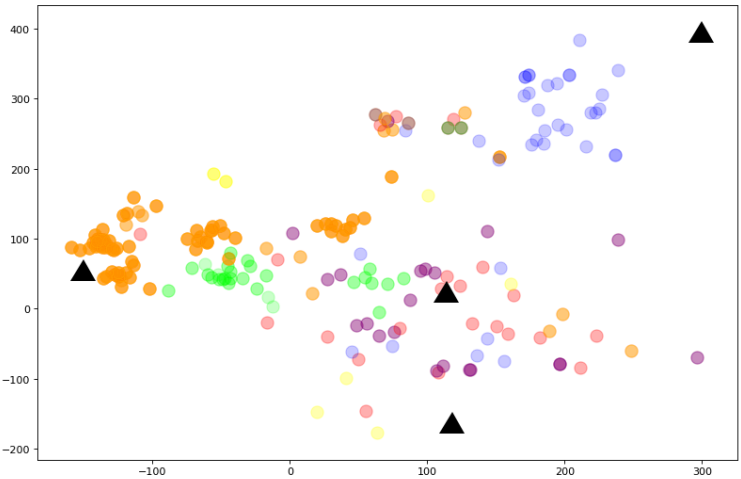


Fig. 5: Visualization of fingerprint vectors, using PCA. The dots mark the radio services, colored by genre. Black triangles represent archetypes.

The third archetype is located at $[-144.58, 50.05]$. *Heart Crawly* has a distance of 8.76 to it, encompassing the genres *Classic/Dance/Pop-rock*, *Disco, Local/Regional, Dance/Dance-pop*, and *Showbiz*. Within a radius of an Euclidean distance of 150, 175 stations could be found in total. The next other station providing genre metadata is *Heart Dorset* with a distance of 11.32, classified by the same genres.

The last archetype is at $[138.10, 60.20]$. *Bayern 2 Nord* is the closest station with a distance of 88.83 and genre *Cultural*. Within a radius of an Euclidean distance of 150, 48 stations could be found in total. The next other station offering genre meta data is *Classic FM* with a distance of 99.47 to the archetype, playing *Classical Music* and *News*.

In summary, the four found archetypes map well on genres and also target groups: The first archetype corresponds to stations with a target group that likes to listen to oldies / old pop music and demands for regional information. The second archetype has nearby stations with a very diverse programme that does not fit into specific music genres (*Noods Radio* for example plays almost everything from post punk to electro) and thus has a very broad target group. The third archetype represents stations with a focus on dance music whereas the fourth archetype is related to classical music and news. Interestingly, there is no specific archetype for contemporary pop music which could be due to the fact that this kind of music is more or less a mixture of existing genres.

4.2.1 Comparing Recommendations by Day Times

It is well known that radio stations change their programme over the day with, for example, fast wake-up music in the morning and slower, relaxing music in the evening. Our hypothesis is, that this effect also manifests in the fingerprints of a station extracted for different day times.

In order to test this hypothesis, we created three additional, normalized fingerprints per times of day, namely night, morning, and day. Night fingerprints consider audio samples from 09:00 pm to 05:00 am, morning fingerprints consider samples from 05:00 am to 09:00 am, and day fingerprints consider samples from between 09:00 am and 09:00 pm. See Figure 6 for a visualization of the different fingerprints using per-station trajectories. Comparing those three fingerprints to each other and to the whole-day fingerprints provides interesting insights into how radio stations change throughout the day.

E.g., large changes appear on *Gold 60s* which had comparably large distances between morning (square) and other times of the day (triangle, diamond) and whole day (circle) fingerprints, leading to the assumption that the style of content changes heavily between morning and rest of the day. This pattern is also visible in other stations' trajectories (e.g. *FFH Die 90er*, *FFH BrandNeu* and *106 Jack FM*). On the other hand, some stations show only minor (*FFH Workout*) or moderate (*Radio City Talk*) changes over the day.

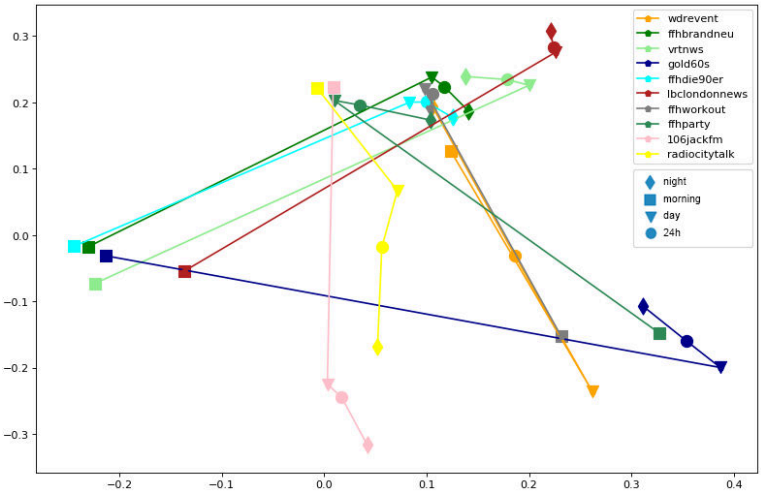


Fig. 6: Fingerprint trajectories visualizing deltas for stations for different times of day. Trajectory color corresponds to the station, circle shapes are whole-day fingerprints (24h), diamonds night, squares morning and triangles day time.

An additional hint for daytime-based changes visible in the extracted fingerprints is depicted in Figure 7. The day time and morning time archetypes (light blue, yellow) significantly differ from the night time archetypes (dark blue) which clearly shows that the assumed programme difference dependent on the time of day is visible in the fingerprint data-set as well.

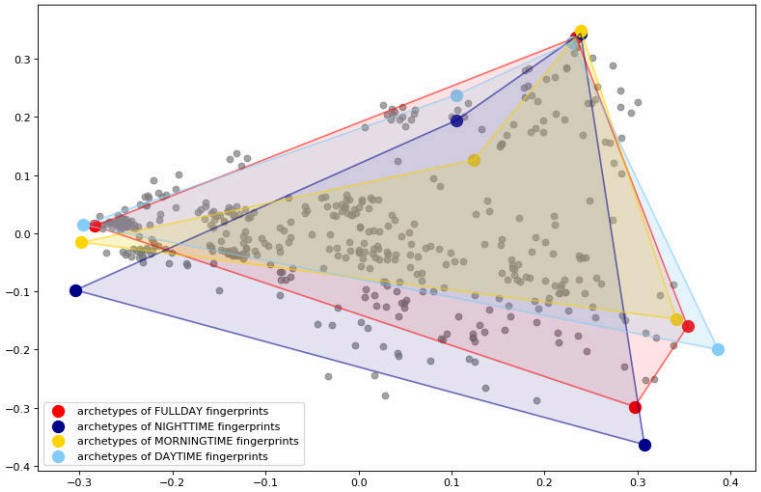


Fig. 7: Extracted archetypes for different times of day.

To summarize, we could prove our hypothesis, that time dependent radio content differences are visible in the fingerprint data by showing exemplary per-station differences and differences in the shapes and locations of the archetypes.

5 Conclusion

In this paper, we presented a holistic pipeline for deep radio station fingerprinting and recommendation. The first component utilized throughout the process is the *HRADIO Communication Platform*, which delivers metadata and bearer addresses of radio stations. The next component, the *DataCollector* recorded 461 radio stations over 24 hours, generating a total of 266,239 audio samples. Following this step, we trained a *Deep Neural Autoencoder* with *auDeep* in order to reduce each sample's dimensions. The *Fingerprinter* then clustered the samples and created a deep fingerprint for each service. The last component, the *Recommender*, is able to give recommendations depending on the calculated distance between the fingerprints.

We evaluated the *Recommender* by analyzing the recommendation results and noticed similar genres for close radio stations. Additionally, we conducted an *Archetypal Analysis* on the fingerprints, leading to four archetypes. By analyzing their closest radio stations, we noticed dissimilar target groups and genres between the archetypes and similar ones between the closest stations surrounding them.

Finally, we compared whole-day fingerprints to fingerprints extracted in the night, morning and day time. The current time of day seems to have a big influence on recommendations for some services. In contrast to that, others stayed somewhat constant throughout 24 hours. The data-set we created offers many more opportunities for further research and analysis. To further validate this, user studies should follow with which the quality of a recommendation can be quantified. In addition, the night-morning-day approach could be evaluated in far more depth.

We see three important practical use cases for the aforementioned techniques and findings:

- The proposed content-based recommender can complement existing recommender engines that mostly use collaborative filtering which is rather domain-unspecific. Within the HRADIO project, a station recommender system was developed, that uses a fingerprint-based recommendation module together with modules based on metadata, location and user behavior.[10]
- The described audio analysis pipeline can be used to automatically extract station and programme metadata (e.g. genre-like classifications) which is a big step forward considering the current situation regarding metadata availability in the radio domain.
- The extracted fingerprints and the proposed distance metric can also be helpful for radio stations that would like to check if their current play-out matches with the station's target programme and groups. This check can be conducted automatically and could provide even more detailed guidance like what to play in the next two hours to match the taste of a certain target group.

In summary, this approach provides valuable recommendations only on basis of audio signals and without the need of additional metadata. Moreover, the presented concepts are included and available within the open *HRADIO Communication Platform* and thereby enable also small businesses and radio stations to actively take part in a radio service landscape driven by smart services.

Bibliography

[AT07] Aranganayagi, S; Thangavel, K: Clustering categorical data using silhouette coefficient as a relocating measure. In: International Conference on Computational Intelligence and Multimedia Applications (ICCIMA 2007). volume 2. IEEE, pp. 13–17, 2007.

[10] Android App that uses the modularized station recommender: `https://play.google.com/store/apps/details?id=lmu.hradio.hradioshowcase`

[Be07] Bennett, James; Lanning, Stan et al.: The netflix prize. In: Proceedings of KDD cup and workshop. volume 2007. New York, NY, USA, p. 35, 2007.

[Bu02] Burke, Robin: Hybrid recommender systems: Survey and experiments. User modeling and user-adapted interaction, 12(4):331–370, 2002.

[ÇA07] Çataltepe, Zehra; Altinel, Berna: Music recommendation based on adaptive feature and user grouping. In: 2007 22nd international symposium on computer and information sciences. IEEE, pp. 1–6, 2007.

[CAS18] Camacho, Lesly Alejandra Gonzalez; Alves-Souza, Solange Nice: Social network data to alleviate cold-start in recommender system: A systematic review. Information Processing & Management, 54(4):529–544, 2018.

[CB94] Cutler, Adele; Breiman, Leo: Archetypal analysis. Technometrics, 36(4):338–347, 1994.

[Ch15] Chang, Shiyu; Zhou, Jiayu; Chubak, Pirooz; Hu, Junling; Huang, Thomas: A space alignment method for cold-start tv show recommendations. In: Twenty-Fourth International Joint Conference on Artificial Intelligence. 2015.

[DBS11] Dieleman, Sander; Brakel, Philémon; Schrauwen, Benjamin: Audio-based music classification with a pretrained convolutional network. In: 12th International Society for Music Information Retrieval Conference (ISMIR-2011). University of Miami, pp. 669–674, 2011.

[EL09] Eugster, Manuel; Leisch, Friedrich: From spider-man to hero-archetypal analysis in R. 2009.

[Fr17] Freitag, Michael; Amiriparian, Shahin; Pugachevskiy, Sergey; Cummins, Nicholas; Schuller, Björn: audeep: Unsupervised learning of representations from audio with deep recurrent neural networks. The Journal of Machine Learning Research, 18(1):6340–6344, 2017.

[Fr19] Friedrich, Markus; Ebert, André; Hahn, Carsten; Schneider, Georg; Obermeier, Liza; Erk, Alexander; Jennes, Iris: A Distributed Metadata Platform for Hybrid Radio Services. In: International Conference on Innovations for Community Services. Springer, pp. 166–183, 2019.

[GG14] Gwardys, Grzegorz; Grzywczak, Daniel: Deep image features in music information retrieval. International Journal of Electronics and Telecommunications, 60(4):321–326, 2014.

[GLH19] Geschke, Daniel; Lorenz, Jan; Holtz, Peter: The triple-filter bubble: Using agent-based modelling to test a meta-theoretical framework for the emergence of filter bubbles and echo chambers. British Journal of Social Psychology, 58(1):129–149, 2019.

[Jo11] Jolliffe, Ian: Principal component analysis. Springer, 2011.

[La08] Lam, Xuan Nhat; Vu, Thuc; Le, Trong Duc; Duong, Anh Duc: Addressing cold-start problem in recommendation systems. In: Proceedings of the 2nd international conference on Ubiquitous information management and communication. ACM, pp. 208–211, 2008.

[LS01] Logan, Beth; Salomon, Ariel: A Music Similarity Function Based on Signal Analysis. In: ICME. pp. 22–25, 2001.

[Or18] Oramas, Sergio; Barbieri, Francesco; Nieto, Oriol; Serra, Xavier: Multimodal deep learning for music genre classification. Transactions of the International Society for Music Information Retrieval. 2018; 1 (1): 4-21., 2018.

[Pa11] Pariser, Eli: The filter bubble: What the Internet is hiding from you. Penguin UK, 2011.

[Pe11] Pedregosa, Fabian; Varoquaux, Gaël; Gramfort, Alexandre; Michel, Vincent; Thirion, Bertrand; Grisel, Olivier; Blondel, Mathieu; Prettenhofer, Peter; Weiss, Ron; Dubourg, Vincent et al.: Scikit-learn: Machine learning in Python. Journal of machine learning research, 12(Oct):2825–2830, 2011.

[RRS11] Ricci, Francesco; Rokach, Lior; Shapira, Bracha: Introduction to recommender systems handbook. In: Recommender systems handbook, pp. 1–35. Springer, 2011.

[RTG00] Rubner, Yossi; Tomasi, Carlo; Guibas, Leonidas J: The earth mover's distance as a metric for image retrieval. International journal of computer vision, 40(2):99–121, 2000.

[Si16] Siddiquee, Md Mahfuzur Rahman; Rahman, Md Saifur; Chowdhury, Shahnewaz Ul Islam; Rahman, Rashedur M: Association rule mining and audio signal processing for music discovery and recommendation. International Journal of Software Innovation (IJSI), 4(2):71–87, 2016.

Towards Collaborative Predictive Maintenance Leveraging Private Cross-Company Data

Marisa Mohr [1,2] Christian Becker[2] Ralf Möller [1] Matthias Richter [2]

Abstract: The accuracy of a predictive maintenance model is largely determined by the available training data. This puts such machine learning systems out of reach for small and medium-sized production engineering companies, as they are often unable to provide training data in sufficient quality and quantity. Building a collaborative model by pooling training data across many companies would solve this issue, but this data cannot simply be consolidated in a central location while at the same time preserving data integrity and security. This paper enables a collaborative model for predictive maintenance on cross-company data without exposing participants' business information by connecting two recent methodologies: blockchain and federated learning.

Keywords: Industrial Internet of Things; Machine Learning; Blockchain; Federated Learning

1 Introduction

Data-driven products and machine learning (ML) methods offer large benefits for production engineering companies. For some, application of such methods may even be necessary to remain competitive. Intelligent planning of maintenance windows, for example, decreases the risk of unwanted production downtimes and helps to keep machines in their optimal conditions. However, development of such products requires a large initial investment in model definition and training data acquisition. The latter is especially important, as the prediction quality of an ML model is largely determined by the data used for its training. If these data do not contain the patterns preceding a failure, the model will be unable to predict impending machine failures from new, previously unseen data. In predictive maintenance, this issue is made worse by the relatively rare occurrence of machine failures. Of course, one could deliberately allow machines to degrade to capture more failure patterns, but this is at the very least fiscally irresponsible – especially for small and medium-sized enterprises (SMEs).

A more sensible approach would be to share data across multiple SMEs or machine manufacturer and machine users in order to train a more powerful model than one contributor could do on their own. There are, however, two key challenges with this approach: data integrity and data security. In this work-in-progress paper, we discuss these

[1] University of Lübeck, Institute of Information Systems, Ratzeburgerallee 160, 23562 Lübeck, Germany, {mohr,moeller}@ifis.uni-luebeck.de, https://orcid.org/0000-0003-0006-6141, https://orcid.org/0000-0002-1174-3323

[2] inovex GmbH, Ludwig-Erhard-Allee 6, 76131 Karlsruhe, Germany, {mmohr,cbecker,mrichter}@inovex.de, https://orcid.org/0000-0002-4917-7574

doi:10.18420/inf2020_39

challenges and combine recent methods, blockchain and federated learning, to enable collaborative predictive maintenance.

2 Use Case: Collaborative Predictive Maintenance

Predictive maintenance systems estimate the time until a machine will likely fail, highlight possible problems with complex machines and identify the parts that need to be repaired. There are several ways to model each of these individual tasks, but the underlying data are typically time series, e.g., of sensor readings. Classification or prediction models, classical or based on deep learning, learn patterns and regularities from a large amount of historical data [Mo20, Ma18]. As more data generally leads to better forecasts [HNP09], companies could work together to pool their data and create collaborative models for predictive maintenance. If machine manufacturers, their machine users and also third parties such as service partners are among the participants, value chains can be expanded to create value networks.

Pooling the data is not without issues though. Consider a production plant equipped with a number of sensors. The data from these sensors are passed through several systems such as the controller, gateway or internal servers and finally stored in a database. Within one company, there is little need to protect the data in this chain against inside manipulation: there is no incentive for manipulation. However, there is a reason for doing so when exchanging data with external partners: a malicious actor could corrupt their data in order to sabotage or subtly change the model predictions in their favour [Ba18]. A second challenge lies in transparent documentation of maintenance procedures. In the event of a machine failure, faulty or incomplete maintenance can be used to draw conclusions about the failure, thus proving both warranty claims by the machine manufacturer and the fault of the machine operator. Finally, the third major challenge is the collaborative training of models for predictive maintenance without disclosing business information from the individual participants [BM20a].

3 Methods

This section introduces the two methodologies that contribute to collaborative predictive maintenance with respect to data integrity and data privacy.

3.1 Blockchain: Protection against Forgery, Consensus, and More Concepts

Sharing data in a consortium requires that the data is verifiable and can't be forged. Ideally, data would be signed where it is recorded (i.e., in the sensor) before it is transferred to a downstream system. However, commercially available systems that implement such approaches may be incompatible between different manufacturers. Korb et al. [Ko19]

propose to install a microcontroller directly between a sensor and a downstream system instead. By dispensing with proprietary operating systems, there is minimal scope for manipulation and unwanted changes in the processing logic of the data. By implementing a blockchain on the microcontroller, it can be ensured that the data is forgery-proof using suitable signing methods and cryptographic methods.

A more important aspect is the consensus mechanism offered by a blockchain. In short, consensus mechanisms are protocols that ensure that all participants are synchronized with each other and agree on which transactions are legitimate and included in the blockchain. This procedure enables a transparent maintenance log based on error messages and maintenance entries. The tuple of an entry in the maintenance log, for example, an error message or a performed maintenance activity, is cryptographically hashed and the hash is written to the blockchain. In the maintenance log, the hashes of the respective entry and the hashes of the previous entry, if any, are linked together. This allows complete traceability of the entries and the participants are able to check the authenticity of the entries themselves at any time. This procedure does not prevent unauthorized manipulation of the data, but would make it obvious to the other participants. Ultimately, this mechanism helps to build trust between parties, which benefits everyone in the end.

3.2 Federated Learning: Collaborated Models with Private Data

Still, participants may be reluctant or even unable to share their data with collaborators, as doing so might expose trade secrets or violate data protection regulation. Federated learning (FL) solves this issue. In FL, each participant trains an ML model on their private data and using their own hardware [Mc16, Li20]. These models are then aggregated by a central curator to form a unified global model that has learned from every participants' private data without ever directly accessing it.

The FL algorithm works in several rounds. Initially, the central curator selects a list of the participating machines – the workers – using some candidate selection algorithm, e.g. [Hu20], to ensure that only cooperative participants with suitable data may take part in the training. The current global model is sent to all participating workers. They then use their private data to derive an updated local model and send it back to the curator. Finally, the curator aggregates the local models to an updated global model, e.g., using a weighted mean, where the weights are proportional to the number of training samples used by the workers. This process is repeated until the model is accurate enough, the number of rounds exceeds a threshold, or some other termination criterion is fulfilled.

With FL, the training data is not required to be centralized, but can instead remains with the owners. Nevertheless, private data may still be extracted from the global model as demonstrated by Carlini et al. [Ca18]. This is a necessary by-product of all ML, because an ML model is essentially a compact representation of the training data. This issue can be mitigated using differential privacy techniques [Ca18, Li20], e.g., by adding noise to the

local models before sending them to the curator. Of course, this will degrade the overall model performance; how to decide the trade off between model quality and data privacy depends on the participants, the use case, the data, etc. and is out of scope of this paper.

4 Integration

The application of blockchain, federated learning and differential privacy operates on two logical components: While some components are accessible via the Internet, others are only present in private networks. Figure 1 shows an exemplary architecture as to be implemented in the research project "KOSMoS". Note that these are only logical components, not physical or digital system parts.

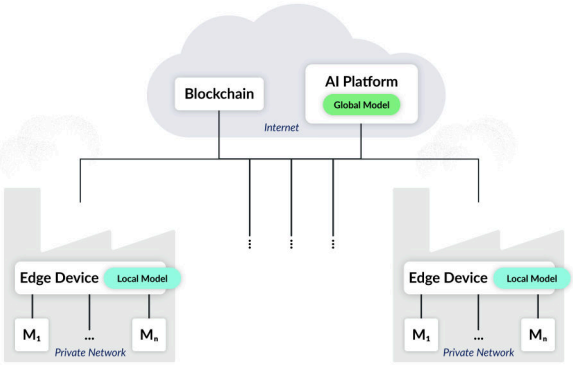


Fig. 1: Exemplary architecture with machines M_i in local networks connected to introduced methods.

The blockchain component of the system serves as a common, trusted distributed database of machine manufacturers and machine users. In the simplest case, the blockchain is centrally operated with only one node. However, there are usually several nodes, where each node contains a complete version of the blockchain. In our case, nodes can be located in the IT of a service provider, a machine manufacturer or even the machine user [BM20b].

In order to realize collaborative predictive maintenance models, the central curator is located in a cloud-based AI platform for prediction models. In the federated learning process, its task is to collect the model updates from all participants, aggregate the new model from them and then distribute the resulting model back to the participants. Each machine operator who wants to take part, needs one edge device locally on their hall floor, which is integrated into the private part of the architecture of KOSMoS. The edge device has access to all machine data required for model training and also serves as a communication interface between the hall floor with a set of $n \in \mathbb{N}$ machines M_i, where $i = 1, ..., n$, and the cloud components. Different data structures of the machines and sensors are translated into topic messages so that the models can be trained with a uniform structure.

5 Conclusion and Future Work

Data-driven business models that are used across company boundaries are important for SMEs, and the challenges discussed here are gaining in importance. The methods presented promise to ensure data integrity and data security at a high level. In detail, however, there are still many open questions that need to be addressed. The KOSMoS approach will be raised to a usable prototype level. In doing so, we will tackle many of the challenges mentioned here and solve them piece by piece. The result will be a reusable system complemented by a framework that will enable SMEs in particular to train collaborative prediction models with their partner companies. The platform will not only be applicable in production-related SMEs but also will be transferable to other industries and use cases.

Acknowledgment The contents of this publication are taken from the research project "KOSMoS - Collaborative Smart Contracting Platform for Digital Value Networks", funded by the Federal Ministry of Education and Research (BMBF) under reference number 02P17D026 and supervised by Projektträger Karlsruhe (PTKA). The responsibility for the content is with the authors.

Bibliography

[Ba18] Bagdasaryan, E.; Veit, A.; Hua, Y.; Estrin, D.; Shmatikov, V.: How To Backdoor Federated Learning. CoRR, abs/1807.00459, 2018.

[BM20a] Becker, C.; Mohr, M.: Federated Machine Learning: über Unternehmensgrenzen hinaus aus Produktionsdaten lernen. atp magazin, (5):18–20, 2020.

[BM20b] Bux, T.; Mohr, M.: Blockchain Lösungen für den produktionstechnischen Mittelstand. WT WERKSTATTSTECHNIK, BD. 111(Nr. 4):201–204, 2020.

[Ca18] Carlini, N.; Liu, C.; Erlingsson, Ú.; Kos, J.; Song, D.: The Secret Sharer: Evaluating and Testing Unintended Memorization in Neural Networks. ArXiv, abs/1802.08232, 2018.

[HNP09] Halevy, A.; Norvig, P.; Pereira, F.: The Unreasonable Effectiveness of Data. IEEE Intelligent Systems, 24:8–12, 2009.

[Hu20] Huang, H.; Lin, K.; Guo, S.; Zhou, P.; Zheng, Z.: Prophet: Proactive Candidate-Selection for Federated Learning by Predicting the Qualities of Training and Reporting Phases. arXiv:2002.00577 [cs, stat], February 2020.

[Ko19] Korb, T.; Michel, D.; Riedel, O.; Lechler, A.: Securing the Data Flow for Blockchain Technology in a Production Environment. IFAC-PapersOnLine, 52(10):125–130, January 2019.

[Li20] Li, T.; Kumar Sahu, A.; Talwalkar, A.; Smith, V.: Federated Learning: Challenges, Methods, and Future Directions. IEEE Signal Processing Magazine, 37:50–60, 2020.

[Ma18] Mao, W.; He, J.; Tang, J.; Li, Y.: Predicting remaining useful life of rolling bearings based on deep feature representation and long short-term memory neural network. Advances in Mechanical Engineering, 10(12), December 2018.

[Mc16] McMahan, H. B.; Moore, E.; Ramage, D.; Hampson, S.; Agüera y Arcas, B.: Federated Learning of Deep Networks using Model Averaging. CoRR, abs/1602.05629, 2016.

[Mo20] Mohr, M.; Wilhelm, F.; Hartwig, M.; Möller, R.; Keller, K.: New Approaches in Ordinal Pattern Representations for Multivariate Time Series. In: Proceedings of FLAIRS-33. 2020.

Vom Feinen ins Grobe

Verwendung einer Self Organizing Map und eines K – Means Clustering Algorithmus zur Clusterbildung der Masse von Schmierfettpulspunkten zur Bestimmung von Mittelwert und Standardabweichung

Stefan Paschek[1], Prof. Dr. Ing. hab. Catherina Burghart[2], Prof. Dr. Ing. Martin Kipfmüller[3]

Abstract: Schmierfette sind komplexe Werkstoffe die starken Parameterschwankungen unterliegen. Moderne Pulsventile ermöglichen das Applizieren von Schmierfetten mit Genauigkeiten im mg und µg Bereich. Die Bestimmung der Prozessfähigkeit erfordert hierbei je nach vorhandenem Messgerät ein manuelles Mehrfachauftragen von Schmierfettpulspunkten, um die Masse sicher zu erfassen und Rauschen zu unterdrücken. Um der Schmierfettmasse einen Erwartungswert und ein gültiges Toleranzintervall mit Standardabweichung zuzuweisen, wurde ein Algorithmus bestehend aus einer Self Organizing Map und eines K – Means Clustering Algorithmus entworfen. Dieser Algorithmus erlaubt sowohl die Bestimmung von Eingangsparametern zu einem definierten Massenerwartungswert, als auch die Vorhersage welcher Erwartungswert unter Verwendung von eingestellten Prozessparametern entstehen wird. Der Algorithmus wurde mit Trainingsdaten aus einem Versuchsstand trainiert und mit Testdaten evaluiert.

Keywords: Schmierfett; Self Organizing Map; K – Means Clustering; Standardabweichung; Pulsventil

1 Einleitung

Schmierfette sind hoch komplexe Werkstoffe die in vielen technischen Bereichen angewendet werden. Dabei ist der klassische Einsatz zur Verminderung von Reibung zweier Werkstoffe bei weitem nicht mehr nur der einzige Anwendungsbereich. Schmierstoffe werden auch häufig eingesetzt um die haptischen Eigenschaften von Drehknöpfen oder Schaltern zu beeinflussen. Ein gezielter Einsatz von Schmiermittel kann dabei den Drück – und Drehwiderstand verändern. Um das gewünschte Schmierverhalten zu erzielen ist eine präzise Dosierung der aufgetragenen Masse essentiell. Neben der Masse kann beim Auftragsprozess auch das entstehende Auftragsbild des Schmierfettes von Entscheidung sein.

[1] Hochschule Karlsruhe Technik und Wirtschaft, IMP, Moltkestraße 30, 76133 Karlsruhe, stefan.paschek@hs-karlsruhe.de
[2] Hochschule Karlsruhe Technik und Wirtschaft, IMP, Moltkestraße 30, 76133 Karlsruhe, catherina.burghart@hs-karlsruhe.de
[3] Hochschule Karlsruhe Technik und Wirtschaft, IMP, Moltkestraße 30, 76133 Karlsruhe, martin.kipfmueller@hs-karlsruhe.de

doi:10.18420/inf2020_40

Der Applikations – und Dosierprozess für Schmierstoffe ist jedoch schwer beherrschbar. Üblicherweise werden Schmierfette in Gebinden gelagert und über Förderpumpen zu ihrer Auftragsstelle gefördert. Das Fließverhalten der Schmierstoffe ist unter anderem von der Temperatur des Schmierstoffes, sowie vom Druck mit dem der Schmierstoff gefördert wird abhängig. Der von der Fasspumpe erzeugte Förderdruck kann jedoch in Abhängigkeit von Geometrie und Länge der Förderschläuche sehr stark schwanken, was sich sowohl auf Auftragsbild als auch auf die Auftragsmasse auswirken kann. Diese Faktoren sind jedoch in technischen Anwendungen beherrschbar und können mit einmaliger Festlegung behoben werden. Problematischer jedoch ist, dass Schmierfette kompressibel sind. Daraus ergibt sich eine zeitliche Abhängigkeit des sich aufbauenden Druckes des Schmiermittels. Zusätzlich dazu kann z.B. das Schmiermittel bei zu langer Lagerung des Gebindes durch sein Eigengewicht eine Separierung von Fettbestandteilen und Grundöl bewirken. Dadurch kann das Schmiermittel im Gebinde seine Viskosität ändern. Bereiche mit mehr Öl werden niedrig viskoser, und Bereiche in denen sich das Öl separiert hat sind höher viskos. Dieser Separierungsprozess kann ebenfalls durch dauerhaftes anliegen eines zu großen Förderdruckes eintreten. Somit können die Einstellparameter die Schmierfetteigenschaften beeinflussen. Außerdem können Schmierfette große Parameterschwankungen selbst innerhalb einer Charge unterliegen. Die Fließeigenschaften von Schmiermittel werden über NLGI Klassen [R05] definiert. Die Einteilungen reichen dabei von der Klasse 000 – sehr flüssig bis zur Klasse 6 – extrem steif. Die Schmierfette in diesem Projekt befinden sich in den Klassen 1 bis 3 – sehr weich bis halbsteif. Da diese Einteilung sehr grob ist, können Hersteller ihre Rezepturen ändern ohne ihre NLGI Klasse zu verlassen. Im aktuellen Stand der Forschung können viele Quellen zur Prädiktion von Schmiermitteleigenschaften während des Produktlebenszyklus gefunden werden. Viele Literaturquellen untersuchen die Alterungserscheinungen von Schmiermitteln im Einsatz [A16] bzw. richten sich an die Vorhersage von Lebensdauer der Schmiermittel [H04], [KA01], [ZVL09]. Es konnte jedoch noch keine Quelle gefunden werden die sich mit der Vorhersage des Verhaltens der aufgebrachten Schmiermittelmasse durch Pulsventile mit Methoden der künstlichen Intelligenz beschäftigt. Dies ist der Ansatzpunkt dieses Papers in dem eine Kombination aus Selforganizing Map und K – Means Clustering Algorithmus verwendet wird um aufgebrachte Masse eines Schmierfettpunktes anhand von Prozessparametern zu bestimmen, einer Gewichtsklasse zuzuweisen und einen Erwartungsbereich anhand einer Standardabweichung festzulegen.

2 Problembeschreibung

Dieses Forschungsprojekt zielt daraufhin ab den Schmiermittelapplikationsprozess zum einen mit dem Projektpartner besser verständlich zu gestalten, eine Prädiktion des Schmiermittelauftragsbildes und eine Prädiktion der Auftragsmasse anhand der Einstellparameter umzusetzen. Zusätzlich dazu ist eine Regelung vorzusehen um Störgrößen die die Schmiermittelaufbringung beeinträchtigen entgegenzuwirken um möglichst gleichbleibende Auftragsbilder zu erhalten. Dieses Paper wird sich im speziellen auf die Prädiktion der

Masse sowie deren statistischer Schwankung befassen. Die Erkenntnisse aus diesem Paper werden im Weiteren in die Entwicklung eines neuen digitalen Pulsventils einfließen um die applizierte Masse als auch Einstellparameter vorhersagen zu können. Im speziellen wird der Schmiermittelaufbringungsprozess über Pulsventile des Typs MPP03Pro zur Analyse betrachtet. Mit einem Pulsventil können Schmierfettpunkte aufgepulst werden. Die im Versuch aufgetragene Masse bewegt sich dabei im Bereich von 2mg bis 15mg. Um Messunsicherheiten des Wägeprozesses zu vermeiden und möglichst viele Messpunkte zu erhalten wurden immer 15 Messpunkte mit denselben Parametern aufgetragen und der Mittelwert der Masse gebildet. Der Nachteil dieses Vorgehens ist dass es nur schwer bestimmbar ist in welchen Massenbereich die Messpunkte schwanken. Somit kann zwar ein Erwartungswert berechnet werden aber keine Aussage getroffen werden welche Schwankungsbreite vorliegt.

Um eine Aussage treffen zu können welche Massenbereiche zu erwarten sind wurde eine Kombination aus Self Organizing Maps (SOM) [K82] und K – Means Clustering [HW79] verwendet. Dabei wird die SOM verwendet um den hochdimensionalen Feature Raum in einen interpretierbaren 2d Raum herunter zu transformieren. Zusätzlich dazu erlaubt eine SOM im Gegensatz von anderen Dimensionsreduktionsverfahren wie Principal Component Analyse die gleichzeitige Klassifizierung, Dimensionsreduktion und Visualisierung. Das K – Means Clustering wird verwendet um automatisiert Cluster aus diesem Raum zu finden. K – Means zählt zu den partitionierenden Clusterverfahren, im Gegensatz zu hierarchischen oder dichtebasierenden Clusterverfahren muss hierbei die Anzahl der Cluster angegeben werden. Dies ist jedoch in dieser Anwendung ein Vorteil, da es dem Benutzer erlaubt wird die Anzahl an Gewichtsunterteilungen vorzugeben. Die Cluster können anschließend verwendet werden um eine Massenschwankungsbreite innerhalb des Clusters zu definieren. Im folgenden Abschnitt wird nun das Vorgehen und die genaue Funktionsweise des Ansatzes vorgestellt.

3 Materialien und Methoden

3.1 Prüfstandsaufbau

Um die notwendige Anzahl an Prüfpunkten zu erhalten wurde ein Prüfstand konzipiert. Die Abb. 1 zeigt dabei das Konzept des Prüfstandes. Das Schmierfett ist in einem Gebinde gelagert. Das Gebinde wird mit einer pneumatischen Kleingerätepumpe mit 50bar zu einem Materialdruckregler weitergeleitet. Dieser kann den Druck des Materials auf einen Bereich von 10 – 25bar regeln. Vom Materialdruckregler wird das Schmierfett anschließend in das Pulsventil geleitet. Das Pulsventil kann über ein 5/2 Wegeventil pneumatisch angesteuert werden. Im Ventil wird die Düse von einer mechanisch vorgespannten Nadel verschlossen. Wird die Nadel pneumatisch angesteuert öffnet sie sich und Material kann aus der Düse austreten. Wird die Ansteuerung beendet schließt die Nadel sich automatisch durch eine vorgespannte Feder. Das Material wird dabei immer auf einem schwarzen Prüfkörper aufgetragen. Während des Applizierens wird der Druckverlauf über einen Drucksensor gemessen und abgespeichert. Über ein Heizelement kann das Material im Ventil noch zusätzlich

vorgewärmt werden. Die aufgetragene Masse wird über eine Präzisionswaage gemessen. Das Auftragsbild wird über eine Kamera festgestellt. Da das verwendete Schmierfett eine weiß – beige Farbe besitzt ergibt sich so ein guter Kontrast zum Prüfkörper. Ein UR10 6 – Achs Knickarm Roboter mit Vakuum Sauger dient zum Transport des Prüfkörpers.

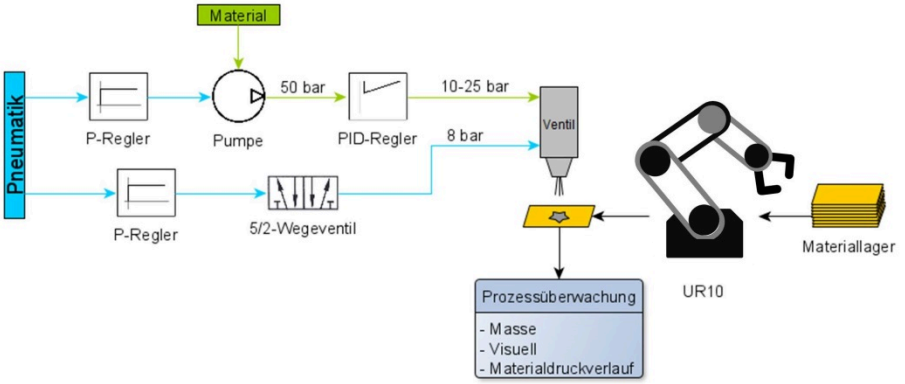


Abb. 1: Konzeptbild des Prüfstandsaufbaus

3.2 Schmierfettaufbringung

Für die Aufbringung des Schmiermittels wurde ein Ablauf definiert und programmiert. Der Ablauf sieht dabei folgendermaßen aus:

1. Datensätze für Messversuch werden geladen
2. Druckregler und Temperaturregler werden eingestellt
3. Roboter entnimmt Prüfkörper von Lager mittels Vakuumgreifer
4. Roboter platziert Prüfkörper auf Waage um Leermessung durchzuführen
5. Roboter positioniert Prüfkörper unter Ventil
   a. 15 Pulspunkte werden gesetzt
   b. Roboter bewegt Prüfkörper nach jedem Pulsvorgang weiter um Überlappung der Prüfpunkte zu vermeiden
   c. Druckverlauf wird für jedem Messpunkt während Pulsvorgang ermittelt
6. Roboter positioniert Prüfkörper über Waage um Vollmessung durchzuführen
   a. Leermessung wird von Vollmessung abgezogen um Masse zu erhalten
   b. Masse wird durch Anzahl an Prüfpunkten dividiert um Einzelmasse zu erhalten

7. Roboter positioniert Prüfkörper unter Kamera um Auftragsbild optisch zu erfassen
8. Roboter positioniert Prüfkörper an Abwurfstelle und wirft Prüfkörper ab => zurück zu Punkt 1

Bei jedem Durchgang wird somit die Masse der Prüfpunkte gemessen, der Kurvenverlauf des Druckabfalls während des Pulsvorgangs, sowie optisch das Auftragsbild. Die Abb. 2 zeigt dabei wie ein typischer runder Pulspunkt nach erfolgreichem Auftragen aussieht.

Abb. 2: Größenvergleich 2 Cent Stück (rechts) zu Pulspunkt (links) mit 2,38mm Durchmesser

3.3 SOM

Eine Self Organizing Map ist eine Spezialform der neuronalen Netze welche ohne Vorgabe von Zieldaten (unsupervised) trainiert werden kann und keine Aktivierungsfunktion besitzen. SOM werden dabei zur Dimensionsreduktion und zur Clusterbildung verwendet. Jedes Neuron der SOM ist mit allen Eingängen verbunden. Die Neuronen werden zu Beginn zufällig im Zustandsraum der Trainingsdaten verteilt. Anschließend wird die Distanz von den Neuronen zu einem ausgewählten Trainingsvektor berechnet. Das Neuron mit der geringsten Distanz ist das Siegerneuron und wird anschließend in Richtung des Trainingsvektors bewegt. Benachbarte Neuronen werden ebenfalls in Richtung des Trainingsvektors bewegt. Wird dieser Prozess während des Trainings mehrfach wiederholt kann ein Gleichgewichtszustand entstehen bei dem es zu keiner Änderung der Neuronenposition mehr kommt. Ist dies erreicht ist das Training beendet. Im Fall des Projektes wurde die vorgefertigte SOM Bibliothek von MathlabR2017 verwendet um einen 7-dimensionalen Feature Raum in eine SOM mit 20x20 Neuronen zu überführen. Die Features umfassen dabei Einstellparameter wie: eingestellter Solldruck des Materialdruckreglers, Ventilöffnungszeit, Rastereinstellung zum Vorspannen der Nadel in der Düse, eingestellte Temperatur des Ventils, zusätzlich gemessene Parameter: minimaler Druckwert während des Pulsens, maximaler Druck vor Pulsen und die Zielgröße: die applizierte Schmierfettmasse. Dabei hat jedes der

Neuronen 7 gewichtete Verbindungen zu den Eingängen. Als Verbindungsstruktur der Neuronen untereinander wurde eine hexagonale Wabenstruktur gewählt. Dies bedeutet, dass Neuronen auch schräg miteinander benachbart sein können. Das Training einer SOM liefert am Ende zwei Ergebnisse, zum einen ein fertiges Netz das genutzt werden kann um Eingangsdaten Neuronen zuzuweisen. Zum anderen liefert es die Gewichtsvektoren der einzelnen Neuronen. Diese Gewichtsvektoren können zu Heatmaps zusammengefügt werden und zur Visualisierung sowie zusätzlich dazu zur Clusterbildung verwendet werden. Mit dieser Technik ist es möglich die Zusammenhänge mehrerer Eingangsvariablen graphisch in nur wenigen Diagrammen darzustellen. Ein Beispiel wurde in [NH11] umgesetzt in dem ein hochdimensionaler Eingangsvektor bestehend aus unterschiedlichen Schadstoffen über Heatmaps der Gewichtungsvektoren visualisiert wurde.

3.4 KMeans Clustering

K – Means Clustering ist ein Clusterverfahren welches automatisch anhand eines Datensatzes Cluster in dem Datensatz ausfindig machen kann. Dadurch können ähnliche Datensätze gruppiert werden. Da der K - Means Algorithmus mit zufällig vordefinierten Clusterzentren startet kann das Ergebnis der Clusterung variieren [A12]. Der Algorithmus selbst ist dabei iterativ und ordnet die Datenpunkte den Clustern anhand ihrer quadrierten euklidischen Distanz zu den Clusterzentren zu. Anschließend werden die Clusterzentren anhand des Mittelwerts aller Datenpunkte innerhalb des Clusters verschoben. Ist dies erledigt beginnt ein neuer Iterationsschritt bis die Clusterung beendet wurde. Die aus der SOM berechneten Gewichtungsvektoren der Features können nun verwendet werden um Cluster zu bilden. Dabei ist es möglich sowohl nur einzelne Heatmaps wie z.B. die Masse oder aber beliebige gewichtete Kombinationen der unterschiedlichen Heatmaps zu verwenden. Das Ergebnis ist in allen Fällen ein Bereich dem ein Cluster zugeordnet ist. Innerhalb des Bereichs kann ein Mittelwert, sowie eine Standardabweichung berechnet werden. Zusätzlich dazu ist es auch möglich abzulesen welche anderen Features zu diesem Bereich gehören. Aus Sicht der Schmiermittelapplikation ist es also möglich, sowohl vorherzusagen wie sich Änderungen der Einstellparameter auf die Masse auswirken, als auch welche Parameter eingestellt werden müssen um die durchschnittliche Masse eines Clusters zu erreichen.

4 Praktische Durchführung

4.1 Datenakquisition

Die Abb. 3 zeigt das Kommunikationsschema des Prüfstandes. Der Controllino Microcontroller dient als zentrale Steuerstelle für Roboter und Ventil. Der Druckverlauf während des Pulsvorgangs wird vom Controllino zum Mess PC gesendet. Die Präzisionswaage zur Erfassung der applizierten Masse sendet die Daten über RS232. Die optische Erfassung der

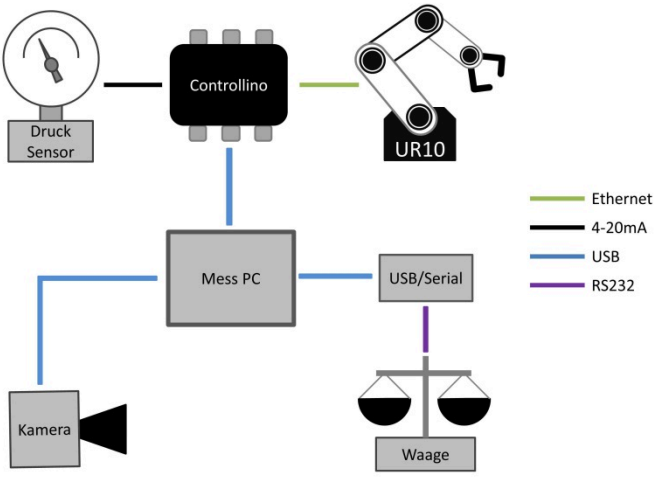


Abb. 3: Kommunikationsschema des Prüfstands

Auftragsbilder erfolgt mit einer Kamera, welche ein Bild der Draufsicht und der Seitenansicht erstellt.

Am Messprogramm des PC's werden die Einstellparameter für Druck, Ventilrasterung, Öffnungszeit und Temperatur sowie verwendetes Schmierfett abgespeichert. Die Druckkurve und die gemessene Masse der Waage werden ebenfalls abgespeichert. Der minimale und maximale Druckwert werden anschließend anhand der Druckkurve ermittelt.

4.2 Trainingsprozess

Der Ablauf des Trainingsprozesses wird in Abb. 4 dargestellt. Dabei wird zu Beginn eine Aufbereitung der Features durchgeführt. Dies inkludiert einen Plausibilitätscheck sowie eine Skalierung auf Mittelwert 0 und Standardabweichung 1. Anschließend werden die Features in einen Test und einen Trainingsdatensatz unterteilt, wobei 30% des originalen Datensatzes zufällig für den Testdatensatz ausgewählt werden. Die restlichen 70% werden als Trainingsdatensatz verwendet. Die Features des Trainingsdatensatzes werden zum Trainieren der SOM überführt. Diese erlernt dabei die optimale Verteilung der Neuronen im Feature Zustandsraum. Die von der SOM trainierten Gewichtungsheatmaps werden anschließend dem K – Means Algorithmus übergeben. Dabei werden ausgewählte Heatmaps gewichtet aufaddiert. Der K - Means Algorithmus unterteilt die resultierende Heatmap in eine vordefinierte Anzahl an Cluster. In einer Optimierungsiterationsschleife werden sowohl Hyperparameter der SOM (z.B. Anzahl an Neuronen) als auch Parameter von K - Means

(z.B. Distanzmetrik) angepasst. Nachdem die Hyperparameter optimiert wurden können die Algorithmen trainiert und evaluiert werden.

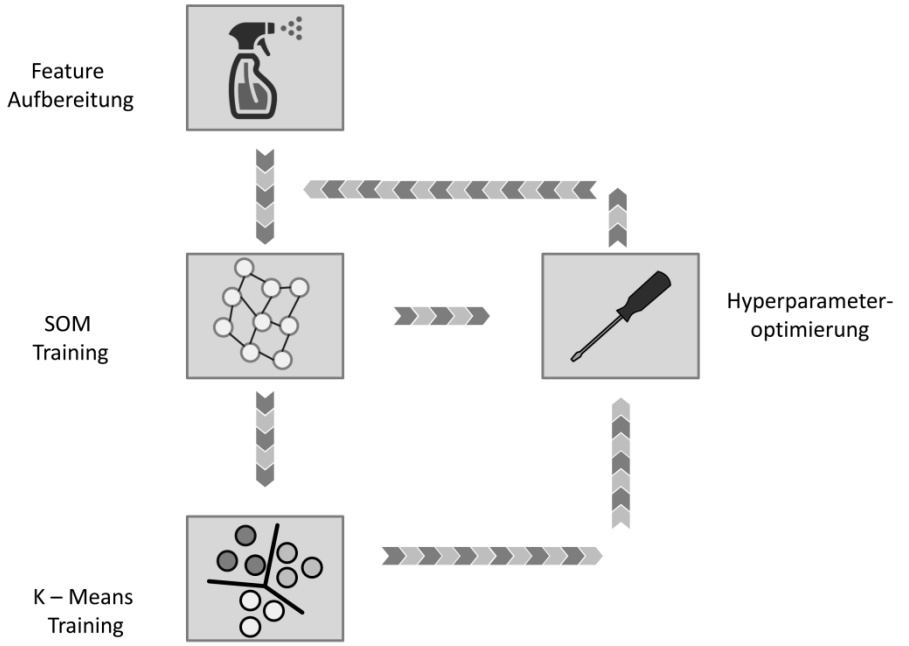


Abb. 4: Prinzipschaubild des Ablaufs des Trainingsprozesses

5 Ergebnisse

Die Abb. 5 zeigt das Ergebnis des Trainings der SOM. Dabei sind alle Features dargestellt. Die einzelnen Neuronen haben sich anhand ihrer Gewichtungsvektoren im Zustandsraum der Eingangsebene platziert. Diese Position kann direkt anhand der Gewichtungsvektoren herausgelesen werden. Die daraus resultierende Reduktion des Eingangsraumes liefert bereits eine Möglichkeit um optische Erkenntnisse und Zusammenhänge von Features zu identifizieren. Zum Beispiel ist erkennbar, dass die Features minP (Minimaldruck während Pulsens), maxP (Startdruck vor Pulsvorgang), Feature Preg (Sollwert des Druckreglers) zusammenhängen. Daraus kann gefolgert werden, dass der eingestellte Druck am Druckregler diese direkt beeinflusst. Im Vergleich dazu kann die Masse nicht direkt mit nur einem einzigen Parameter in Verbindung gebracht werden, sie hängt von einer Vielzahl an Parametern ab.

Wird nun dieses Ergebnis weiterverwendet und in den K – Means Algorithmus überführt so kann die Clusterung der SOM Heatmaps erfolgen. Das Ergebnis dazu wird in der Abb. 6 dargetstellt. Dabei ist einer der gefundenen Cluster in allen Features eingezeichnet.

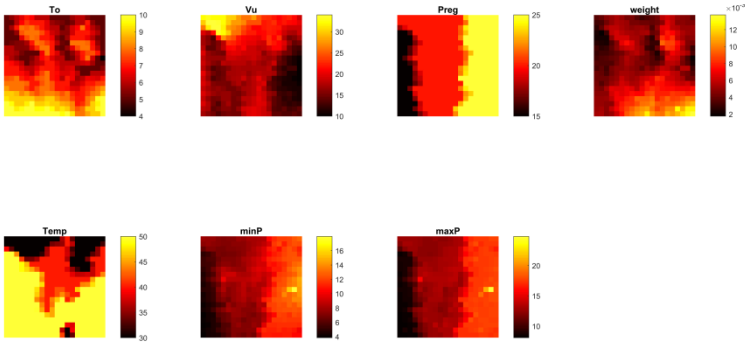


Abb. 5: Darstellung aller Gewichtungsfaktoren aller Neuronen für alle Features. Ein Neuron hat dabei einen Gewichtungsvektor der aus seiner Position in jeder einzelnen Heatmap besteht. Dabei entspricht jedes Rechteck einem Neuron der SOM. To . . . Ventilöffnungszeit, Vu. . . Rastereinstellung zum Vorspannen der Nadel der Düse, Preg. . . Sollwert des Materialdruckreglers, weight . . . gemessene Masse des applizierten Schmierfettpunktes, Temp. . . eingestellte Temperatur, minP. . . minimaler Druck während des Pulsvorgangs, maxP . . . Startdruck beim Pulsen

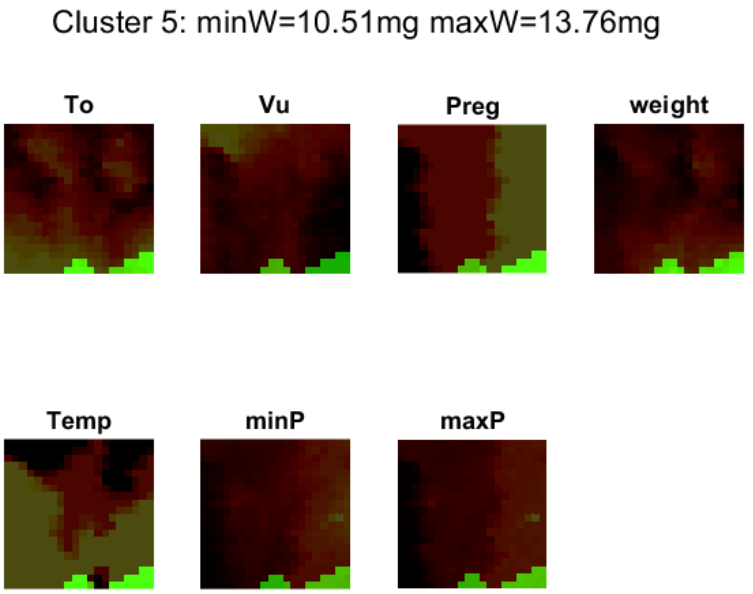


Abb. 6: Darstellung des ersten Clusters des K- Mean Algorithmus. Zur Clusterbildung wurde dabei nur die Masse des Schmierfettes gewählt. Der grüne Bereich kennzeichnet den Bereich des Clusters und wurde auf alle Features übertragen.

Für jeden Cluster kann nun Mittelwert und Standardabweichung berechnet werden. Dies liefert die notwendigen Informationen, um die Auswirkungen einer Systemparameterauswahl auf die gewünschte Qualität des Pulspunktes abzuschätzen. Die Tab. 1 zeigt die resultierenden Massenwerte für jeden Cluster auf. Die Cluster Nummer wurde dabei vom Ergebnis des K – Means Algorithmus entnommen. Diese Tabelle wurde direkt mit der trainierten Heatmap der SOM erstellt. Sie stellt einen Optimalfall dar, berücksichtigt aber noch nicht die Schwankungen des tatsächlichen Trainingsdatensatzes.

Tab. 1: Auflistung des Minimalgewichtes, des Maximalgewichtes, des Durchschnittsgewichtes und der Standardabweichung in jedem Cluster unter Verwendung der entstandenen Heatmap zur Klassifikation, sortiert nach steigender Masse. Die Masse ist dabei in mg angegeben.

minW/mg	maxW/mg	avW/mg	stdW/mg	ClusterNr
1,762	4,196	3,394	0,615	3
4,226	5,859	5,030	0,447	2
5,898	7,789	6,714	0,552	4
7,839	10,226	8,938	0,710	1
10,506	13,757	11,637	1,025	5

Um die Schwankungen des Trainingsdatensatzes ebenfalls zu berücksichtigen ist es nun notwendig den gesamten Trainingsdatensatz mit dem Algorithmus zu prüfen. Dies stellt auch den tatsächlichen Feldeinsatz dar. Der Trainingsdatensatz wird dazu in die SOM überführt. Diese führt ihre interne Klassifizierung durch und weist jedem Eintrag des Trainingsdatenansatzes einem Neuron zu. Da die Cluster des K – Means Algorithmus bereits in der Trainingsphase erstellt wurden und es möglich ist jedem SOM Neuron einen K – Means Cluster zuzuweisen, kann somit bestimmt werden welcher K – Means Cluster für den jeweiligen Datensatz zutrifft. Tab. 2 zeigt das Ergebnis aufgelistet. Dabei ist vor allem der neue Mittelwert für die Cluster von Interesse. Cluster 3 zeigt die größte Abweichung mit 3,62% zu dem Heatmap Cluster aus Tab. 1.

Tab. 2: Auflistung des Minimalgewichtes, des Maximalgewichtes, des Durchschnittsgewichtes und der Standardabweichung in jedem Cluster für Werte des Trainingsdatensatzes, sortiert nach steigender Masse. Die Masse ist dabei in mg angegeben

minW/mg	maxW/mg	avW/mg	stdW/mg	Cluster
1,584	4,607	3,271	0,688	3
4,042	6,216	4,948	0,466	2
5,598	7,887	6,764	0,577	4
7,907	10,607	8,970	0,724	1
10,300	14,231	11,837	1,186	5

Um das Ergebnis auf seine Generalisierbarkeit zu überprüfen, wird der Algorithmus mit dem Testdatensatz getestet. Damit wird überprüft wie gut der Algorithmus mit Daten arbeiten kann, die er zuvor noch nie während des Trainings verwendet hat. Das Ergebnis kann in Tab. 3 eingesehen werden. Das Resultat ist dabei sehr ähnlich zum Trainingsdatensatz. Der Cluster 3 hat eine um 4,24% größere Abweichung als in der Evaluierung mit Heatmap aus Tab. 1. Die Zunahme des Fehlers im Testdatensatz ist bei Machine Learning Algorithmen

normal. Der prozentuale Fehler ist im akzeptablen Bereich wodurch gezeigt werden konnte, dass der Algorithmus für diese Aufgabenstellung funktionsfähig ist.

Tab. 3: Auflistung des Minimalgewichtes, des Maximalgewichtes, des Durchschnittsgewichtes und der Standardabweichung in jedem Cluster für Werte des Testdatensatzes, sortiert nach steigender Masse. Die Masse ist dabei in mg angegeben

minW/mg	maxW/mg	avW/mg	stdW/mg	Cluster
1,584	4,607	3,250	0,702	3
3,907	6,104	4,975	0,470	2
5,647	7,887	6,754	0,558	4
7,907	10,267	8,923	0,735	1
10,300	14,231	11,941	1,196	5

Zusätzlich zu den Vergleichstabellen wurde das Bestimmtheitsmaß (R^2) [Y08] berechnet, um den Algorithmus bewerten zu können. Das Bestimmtheitsmaß gibt dabei Aufschluss darüber wie viel Varianz durch das Regressionsmodell erklärt werden kann und entspricht dem Quadrat des Korrelationskoeffizienten. Dabei deutet ein R^2 Wert nahe 1 an, dass die Varianz der abhängigen Variable von der unabhängigen Variable vollständig erklärt wird. Ein R^2 Wert von 0 deutet an dass die Varianz der abhängigen Variable nicht von der unabhängigen Variable erklärt werden kann. Somit spiegelt das Modell den tatsächlichen Verlauf umso besser wieder je näher R^2 gegen 1 geht. Für den entwickelten Algorithmus mit 5 Clustern ergibt sich ein R^2 Wert von 0.925.

Aus dem entwickelnden Algorithmus kann nun ein Workflow abgeleitet werden. Das Vorgehen zur Prädiktion des Clusters anhand von Eingabedaten wird in Abb. 7 dargestellt. Dabei kann anhand von Eingangsdaten unter Zuhilfenahme der SOM ein Gewinnerneuron ermittelt werden. Der bereits trainierte K – Means Algorithmus besitzt einen Cluster Index Vektor der jedem Neuron der SOM einen Cluster zuweist. Somit kann das Gewinnerneuron als Index des Clustervektors verwendet werden, um festzustellen welcher K – Means Cluster für den Datensatz zutrifft. Diese Information kann anschließend dem originalen Datensatz angehängt werden. Daraus ist es möglich abzuleiten welcher Erwartungswert und welche Standardabweichung zutreffen.

Der Algorithmus erlaubt es auch die umgekehrte Richtung abzuarbeiten. Dabei kann von einem ausgewählten Cluster einer vorgegebenen Pulspunktmasse auf die dafür notwendigen und möglichen Eingangsparameter geschlossen werden. Der Vorgang ist in Abb. 8 dargestellt. Zu Beginn wird ein Cluster anhand von der vorgegebenen Masse ausgewählt. Alle Neuronen der SOM die innerhalb des Clusters liegen werden anschließend ausgewählt. Aus den Neuronen können nun die dafür bereits trainierten Eingangsgewichtungen extrahiert werden. Diese Gewichtungen entsprechen den von der SOM gefundenen Positionen innerhalb des ursprünglichen Eingangszustandsraumes. Deshalb können die Gewichtungsvektoren der Neuronen direkt als Eingangsgrößen für die Prozesssteuerung verwendet werden.

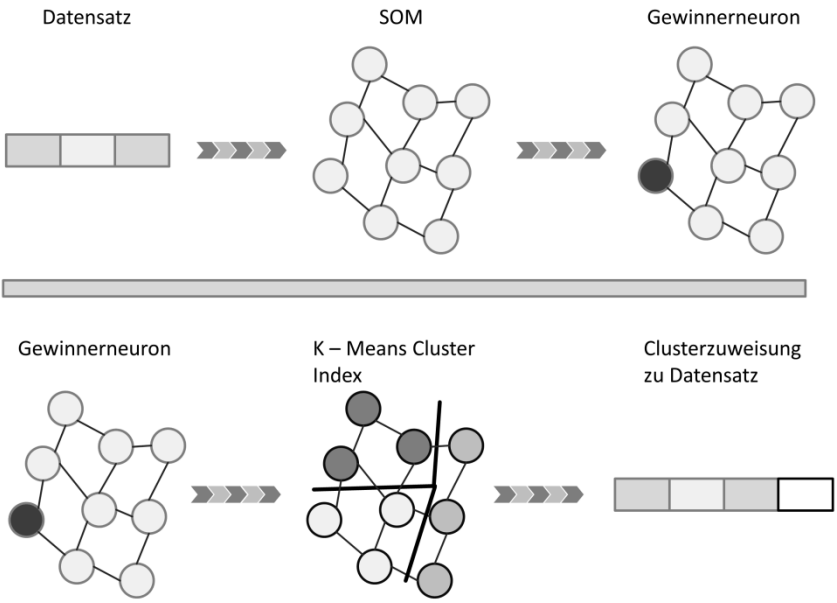


Abb. 7: Workflow des entwickelten Algorithmus zur Clusterprädiktion

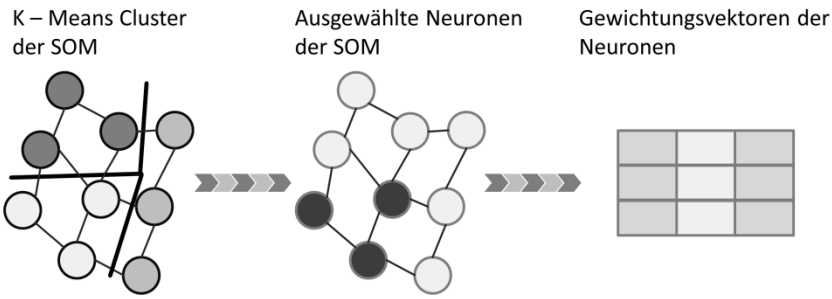


Abb. 8: Workflow um von einem ausgewählten Cluster zu den notwendigen Einstellparametern zu gelangen

6 Zusammenfassung und Ausblick

Eine SOM und ein K – Means Algorithmus wurden trainiert, um die Masse von Schmierfettpulspunkten in Cluster zu unterteilen. Dies bietet die Möglichkeit unter Kenntnis der Eingabeparameter sowohl den Erwartungswert der Masse als auch die Streuung über die

Standardabweichung vorherzusagen. Es ist genauso umgekehrt möglich durch Auswahl eines Clusters zu bestimmen welche Eingabeparameter möglich sind, um einen Pulspunkt zu erstellen der innerhalb des ausgewählten Clusters liegt.

Der Algorithmus wurde erfolgreich umgesetzt und mit Testdaten validiert. Es konnte ein Bestimmtheitsmaß von 0.925 erreicht werden welches zeigt dass die entstandene Regression die Varianz der Daten ausreichend erklärt.

In den nächsten Schritten werden nun zusätzliche Messdaten mit dem Prüfstand aufgenommen. Dabei sollen Faktoren wie unterschiedliche Düsendurchmesser sowie Schmierfette mit anderen Fließverhalten mit untersucht werden. Der Algorithmus wird dabei erweitert und auf die neuen Testdaten übertragen. Zusätzlich wird der Algorithmus als Massenprädiktionsmodell dienen um eine anpassungsfähige Regelung zu entwickeln.

Literaturverzeichnis

[R05] Rudnick, L. R: Synthetics, Mineral Oils, and Bio-Based Lubricants: Chemistry and Technology (Chemical Industries), CRC Press Inc, 2005.

[K82] Kohonen, T: Self-Organized Formation of Topologically Correct Feature Maps, Biological Cybernetics, 43 (1), Seite 59–69, 1982

[HW79] Hartigan, M. A.; Wong, M.A.: Algorithm AS 136: A K-Means Clustering Algorithm. In: Journal of the Royal Statistical Society, Series C (Applied Statistics). 1. Auflage. Band 28, Seite 100–108, 1979

[Y08] Yadolah, D: The Concise Encyclopedia of Statistics: Springer Science + Business Media, LLC, Seite 88, 2008

[NH11] Neme, A.; Hernandez, L.: Advances in Self - Organizing maps, 8th International Worshop, WSIM 2011, Espoo, Finland, June2011, Proceedings: Visualizing Patterns in the Air Quality in MexicoCity with Self-Organizing Maps, Springer Science + Business Media, LLC, Seite S318 – 327, 2011

[A12] Armin, S.: Knowledge Discovery in Databases, Eine Analyse des Änderungsmanagements in der Produktentwicklung, Springer Science + Business Media LCC, Seite 70, 2012

[A16] Asghar R., On the Degradation of Lubricating Grease, LSU Doctoral Dissertations, Quelle: `https://digitalcommons.lsu.edu`, Zugriff: 04.05.2020, 2016

[H04] Huiskamp, B: Grease life in lubricated-for-life deep groove ball bearings. Evolution, 2:26–28, 2004.

[KA01] Van den Kommer, A;. Ameye J.: Prediction of remaining grease life- a new approach and method by linear sweep voltammetry. Proceedings Esslingen Conference, Seite 891–896, 2001.

[ZVL09] Van Zoelen, M.T.; Venner, C.H; Lugt, P.M.: Prediction of film thickness decay in starved elasto-hydrodynamically lubricated contacts using a thin-film layer model. Proceedings of the Institution of Mechanical Engineers. Part J, Journal of engineering tribology, Seite 541-552, 2009.

Hochschule 2030

Hochschule 2030 - Welche Entwicklungen der IT prägen die Zukunft der Hochschulen?

Vorwort zum Workshop

Gunnar Auth[1], Markus von der Heyde[2], Ulrike Lucke[3]

Hochschule ist wie jeder andere Bereich des öffentlichen Lebens ohne Digitalität nicht mehr vorstellbar. Covid19 hat dieser schon länger andauernden Entwicklung einen neuen Schub verliehen. Der Rückblick auf diese Entwicklung erlaubt uns auch eine Einschätzung, welche zukünftigen Technologien und Organisationsformen die Kernprozesse unserer Hochschulen verändern können.

Zum einen werden Technologien, auch durch wirtschaftliche Interessen getrieben, in immer engeren Innovationszyklen und kürzeren Laufzeiten mit Hochdruck zum Einsatz gebracht. Zum anderen werden eigene Forschungsleistungen in verschiedenen Bereichen einer Hochschule (bspw. Institute, Verwaltungsdezernate, IT-Service-Einrichtungen) umgesetzt und manifestieren damit die Dualität von Forschungsleistung und nutzbaren Services. Hinzu kommen neuartige Managementkonzepte und Organisationsmodelle zur Bewältigung dieser Veränderungen.

In diesem Spannungsfeld des Innovationsmanagements untersucht diese Workshop-Reihe zur Zukunft der IT an Hochschulen aus unterschiedlichen Blickwinkeln, welche aktuellen Phänomene und Entwicklungen maßgeblich zur digitalen Transformation der Hochschulen beitragen werden. Durch das Verstehen dieser Entwicklungen eröffnet der Workshop Gestaltungsoptionen für die Hochschule der Zukunft. Mittels konstruktiver Forschungsdesigns erarbeitete Modelle, Methoden oder Prototypen werden aus theoretischer und praktischer Sicht diskutiert und weiterentwickelt. Wir adressieren dabei sowohl Forschung und Studium & Lehre als primäre Geschäftsfelder der Hochschulen als auch unterstützende Verwaltungsprozesse. Dabei stellt der Workshop „Hochschule 2030“ diese Themen zur Diskussion:

[1] Hochschule Meißen (FH) und Fortbildungszentrum, Herbert-Böhme-Str. 11, 01665 Meißen, gunnar.auth@hsf.sachsen.de
[2] vdH-IT, Rainer-Maria-Rilke-Str. 10, 99425 Weimar, info@vdh-it.de
[3] Universität Potsdam, Institut für Informatik & Computational Science, A.-Bebel-St. 89, 14482 Potsdam, ulrike.lucke@uni-potsdam.de

doi:10.18420/inf2020_41

- Durch die derzeitige Pandemie-Situation werden vermehrt digitale Formate für Lehr- und Lernorte gewählt. Traditionelle Präsenzveranstaltungen müssen neu geplant werden und stellen Herausforderungen an verschiedene Phasen im Bildungsprozess. Darüber hinaus müssen weiterhin internationale Kurse angeboten werden können um die globale Lehr-/Lern-Community im akademischen Austausch zu stärken. Dafür werden neue Plattformen genutzt und bekannte Lernumgebungen neu strukturiert um die von Bologna geforderte Mobilität der Studierenden zu gewährleisten.
- Aus der technischen Perspektive präsentieren Globalisierung, Mobilität und kontextbewusste Endgeräte neue Möglichkeiten für die akademische Lehre. Dafür muss eine adäquate IT-Infrastruktur erstellt werden, die nutzergenerierte Inhalte erreichbar macht und eine verlässliche Kommunikation zwischen Individuen und Communitys sicherstellt. Ein Schwerpunkt dieses Workshops ist es daher den Einsatz verschiedener Medienformate und Wege der Erreichbarkeit zu diskutieren.
- Da Hochschulen diverse Aufgaben zu bewältigen haben, ist das Datenmanagement in diesem Workshop auch von besonderem Interesse. Durch Verwaltung, Planung, Abrechnung, Publikationen und weitere Gebiete, kommt es zur Redundanz in digitalen Systemen, die zu mangelhafter Datenqualität und Mehrarbeit führen. Um diese Probleme zu diskutieren werden hier Wege zur Darstellung der Daten dargestellt und im Besonderen ein Augenmerk auf offene Bildungsressourcen gelegt.
- Neben den Orten für Lehre und Lernen stellt die derzeitige Situation auch besondere Herausforderungen an die Verwaltung und Organisation an Hochschulen. Insbesondere das Campus-Management und die Organisation von Präsenzstudiengängen müssen in dieser Zeit überdacht werden. Innovative Ansätze, aktuelle Entwicklungen und Ergebnisse in Forschung und Lehre sind Schwerpunkt dieses Workshops

Die im Workshop präsentierten aktuellen Entwicklungen, Herausforderungen und Trends zu den Bereichen der Digitalisierung der Hochschule Organisation, Lehr- und Lernorte, Datenmanagement und Unterstützung der Lehre umfassen in diesem Jahr zehn Beiträge, die vom Programmkomitee begutachtet und ausgewählt wurden:

- Auf den Schwerpunkt Organisation und Verwaltung beziehen sich „COBIT 2019 Application in Higher Education in Bavaria“ sowie „Start des neuen weiterbildenden Masters ‚Digitales Datenmanagement – DDM‘ während des Corona Lockdowns“.
- Der Punkt Lehr- und Lernorte wird beleuchtet von „Persönliche Lernumgebungen von Studierenden im Corona-Semester“, „Community-basierte Methode zur transdisziplinären Gestaltung von Lernräumen an Hochschulen“ und „Das Stud.IP ePortfolio Plugin als digitaler Lehr-Lern- und Prüfungsort“.
- Das Thema Datenmanagement wird besprochen in „Entwicklung und Implementierung eines Plug-Ins und von APIs für offene Bildungsressourcen“, in „Die Sprache «SemaLogic» als semantische Repräsentation - Eine anforderungsbasierte Sprache zur

Modellierung von Prüfungsordnungen und Abbildung von Studienverläufen“ und in „Fachanwendung für digitale Modulkataloge - Eine Untersuchung zu Graph-basierter Daten-Modellierung und Navigation“.

- Näheres zum Thema Unterstützung in der Lehre erfahren Sie in „Videoproduktion: Entwicklung eines adaptiven Wegweisers für Hochschullehrende“ sowie in „Voneinander lernen – miteinander gestalten. Hochschulübergreifende Netzwerke für die Digitalisierung der Lehre“.

Die wissenschaftlichen Beiträge wurden während des Workshops durch Diskussionen abgerundet und bereichert. Die thematische Spannbreite des Workshops spiegelte sich in der erfreulich hohen Diversität der Teilnehmenden wider. Dadurch bekamen die vorgetragenen Perspektiven eine vielfältige Einordnung in den jeweiligen fachlichen Kontext der Kommentare und Rückfragen. Allen Teilnehmenden sei für die Disziplin in der digitalen Umsetzung des Workshops herzlich gedankt.

Wir danken allen Einreichern für die sorgfältige Aufbereitung ihrer Arbeitsergebnisse sowie den Mitgliedern des Programmkomitees für die Mitwirkung bei der Begutachtung und Auswahl der Beiträge. Den Organisatoren der GI-Jahrestagung danken wir für die Unterstützung bei der Ausrichtung der Veranstaltung. Nicht zuletzt sei Axel Wiepke unser herzlicher Dank ausgesprochen für die große Unterstützung bei der Organisation des Workshops.

G. Auth, M. von der Heyde und U. Lucke

Karlsruhe, im September 2020

Programmkommitee:

Arndt Bode (TU München), Michael Brinkwerth (TU Clausthal), Jan Eden (Uni Köln), Christian Erfurth (EAH Jena), Torsten Eymann (Uni Bayreuth), Marc Göcks (Multimedia Kontor Hamburg), Yvonne Groening (myconsult), Andreas Hartmann (HfT Leipzig), Odej Kao (TU Berlin), Frank Klapper (Uni Bielefeld), Andreas Knaden (Uni Osnabrück), Michael Koch (UniBW München), Harald Kosch (Uni Passau), Susanne Leist (Uni Regensburg), Sören Lorenz (GEOMAR Kiel), Vera Meister (TH Brandenburg), Heike Neuroth (FH Potsdam), Gudrun Oevel (Uni Paderborn), Hans Pongratz (TU München), Simone Rehm (Uni Stuttgart), Sabine Roller (Uni Siegen), Guido Wirtz (Uni Bamberg), Ramin Yahyapour (Uni Göttingen), Markus Zahn (Uni Augsburg)

Entwicklung und Implementierung eines Plug-Ins und von APIs für offene Bildungsressourcen (OER)

Entwicklungen der Initiative „Open Education Austria Advanced“ für die Verknüpfung von LMS und OER-Repositorium einer Universität sowie die Metadatenweitergabe an das österreichweite OER-Fachportal

Christoph Ladurner [1], Christian Ortner [2], Karin Lach [3], Martin Ebner [4], Maria Haas [5], Markus Ebner [6], Raman Ganguly [7], Sandra Schön [8]

Abstract: Um einen breiten Zugang zur Bildung und großzügige Nutzung von Bildungsressourcen zu ermöglichen, setzt auch die Technische Universität Graz (TU Graz) auf offene Bildungsressourcen (Open Educational Resources, kurz OER). Der Beitrag beschreibt die technologischen Entwicklungen und Prozesse, damit Lehrende der TU Graz das eigene Lernmanagementsystem für die Veröffentlichung von OER nutzen können. Es wird im Beitrag nachgezeichnet wie Schnittstellen und Prozesse gestaltet wurden, um Lern- und Lehrressourcen der TU Graz mit entsprechenden Metadaten auszuzeichnen, um sie über das universitätseigene OER-Repositorium und entsprechenden Schnittstellen für das OER-Fachportal der Universität Wien einer breiten Öffentlichkeit recherchierbar anzubieten. Nur entsprechend qualifizierte Lehrende der TU Graz erhalten die Berechtigung für die Nutzung des neuen OER-Plug-In. Der Beitrag schließt mit Empfehlungen für Nachahmer/innen.

Keywords: Open Educational Resources (OER); Metadaten; LOM; Repositorium; Referatorium; Lernmanagementsystem; Qualifizierung; Hochschulen; Zertifizierung; Schnittstelle; Plug-In

[1] Technische Universität Graz, Bibliothek und Archiv, Technikerstr. 4, 8010 Graz, Österreich, christoph.ladurner@tugraz.at, https://orcid.org/0000-0003-3653-7558
[2] Technische Universität Graz, Lern- und Lehrtechnologie, Münzgrabenstraße 36, 8010 Graz, Österreich, christian.ortner@tugraz.at, https://orcid.org/0000-0002-6728-7574
[3] Universität Wien, Bibliotheks- und Archivwesen, Universitätsring 1, 1010 Wien, Österreich, karin.lach@univie.ac.at
[4] Technische Universität Graz, Lern- und Lehrtechnologie, Münzgrabenstraße 36, 8010 Graz, Österreich, martin.ebner@tugraz.at, https://orcid.org/0000-0001-5789-5296
[5] Technische Universität Graz, Lern- und Lehrtechnologie, Münzgrabenstraße 36, 8010 Graz, Österreich, maria.haas@tugraz.at, https://orcid.org/0000-0002-3028-0803
[6] Technische Universität Graz, Lern- und Lehrtechnologie, Münzgrabenstraße 36, 8010 Graz, Österreich, markus.ebner@tugraz.at, https://orcid.org/0000-0002-5445-1590
[7] Universität Wien, Zentraler Informatikdienst, Universitätsstr. 7, 1010 Wien, Österreich, raman.ganguly@univie.ac.at, https://orcid.org/0000-0002-9837-0047
[8] Technische Universität Graz, Lern- und Lehrtechnologie, Münzgrabenstraße 36, 8010 Graz, Österreich, sandra.schoen@tugraz.at, https://orcid.org/0000-0003-0267-5215

doi:10.18420/inf2020_42

1 Einleitung: OER an Hochschulen

Um einen breiten Zugang zur Bildung und großzügige Nutzung von Bildungsressourcen zu ermöglichen, setzen zahlreiche weltweite Organisationen und Agenturen auf sog. „Open Educational Resources“ (kurz OER). Als „offene Bildungsressourcen“ bezeichnet die UNESCO, festgehalten in der Pariser Erklärung der „Weltkonferenz zu OER“ im Jahr 2012: „[OER sind] Lehr-, Lern- und Forschungsressourcen in Form jeden Mediums, digital oder anderweitig, die gemeinfrei sind oder unter einer offenen Lizenz veröffentlicht wurden, welche den kostenlosen Zugang sowie die kostenlose Nutzung, Bearbeitung und Weiterverbreitung durch Andere ohne oder mit geringfügigen Einschränkungen erlaubt.” [UN12, Bu13]. Auch die Europäische Kommission fördert OER und will damit „Bildung öffnen“ und die Vermittlung digitaler Kompetenzen an Schulen und Hochschulen verbessern [EC13].

Damit Bildungsressourcen von Dritten in rechtlich einwandfreier Weise modifiziert und genutzt werden können, müssen sie mit einer sog. „offenen Lizenz“ veröffentlicht worden sein. Im entsprechenden Lizenztext werden die genannten Nutzungsmöglichkeiten zugestanden und ggf. Bedingungen dazu genannt. Ohne eine solche offene Lizenz können Ressourcen von Dritten im Hochschulkontext nur sehr eingeschränkt genutzt werden, die Fair-Use-Regel der USA erlaubt dort weitaus mehr Spielraum [ESK16].

In Bezug auf den Kontext der Universitäten, gibt es weitere Spezifika für OER im deutschsprachigen Europa im Vergleich mit anderen Ländern [ESK16, Mr13]: Zum einen ist der Besuch öffentlicher Hochschulen nur mit geringen Kosten verbunden, OER ist also kein potentielles Marketingmittel für zukünftige Studierende. Zum anderen ist die wissenschaftliche Freiheit an Universitäten ein hohes Gut, so dass zumindest für Lehrende an Universitäten kaum Vorgaben möglich sind, dass sie z. B. Lehrmaterialien als OER veröffentlichen müssen. Vor diesem Hintergrund erklärt sich zum einen, dass es nicht einfach ist, an europäischen Hochschulen OER-Strategien einzuführen und Prozesse zu implementieren, die die Erstellung und Veröffentlichung von OER aktiv unterstützen.

Die wichtigste Initiative zu OER im Bereich der Hochschulen in Österreich war bislang das Projekt „Open Education Austria”, ein Projekt mit Ko-Finanzierung des Bundesministeriums für Bildung, Wissenschaft und Forschung. Das Vorhaben wurde im März 2020 unter dem Titel „Open Education Austria Advanced“ mit weiteren Partnern fortgesetzt und wird die OER-Infrastruktur für österreichische Hochschulen gezielt erweitern. Ein Teilvorhaben ist dabei die Service- und Technologie-Infrastrukturen rund um Veröffentlichung von OER innerhalb der Partner-Universitäten auszubauen und die Erfahrungen und Lösungen damit mit anderen zu teilen (s. Abb. 1). Zu den Einzelvorhaben gehört dabei die Entwicklung eines Prototypens für eine Anwendung, welche die automatische Überführung von OER aus dem Lernmanagementsystems (LMS) in ein Bibliothekssystem ermöglicht. Die Entwicklung des hier vorgestellten Prototyps eines Plug-Ins sowie von Schnittstellen wurde für das LMS der Technischen Universität Graz (TU Graz) entwickelt. Die Metadaten-Auswahl

und Schnittstellen-Entwicklung erfolgte dabei mit Unterstützung von MitarbeiterInnen der Universität Wien, die für die Entwicklung des OER-Fachportals verantwortlich sind.

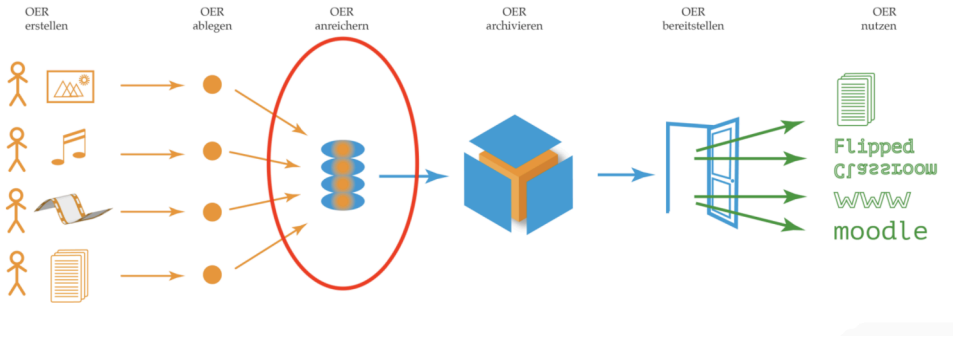


Abb. 1: Schematische Darstellung der notwendigen technischen Infrastruktur zu Verfügbarmachung von OER im Rahmen des Projekts Open Educational Austria Advanced. Quelle: OEAA, `https://www.openeducation.at/das-projekt/ziele/` (2020-05-19)

In diesem Beitrag wird die Entwicklung eines OER-Plug-In und von Schnittstellen beschrieben, die es den einschlägig qualifizierten Lehrenden der TU Graz ermöglicht, Lernressourcen im Lernmanagementsystem mit den entsprechenden OER-Metadaten auszuzeichnen, sie ins universitätseigene Repositorium zu übertragen und über Schnittstellen für Recherchen am OER-Fachportal zentral recherchierbar zu machen.

Bevor wir die genaueren Methoden und Ergebnisse unserer Entwicklung vorstellen, möchten wir einen Einblick in die Hintergründe und den aktuellen Stand in die Debatte um Bildungsressourcen und passende Metadaten geben.

2 Metadatenstandards für offene Bildungsressourcen an Hochschulen

Um offene Bildungsressourcen für andere nutzbar zu machen, müssen sie auch auffindbar und recherchierbar sein. Die Verwendung und Einbettung einer entsprechenden offenen Lizenzierung in den Quelltext ist dabei nur ein erster Schritt. Grundlage für die angebotsübergreifende Recherchierbarkeit von Ressourcen sind einheitliche Beschreibungen der Materialien, also Standards für die Metadaten zu den Ressourcen. Es gibt unterschiedliche Ansätze und Vorschläge für Systematiken von Metadaten von (freien) Bildungsmaterialien bzw. Lernobjekten [Po14]. Die Herausforderung liegt darin, dass OER bzw. auch allgemein Lern- und Lehrmaterialien sehr variantenreich sind [Eb15] Einen Überblick über (englischsprachige) Standards von Metadaten von Bildungsressourcen geben [BC10]. [ZDN13] haben eine entsprechende Übersicht über nutzbare Standards für Metadaten für OER zusammengestellt. In Deutschland arbeitet die OER-Metadaten-Gruppe daran, „eine Harmonisierung der OER-Metadaten im deutschsprachigen Raum zu erreichen und hierzu eine Empfehlung zu erarbeiten“ [OER15]. Bei digitalen OER wird in vielen Projekten auf den Standard

„Learning Objects Metadata" (kurz LOM) [Re13] zurückgegriffen. LOM ist ein offener Standard, der von der Organisation IEEE (Institute of Electrical and Electronics Engineers) entwickelt und veröffentlicht wurde [Wi19]. LOM ist in verschiedene Kategorien gegliedert, die Teilaspekte der Metadaten abdecken. In anderen Veröffentlichungen zu Metadaten und OER wird auch der alternative Ansatz des Learning Registry Metadata Initiative (LRMI) favorisiert [St17]. Im universitären Raum, ist LOM aber weiterhin dominierend, so wurde 2020 wurde „LOM for Higher Education OER Repositories", also eine „Beschreibung zur XML Schema Definition des Metadatenprofils für Open Educational Resources im Hochschulbereich" der OER-Metadatengruppe, einer Arbeitsgruppe deutschsprachiger Universitäten veröffentlicht [MP20, KIM20].

3 Vorgehen

Dieser Beitrag dokumentiert die Entwicklung und Implementierung der technischen Infrastruktur und der Prozesse, damit Lehrende der TU Graz ihre selbst erstellten Lern- und Lehrressourcen mit einer offenen Lizenz versehen können und diese OER dann aus dem Lernmanagementsystem in das Repositorium der TU Graz übertragen werden können um schließlich auf dem (geplanten) österreichweiten Referatorium für OER, einem OER-Fachportal der Universität Wien, recherchierbar und auffindbar zu sein. Neben technischen Lösungen sind dabei auch Qualifikationen und Berechtigungen auf Seite der Lehrenden notwendig, die in der Umsetzung berücksichtigt werden müssen.

In diesem Beitrag beschreiben wir die technischen Analysen und Entwicklungen der Auszeichnung der Lern- und Lehrressourcen der TU Graz als OER, also die Auswahl des entsprechenden Metadatenstandards und der genutzten Auszeichnungen, sowie die technischen Umsetzungen in Form eines Plug-In für das hochschuleigene Lernmanagementsystem und Entwicklungen von Application Programming Interfaces (API). Methodisch werden dabei Verfahren der technischen Analyse und Prototypenentwicklung in der Softwareentwicklung genutzt. Zudem beschreiben wir ergänzend auch den Qualifikations- und Zertifizierungsprozess, der Lehrende dazu qualifiziert, kompetent zu entscheiden, welche der von ihnen erstellten und genutzten Lern- und Lehrressourcen der Allgemeinheit als OER zur Verfügung gestellt werden sollen. Als Grundlage haben wir dazu auch interne Arbeitspapiere und Dokumentationen sowie eine Projektpräsentation genutzt [La19, EHO17, Ha18].

4 Ergebnisse

Im Folgenden beschreiben wir die einzelnen Entwicklungsschritte und ihre Ergebnisse bei der Entwicklung der technischen und sozialen Implementierung des Plug-Ins und der APIs in die technische Infrastruktur und Prozesse der OER-Publikation an der TU Graz.

4.1 Analyse der Ausgangssituation

Folgende Technologien und Prozesse waren Ende 2017, zu Beginn der Implementierung, an der TU Graz rund um die Erstellung von Lern- und Lehrressourcen etabliert:

„TeachCenter" heißt das Lernmanagementsystem (LMS) der TU Graz. Es basiert Ende 2017 auf der Open-Source-Software Moodle in der Version 3.1. Für den Betrieb an der TU Graz wurde die Software Moodle um einen Webservice für die Benutzer/innen- Synchronisation, und Synchronisation von Kurs An- und Abmeldungen erweitert. Zudem wurde eine eigene Oberfläche, welche der Corporate Identity der TU Graz entspricht, entwickelt. Auf Anfrage der Lehrenden werden diese Kurse erstellt und gewartet. TeachCenter umfasste 2017 etwa 1.200 Kurse, derzeit (Mai 2020) über 2.000 Kurse [Eb20]. Lehrende können in ihren Kursen eigene Materialien unterschiedlicher Formate einbinden und hochladen.

Ein Kurs ist mit einer oder mehreren Lehrveranstaltungen aus „TUGRAZ online", dem Campusmanagementsystem mit ´der entsprechenden Nutzer/innen-Verwaltung, verbunden. Bei TUGRAZ online handelt es sich um das zentrale Verwaltungssystem für Bedienstete und Studierende der TU Graz. Studierende können sich hier für ihre Lehrveranstaltungen anmelden und Lehrende können administrative Tätigkeiten (z. B. Prüfungsergebnisse) zur Verwaltung von Lehrveranstaltung durchführen.

Das Repositorium der TU Graz ist eine Eigenentwicklung, geschrieben in PHP. Damit die Ressourcen im Repositorium auch geordnet und recherchiert werden können, sind zusätzliche Informationen über die Materialien notwendig, also „Metadaten", die die Materialien beschreiben (s. Abschnitt 2). Das Repositorium der TU Graz hat dafür das Maschinelle Austauschformat für Bibliotheken (MAB) [DN19] implementiert. Es wird unter anderem in dem Bibliotheksprogramm Aleph als Datenbankformat für die Speicherung von bibliographischen Daten verwendet. Die Entwicklung des Formats wurde eingestellt und wird laufend durch MARC21 abgelöst [DN19].

In der Ausgangssituation 2017 fehlen also Schnittstellen, die zum einen die Materialien aus dem Lernmanagementsystem (LMS) der TU Graz in das Repositorium der Universitätsbibliothek der Technischen Universität Graz überführen und zum anderen die Weitergabe der Metadaten an das zentrale OER-Fachportal ermöglichen.

Eine Besonderheit ist das Videoportal TUbe der TU Graz, bei dem Lehrveranstaltungsaufzeichnungen und Videos der Lehrenden abgelegt und gespeichert werden können und im Lernmanagementsystem eingebettet werden können. Der Vollständigkeit halber ist auch darauf hinzuweisen, dass es an der TU Graz mit der MOOC-Plattform imoox.at eine weitere Plattform für offen lizenzierte Materialien gibt, bei der Lehrende OER erstellen. Wie beim LMS TeachCenter sowie dem Videoportal TUbe fehlt hier eine Möglichkeit, die Materialien über das universitätseigene Repositorium bzw. der OER-Plattform der Universität Wien anderen leichter für die Recherche zur Verfügung zu stellen.

4.2 Grobkonzept für die technische Lösung und Vorgehen im Überblick

Um die Daten aus dem TeachCenter ins TU Graz Repositorium bzw. in das OER-Fachportal zu transferieren ist es notwendig, Lehrenden die Möglichkeit geben, die entsprechenden Metadaten zu ergänzen und Schnittstellen (API) zu entwickeln. Abb. 2 stellt das notwendige LMS-Plug-In und Verortung der APIs dar.

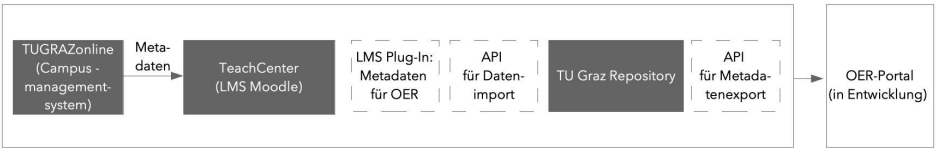


Abb. 2: Grobkonzept der technischen Lösung zur Schließung der Lücken der Infrastruktur

Das Vorgehen bei der Entwicklung war dabei folgendes: Zunächst musste festgelegt werden, auf welche Weise die OER im Repositorium beschrieben werden sollen, also welche Metadaten genutzt werden sollen. Dazu musste ein Standard gewählt werden und analysiert werden, welche Daten bereits vorhanden sind, welche unbedingt notwendig sind und welche ggf. von den Lehrenden ergänzt werden müssen.

4.3 Wahl des Standards für Metadaten und Analyse vorhandener und fehlender Metadaten

Für die Umsetzung der OER ist MAB allerdings nicht ausgelegt. Dies führt dazu, dass ein Standard gesucht werden musste, der Lern- und Lehrressourcen beschreiben kann. Der Metadaten-Standard Learning Object Metadata (LOM) war zum Analysezeitpunkt schon in mehreren Projekten eingesetzt, die LOM-Metdaten lassen sich auch übersetzen (Educa.ch, 2017, siehe auch Abschnitt 2). LOM wurde daher ausgewählt, und das Datenmodel des Repositoriums wurde an LOM angepasst. Das heißt es wurden zusätzliche Felder, entsprechend der LOM-Semantik, implementiert.

Da wir davon ausgehen, dass Lehrende nur wenig erpicht darauf sind, zusätzliche Metadaten zu ihren Lernobjekten und -einheiten in ein System einzugeben, stellte sich nun die Frage, welche LOM-Metadaten bereits im Campusmanagementsystem erfasst sind. Um ein kohärentes Vorgehen zu ermöglichen, erfolgte diese Äquivalenzprüfung der Metadaten und LOM Analyse und Auswahl der relevanten Metadaten in Kooperation mit der Universität Wien, die für das OER-Fachportal und das eigene Repositorium an einem gemeinsamen Vorgehen und Auswahl interessiert ist. Es wurden daher die Metadaten aus den Systemen der TU Graz und die Metadaten des OER-Fachportals der Universität Wien mit LOM verglichen.

Es wurde eine Äquivalenzliste (s. Tabelle 1) ausgearbeitet. Sie zeigt welche der LOM-Daten in den Informationssystemen der beiden Universitäten vorhanden sind.

Documentation			*Repository: OER API*		
teachcenter		openlib	description[static]	vienna	lom
		1000	unique id		
		1050	children ids		
		1051	all child ids		
	C	1102	child type [course]		
	O	1600	creation date node		
	U	1601	modification date node		
title course	R	331	title	title course	
science field	S	1800	science field	science field	
location	E	1507	location [graz]	location	coverage [1.6]
intended end user role		1801	intended end user role [learner]	intended end user role	intended end user role [5.5]
context		1802	context [university]	context	context [5.6]
course language		1301	language	course language	
		1000	unique id		
		1050	children ids		
		1051	all child ids		
		1052	root ids		
		1053	parent ids		
		1102	child type [unit]		
		1600	creation date node		
		1602	modification date node		
institute		1803	institute	institute	
year		1401	year	year	
semester	U	1409	semester	semester	
description	N	1500	abstract	description	
university lecturer	I	100	author 1	university lecturer	contribute [2.3]
	T	103	corporate body 1	corporate body lecturer	role [2.3.1]
contributor		104	author 2	contributor	
categories		710	subject	categories	
course type		1804	course type	course type	structure [1.7]
		1000	unique id	unique identifier	identifier [1.1] [3.1]
		1052	root ids		
		1053	parent ids		
		1102	child type [file]		
		1600	creation date node		contribution to metadata [3.2]
		1601	modification date node		contribution to metadata [3.2]
author		100	author	author	author [2.3]
		103	corporate body	corporate body	role [2.3.1]
filename		331	title	filename	title [1.2]
filesize		1200pb	filesize	filesize	size [4.2]
		1200pa	filename[hash value of file]		
abstract	F	1500	abstract	abstract	
file language	I	1301	language	file language	language [1.3]
cost	L	1709	cost [none]	cost	cost [6.1]
	E		copyright a. other restrictions [yes]	copyright other restrictions	copyright and other restrictions [6.2]
licence		1701	licence [cc-*]	licence	description of rights [6.3]
reference program		1214	reference program	reference program	name of required technology [4.4.1.2]
categories		710	subject	categories	
subjects		710	subject	subjects	
file creation date		1604	file creation date	date	version [2.1] [2.3.3]
		1602	file modification date		
oefos		1508	oefos	oefos	
resourceType		1109	resource type		
	L			metadaten link	
	I			download link	location [4.3]
	N			id	
	K				

Tab. 1: Äquivalenzliste der Felder des Campusmanagementsystems der TU Graz, des OER-Fachportals der Universität Wien sowie des LOM-Standards. Quelle: Ladurner, 2018, Tabelle 1

Der Vergleich von LOM der Metadaten der Informationssysteme der TU Graz und der Universität Wien zeigt, dass es große Überschneidungen gibt. Allerdings gibt es auch Felder in den Systemen der TU Graz, die bei der Universität Wien nicht vorhanden sind (z. B. *resourceType*). Da die Universitäten eine kompatible Lösung anstreben wird dieses Feld nicht weiter berücksichtigt. Einige der Felder von LOM, die in der Äquivalenzliste auch in den Informationssystemen beider Universitäten vorhanden sind wurden nicht ausgewählt (z. B. *cost, reference program*), weil sie als nicht als relevant erscheinen.

Schematisch gibt es also unterschiedliche als notwendig identifizierte Metadaten auf Basis einer Auswahl von Metadaten auf Grundlage des LOM-Schemas. Sie können aus unterschiedlichen vorhandenen Quellen, nämlich den Informationssystemen der TU Graz wie der Datei selbst übernommen werden. Ein Teil muss aber weiterhin durch die Autor/inn/en selbst ergänzt werden bzw. muss durch sie editierbar sein.

4.4 Automatische Bereitstellung und Ergänzung der Metadaten für OER in einem Plug-In für das Lernmanagementsystems

Die Analyse und Auswahl ergab für die TU Graz, das viele Metadaten vom Lernmanagementsystem, der Datei selbst bzw. dem Campusmanagementsystem bereitgestellt werden können, z. B. die/der Autor/in (wenn nicht explizit eingetragen, wird hier der/die Person verwendet, die/der die Datei hochgeladen hat), Lizenz, der Name der Datei, Dateigröße und Dateityp, Hochladedatum und Änderungsdatum, Kurs (Sprache, Lehrveranstaltungs-Typ, Lehrende), Fakultät/Institut (Studium, Semester), Name der Person, die die Datei hochgeladen hat sowie Schlagwörter (Tags, falls verwendet). Für einige Felder und Metadaten musste jedoch eine Eingabemaske im LMS erstellt werden, um den Benutzer/innen die Möglichkeit zu geben, die Metadaten anzugeben bzw. anzupassen.

Für die Lehrenden wurde daher ein Plug-In für das Lernmanagementsystem entwickelt, in dem sie angeben können, welche Dateien unter einer Creative Commons (CC)-Lizenz gestellt und in das Repositorium exportiert werden dürfen. Da im LMS keine Metadaten für einzelne Dateien nötig sind, müssen noch einige von der Kursleitung hinzugefügt werden. Alle Lehrenden sehen Menüpunkt „OER Plug-In". Nur Personen mit Berechtigung können Dateien hochladen. Personen ohne Berechtigung erhalten Informationen darüber was OER sind und wie sie die Berechtigung erhalten, OER zu veröffentlichen. Abb. 3 zeigt das Plug-In ausschnittsweise, und verdeutlicht exemplarisch, dass die Metadaten im Plug-In unterschiedlichen Quellen stammen: So wird das Semester und der Kontext aus dem Lernmanagementsystem (TeachCenter) übernommen, die komplette Kursbeschreibung stammt aus dem Campusmanagementsystem (TUGRAZ online). Die Angabe über die Größe der Datei wird von der Datei selbst ausgelesen. Lehrende müssen und können also nur relativ wenige Metadaten editieren: Dateiname, Sprache, Ressourcentyp, Rolle, Autor (Urheber/in), CC-Lizenz, Schlagworte sowie die OEFOS-Klassifikation.

4.5 Entwicklung der Application Programming Interfaces

Die Application Programming Interfaces (API) gliedern sich in eine Import- und eine Export-Richtung. Es wird dem LMS ein REST-API angeboten, die es ermöglicht den gesamten Kurs als eine ZIP-Datei in das Repositorium der TU Graz einzuspielen. Die API wurde dafür sehr einfach gehalten. Ein Token ist für die Authentifizierung zuständig und die Dateipaare sind in einer ZIP-Datei verpackt. Ein Dateipaar besteht aus einer Datei für

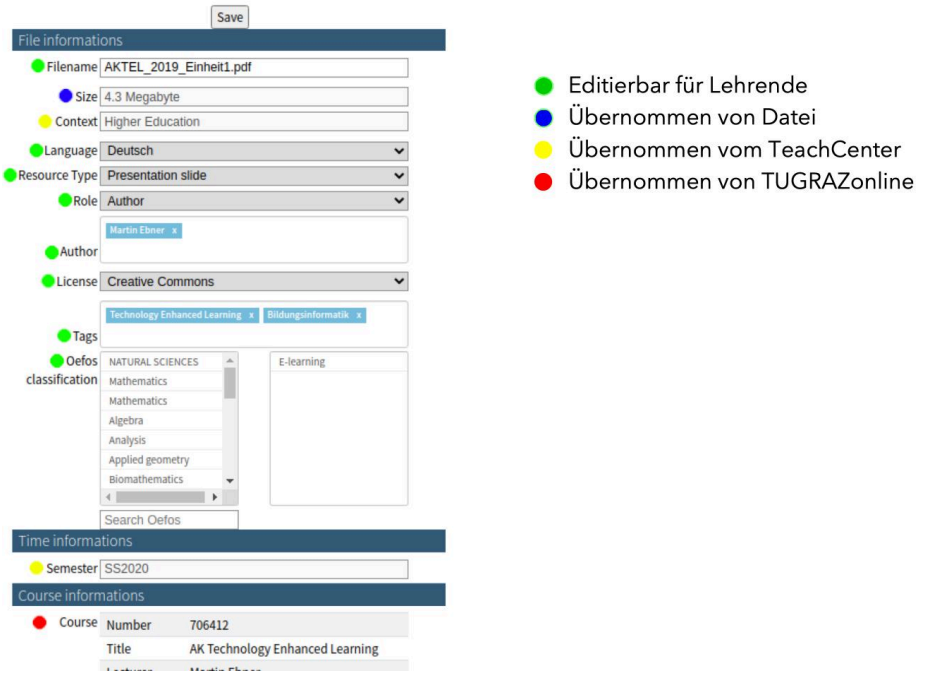


Abb. 3: Screenshot des Plug-Ins und Legende zum Ursprung der Daten und Möglichkeiten der Dateneingabe durch die Lehrenden (Auswahl)

Metadaten und einer downloadbaren Datei. Wie das Repositorium der TU Graz wurde auch die API in PHP programmiert. Im Detail wurde für die „API Import“ so zunächst die URL festgelegt (`https://openlib.tugraz.at/upload.php`). Der *token* identifiziert die importierende Institution und gibt ihr damit das Recht, Dateien hochzuladen. Als *package* wird eine ZIP-Datei definiert, die Dateipaare (JSON-Datei für die Metadaten und eine Datei ohne Dateiendung, welche die beschriebene Datei darstellt) enthält. Die Fehlermeldungen lehnen sich an SWORD an. Die Metadaten werden wie in Tabelle 1 dargestellt für den Kurs (course), die Einheit (unit) sowie die einzelne Datei (file) dargestellt.

Auch die Schnittstelle für den Export der Metadaten, insbesondere für das OER-Fachportal, wurde bewusst einfach gehalten. Die Metadaten werden in eine JavaScript Object Notation (JSON)-Datei verpackt und in das Discovery-Tool per Representational-State-Transfer (REST) exportiert. Die Metadaten haben dieselbe Struktur wie die Dateien, die importiert werden. Es wird jedoch das Attribut *links* hinzugefügt. Dies enthält *id*, *course* und *file*. Zudem wird in *course* das Attribut *location* mit *Technische Universität Graz* hinzugefügt. Die Dateien selbst bleiben im Repositorium und sind über einen Persistent Identifier erreichbar. Der Upload erfolgt dann wieder über REST auf die Testinstanz (`https://portal.openeducation.at/upload/json/v1/openlib.tugraz.at`).

4.6 Prozessgestaltung: OER-Zertifizierung von Lehrenden

Neben den technischen Lösungen ist es auch notwendig, auf Seiten der Lehrenden Prozesse zu schaffen, dass diese kompetent OER erstellen können und rechtliche Schwierigkeiten vermieden werden können. Wie in Abb. 4 dargestellt, wurde dafür an der TU Graz eine OER-Weiterbildung angeboten. Lehrende, die diese Weiterbildung im Umfang von einem ECTS (entspricht 25 Stunden) erfolgreich abgeschlossen haben und OER erstellt haben bekommen das Plug-In freigeschalten. Die Weiterbildung umfasst Präsenztraining und die erfolgreiche Teilnahme am MOOC zu OER, der über die Plattform iMooX.at zur Verfügung steht. Für die erfolgreiche Teilnahme am MOOC ist das Ablegen von mehreren Tests je Einheit notwendig. Zertifiziert werden also nicht einzelne OER im Sinne einer Kontrolle, sondern es werden Lehrende dazu weitergebildet, dass sie die rechtlichen Voraussetzungen für den Umgang mit und die Erstellung von OER kennen. Die Inhalte der Weiterbildung bzw. die OER-Zertifizierung der TU Graz orientiert sich dabei an den Vorschlägen des Forum Neue Medien in der Lehre Austria zu offenen Bildungsressourcen (Ebner et al., 2016a) und dem Whitepaper zur OER-Zertifizierung in Österreich (Ebner et al., 2016b; Ebner, 2018).

Abb. 4: Prozessmodellierung der Qualifikation und Rechtevergabe zur Veröffentlichung von OER an der TU Graz

4.7 Implementierung und bisherige Nutzung

Alle Systeme und Prozesse sind an der TU Graz vollständig implementiert und im Produktiveinsatz:

An der TU Graz ist bisher nach der beschriebenen OER-Weiterbildung und „OER-Zertifizierung“ für sieben Lehrende das Plug-In freigeschalten worden, ein Teil von ihnen hat es bereits im letzten Semester genutzt, so dass die entsprechenden Dateien und Metadaten im Repositorium der TU Graz zu finden und bereits im OER-Fachportal recherchierbar sind. Im ersten Semester wurden so Daten aus vier Lehrveranstaltungen aus dem LMS in das OER-Repositorium der TU Graz übertragen. Auch die Übertragung der Metadaten der OER-Materialien aus der TU Graz OPEN Library ist implementiert und lauffähig.

5 Ausblick und Empfehlungen

Wir haben unsere einzelnen Entwicklungsschritte hier detailliert beschrieben und vorgestellt um anderen ggf. die Möglichkeit zu geben, an ihren Universitäten und Hochschulen

ähnliche Schnittstellen zu entwickeln und Prozesse zu implementieren. Die entwickelten Technologien und Prozesse sind im produktiven Einsatz; die entsprechenden Prozesse und Technologien wurden bei der TU Graz implementiert: Das Plug-In ist im Einsatz und der Export der Materialien wurde schon wie dargestellt erfolgreich durchgeführt. Eine weitreichende interne Ausrollung der OER-Zertifizierung und Nutzung des OER-Plug-In in der TU Graz ist noch nicht gestartet. Die nächsten Entwicklungsschritte an der TU Graz sind u. a. eine Anpassung des OER-Plug-Ins an Moodle 3.9. Denkbar sind auch ähnliche Plug-Ins für die Videoplattform TUbe. Ähnliches gilt für das OER-Fachportal der Universität Wien, das bereits im Produktivmodus ist, es stehen aber noch Erweiterungen und der öffentliche Launch aus. Diese Weiterentwicklungen auch die Implementierung einer österreichweiten OER-Zertifizierung werden dazu im Rahmen des Projekts „Open Education Austria Advanced" bis 2024 fortgeführt.

Abschließend möchten wir Nachahmer/innen folgende Empfehlungen geben: Erstens sollte der Mehraufwand für die Ersteller/innen von OER im Lernmanagementsystem bzw. Repositorium möglichst klein bleiben: Es sollte sich daher auf die notwendigsten Daten konzentriert werden. Da lohnt es sich unserer Sicht, genau hinsehen: Wir waren überrascht, dass es doch sehr viele Metadaten gibt, die bereits zu den Bildungsressourcen zur Verfügung stehen, z. B. für welchen Studiengang oder in welchem Semester sie genutzt werden. Bei der Auswahl der Metadaten empfehlen wir darauf zu achten, welche Metadaten von anderen OER-Infrastrukturpartnern benötigt werden. Schließlich sollten Standards genutzt werden, bei den Metadaten wie bei der Entwicklung von APIs.

Weitere Details zu Planng und Umsetzung sind in einem ausführlichen Arbeitsbericht zu entnehmen, der ab Oktober 2020 auf der Projektwebsite zu finden ist (https://openeducation.at).

Danksagung

Die hier vorgestellte Entwicklungsarbeit wurde durch Fördermittel des Bundesministerium für Bildung, Wissenschaft und Forschung, Österreich, im Rahmen der Ausschreibung zur digitalen und sozialen Transformation in der Hochschulbildung 2019 für das Vorhaben „Open Education Austria Advanced" (2021-2024) ko-finanziert; Partner: Universität Wien, TU Graz, Universität Graz, Universität Innsbruck, Forum Neue Medien in der Lehre Austria, ÖIBF.

Literaturverzeichnis

[BC10] Barker, P.A. & Campbell, L.M.: Metadata for learning materials: An overview of existing standards and current developments. Technology, Instruction, Cognition and Learning, 7 (3–4), 225–243, 2010.

[Bu13] Butcher, N.: Was sind Open Educational Resources? Und andere häufig gestellte Fragen zu OER. Bonn: Deutsche UNESCO-Kommission, 2013. URL: https://www.unesco.de/fileadmin/medien/Dokumente/Bildung/Was_sind_OER__cc.pdf

[CC20] Creative Commons. http://de.creativecommons.org/ (2020-05-05).

[DN19] Deutsche Nationalbibliothek (2019). Maschinelles Austauschformat für Bibliotheken (MAB). URL https://www.dnb.de/DE/Standardisierung/Formate/MAB/mab_node.html. (2019-05-31)

[Eb18] Ebner, M.: OER-Certification in Higher Education. In Proceedings of EdMedia: World Conference on Educational Media and Technology. Amsterdam, Netherlands: Association for the Advancement of Computing in Education (AACE). S. 1-6, 2018, URL: https://www.researchgate.net/publication/326034702_OER-Certification_in_Higher_Education

[EHO17] Ebner, M.; Haas, M. & Ortner, C.: IST-Zustand Learningmanagementsystem TU Graz. Interner Arbeitsbericht für das Projekt "Open Education Austria" vom 17.4.2017.

[Eb15] Ebner, M., Muuß-Merholz, J., Schön, M. und Schön, S.: Bildungsbereichsübergreifende Entwicklungen. In: Martin Ebner, Elly Köpf, Jöran Muuß-Merholz, Martin Schön, Sandra Schön und Nils Weichert (Hrsg.), Ist-Analyse zu freien Bildungsmaterialien (OER), Wikimedia: Berlin, S. 10-34, 2015,

[Eb16a] Ebner, M., Kopp, M., Freisleben-Teutscher, C., Gröblinger, O., Rieck, K., Schön, S., Seitz, P, Seissl, M., Ofner, S., Zimmermann, C., Zwiauer, C.: Recommendations for OER Integration in Austrian Higher Education. In: Conference Proceedings: The Online, Open and Flexible Higher Education Conference, EADTU 2016, S. 34-44.

[Eb16b] Ebner, M., Freisleben-Teutscher, C., Gröblinger, O., Kopp, M., Rieck, K., Schön, S., Seitz, P., Seissl, M., Ofner, S. & Zwiauer, C.: Empfehlungen für die Integration von Open Educational Resources an Hochschulen in Österreich. Forum Neue Medien in der Lehre Austria, 2016.

[ES13] Ebner, M. & Schön, S.: Offene Bildungsressourcen als Auftrag und Chance –Leitlinien für (medien-)didaktische Einrichtungen an Hochschulen. In: G. Reinmann, M. Ebner & S. Schön (Hrsg.), Hochschuldidaktik im Zeichen von Heterogenität und Vielfalt. Doppelfestschrift für Peter Baumgartner und Rolf Schulmeister, Norderstedt: BoD, 2013, S. 7-28, URL: http://bimsev.de/festschrift (2015-05-05).

[Eb20] Ebner, M.; Schön, S.; Braun, C.; Ebner, M.; Grigoriadis, Y.; Haas, M.; Leitner, P.; Taraghi, B.: COVID-19 Epidemic as E-Learning Boost? Chronological Development and Effects at an Austrian University against the Background of the Concept of "E-Learning Readiness". Future Internet 2020, 12, 94.

[ESK16] Ebner, M., Schön, S., & Kumar, S.: Guidelines for leveraging university didactics centers to support OER uptake in German-speaking Europe. Education Policy Analysis Archives, 2016, 24 (39). http://dx.doi.org/10.14507/epaa.24.1856

[Ed17] Educa.ch: Applikationsprofil lom-ch, 2017. URL: https://www.educa.ch/de/online-zugang/lom-ch (Stand 2017)

[EC13] Europäische Kommission: Die Bildung öffnen: Innovative Lernen für alle mithilfe neuer Technologien und frei zugänglicher Lehr-und Lernmaterialien. Mitteilung der Kommission an das Europäische Parlament, den Rat, den Europäischen Wirtschafts-und Sozialausschuss und den Ausschuss der Regionen, 25.9.13. URL: http://eur-lex.europa.eu/legal-content/DE/ALL/?uri=CELEX:52013DC0654 (2015-05-05).

[Ha19] Haas, M.: Die Schnittstelle zur Übergabe von OER an das Bibliothekssystem und einer möglichst automatisierten Erfassung von Metadaten. Vortrag bei den Open-Access-Tagen 2019, 25.9.2019, Graz.

[KIM20] KIM-AG: OER-Metadatengruppe. URL: https://wiki.dnb.de/display/DINIAGKIM/OER-Metadatengruppe (2020-07-03)

[La19] Ladurner, C.: Repository: OER API. Documentation. Version vom 29.7.2019. Universitätsbibliothek, Technische Universität Graz.

[MP20] Menzel, M. & Pohl, A.: LOM for Higher Education OER Repositories. Beschreibung zur XML Schema Definition des Metadatenprofils für Open Educational Resources im Hochschulbereich. Spezifikation vom 28. Februar 2020, https://dini-ag-kim.github.io/hs-oer-lom-profil/latest/

[Mr13] Mruck, K.; Mey, G.; Schön, S.; Idensen, H. & Purgathofer, P.: Offene Lehr-und Forschungsressourcen. Open Access und Open Educational Resources. In: M. Ebner & S. Schön (Hrsg.), Lernen und Lehren mit Technologien (L3T). Ein interdisziplinäres Lehrbuch, Berlin: epubli, 2013. URL: http://l3t.eu (2015-05-05).

[OER15] OER Metadatengruppe: OER Metdatengruppe, DNB, https://wiki.dnb.de/display/DINIAGKIM/OER-Metadaten-Gruppe (2015-05-05)

[Po14] Pohl, A.: Empfehlungen zur Publikation von OER-Metadaten (Entwurf), Stand 2014-12-19, URL: https://wiki.dnb.de/pages/viewpage.action?pageId=94678918 (2015-05-05).

[Re13] Rensing, C.: Standards für Lehr-und Lerntechnologien. Metadaten, Inhaltsformate und Beschreibung von Lernprozessen. In: M. Ebner & S. Schön (Hrsg.), Lehrbuch für Lernen und Lehren mit Technologien (L3T). URL: http://l3t.eu (2015-05-05), 2013.

[St17] Steiner, T.: Metadaten und OER: Geschichte einer Beziehung. In: Synergie : Fachmagazin für Digitalisierung in der Lehre, 2017. 4, S. 51-55, URL: https://doi.org/10.17613/M6P81G - ISSN: 2509-3088; 2509-3096

[Sw19] Sword: Swordapp. URL: http://swordapp.org/ (Stand Mai 2019)

[UN12] UNESCO: Pariser Erklärung zu OER. Weltkongress zu Open Educational Resources (OER), Paris, Juni 2012. Übersetzung nach Deutsche UNESCO-Kommission (siehe Butcher, 2013, Anhang).

[ZDN13] Ziedorn, F.; Derr, E. & Neumann, J.; Metadaten für Open Educational Resources (OER). Eine Handreichung für die öffentliche Hand, Hannover: Technische Informationsbibliothek (TIB), 2013 URL: http://www.pedocs.de/volltexte/2013/8024/pdf/TIB_2013_Metadaten_OER.pdf

[Wi19] Wikipedia: Learning objects metadata. In: Wikipedia, die freie enzyklopädie. URL: https://de.wikipedia.org/w/index.php?title=Learning_Objects_Metadata&oldid=177337688, 2018. (2019-05-20)

Fachanwendung für digitale Modulkataloge

Eine Untersuchung zu Graph-basierter Daten-Modellierung und Navigation

Vera G. Meister [1], Wenxin Hu [2], Aleksandra Revina [3], Marcel Cikus [4], Johannes Müller[5]

Abstract: Hochschulen nutzen eine Vielzahl von Fachanwendungen für die verschiedensten Prozesse zur Konzeption, Publikation, Verwaltung, Planung, Durchführung und Leistungsabrechnung von Modulen, oftmals mit redundanter Datenhaltung. Fehlende Integration der Anwendungen sorgt für mangelhafte Datenqualität und substanzielle Mehrarbeit. Das Paper untersucht mit Hilfe eines Prototyps, inwiefern eine Graph-basierte Datenhaltung ein Ansatz zur Lösung des Integrationsproblems sein kann. Zugleich wird erforscht, ob eine Graph-Navigation bei den Nutzer*innen auf Akzeptanz stößt und in welchem Maße sie die Orientierung in vernetzten Strukturen fördert.

Keywords: Digitaler Modulkatalog; Wissensgraph-Anwendung; Hochschul-IT; integrierte Datenhaltung; Graph-basierte Navigation; Prototyp; Softwareevaluation

1 Einleitung

Im Ergebnis der Bologna-Reform wurden die Studiengänge an Hochschulen modularisiert, d. h. sie setzen sich aus einzelnen, fachlich differenzierbaren, durch Prüfung abschließbaren Lerneinheiten – den Modulen – zusammen [Ku10]. Diese Lerneinheiten sind zu konzipieren, zu spezifizieren, zu publizieren, mit Ressourcen zu belegen und schließlich durchzuführen und abzuschließen. Ein Modul ist damit zum einen ein ideelles Konzept, zum anderen eine Veranstaltung in Raum und Zeit.

In einer digitalen Hochschullandschaft sollten alle Prozesse rund um diese zentralen Entitäten durch digitale Werkzeuge – sogenannte Fachanwendungen – unterstützt werden. Idealerweise sollten die Daten zu den Modulen und weiteren verbundenen Entitäten integriert sein, d. h. nicht mehrfach in verschiedenen, isolierten Datenbanken verarbeitet und gespeichert werden. Die IT-Landschaften deutscher Hochschulen sind häufig aufgrund verteilter Verantwortlichkeiten in den Instituten mäßig bis gar nicht integriert.

[1] Technische Hochschule Brandenburg, Fachbereich Wirtschaft, Magdeburger Straße 50, 14770 Brandenburg an der Havel, Deutschland, vera.meister@th-brandenburg.de, https://orcid.org/0000-0002-2780-0222
[2] ebenda, hu@th-brandenburg.de, https://orcid.org/0000-0003-3449-5980
[3] ebenda, revina@th-brandenburg.de, https://orcid.org/0000-0002-8405-0018
[4] ebenda, cikus@th-brandenburg.de, https://orcid.org/0000-0001-6715-2146
[5] ebenda, johannes.mueller@th-brandenburg.de

doi:10.18420/inf2020_43

Aufbauend auf Vorarbeiten [MB18; MHP19] soll in diesem Paper die prototypische Implementierung einer Fachanwendung für digitale Modulkataloge untersucht werden. Der Fokus wird auf die integrativen Fähigkeiten einer Graph-basierten Datenhaltung und den Einfluss einer solchen Navigation auf Nutzerakzeptanz gelegt. Dafür werden in Abschnitt 2 Motivation und Forschungsfragen ausgearbeitet sowie in Abschnitt 3 der Stand der Forschung reflektiert. In den Abschnitten 4 Prototypentwicklung und 5 Evaluation werden die Methoden zur Beantwortung der Forschungsfragen eingeführt und die Ergebnisse dargestellt. Das Paper schließt mit einem Fazit und Ausblick in Abschnitt 6.

2 Motivation und Forschungsfragen

Wie in [MHP19] gezeigt, ist die Bandbreite an eingesetzten Fachanwendungen für Modulkataloge an Hochschulen sehr groß. Im nicht seltenen Extremfall gibt es für jeden Prozess rund um Module ein anderes Werkzeug: Office-Tools und/oder Dokumenten-Management-Systeme (DMS) für die Konzeption und Bereitstellung von Modulbeschreibungen, Content-Management-Systeme (CMS) für die Veröffentlichung auf Hochschulwebseiten, komplexe Tabellenkalkulationen für die Planung von Lehrbedarfen und -verpflichtungen, Spezialanwendungen für die Stunden- und Raumplanung, Lernmanagement-Systeme (LMS) für die didaktische und organisatorische Steuerung der Lernprozesse sowie Campus-Management-Systeme (CaMS) für das Prüfungsmanagement im Student Life Cycle. Daten zu Modulen werden in einem solchen Setting vielfach redundant erfasst und gespeichert, was zu mangelhafter Datenqualität und substanziellem Mehraufwand führt.

Um eine integrierte Datenhaltung herzustellen gibt es vier verschiedene Ansätze [Kr20]:

1. Ablösung der einzelnen Fachanwendungen durch ein integriertes komplexes System mit einheitlicher Datenhaltung,
2. Einführung einer integrierenden Middleware basierend auf einem vorkonfigurierten, einheitlichen Datenschema,
3. Programmierung dedizierter Schnittstellen für die Point-to-Point-Integration der verschiedenen Fachanwendungen,
4. Einführung eines flexibel erweiterbaren, Graph-basierten Daten-Hubs, der die Relationen und Interdependenzen zwischen den verschiedenen Datenquellen expliziert und zugänglich macht.

Einige große Hersteller von CaMS versuchen sich an Lösungsansatz 1. Die Praxis zeigt jedoch, dass nicht nur die komplexen Einführungsprozesse, sondern auch die Spezifik der verschiedenen Hochschulen diesem One-Fits-All-Ansatz zuwiderlaufen. Ansatz 2 erfordert ein vollständig ausgearbeitetes Integrationsmodell, das nur mit hohem Aufwand an Änderungen angepasst werden kann. Der wesentliche Nachteil von Ansatz 3 besteht in seinem hohen Pflegeaufwand, der quadratisch zu jeder weiteren zu integrierenden Anwendung ansteigt. Ansatz 4 wiederum ist offen im Hinblick auf die Datenmodelle der zu integrierenden

Anwendungen und zugleich flexibel erweiterbar. Der größte „Nachteil" besteht hier darin, dass es sich um eine Technologie handelt, der unter IT-Verantwortlichen und Softwareentwicklern eine steile Lernkurve bescheinigt wird [Tu20]. Zugleich weist [Tu20] nach, dass im letzten Jahrzehnt Graph-Technologien in einer Reihe von Forschungsfeldern (wie Biomedizin) und Branchen (Finanzwirtschaft, Maschinenbau u. a.) eine weite Verbreitung gefunden haben.

Der Einsatz von Graph-Technologien ist besonders dann angezeigt, wenn es sich um stark vernetzte Daten handelt, die zudem verschiedenen Fachsichten unterliegen, wie es sich auch im vorliegenden Anwendungsfall darstellt. So sind aus Sicht des Studiengangmanagements Module ideelle Konzepte, die mit Lehrenden, Organisationseinheiten, didaktischen Methoden, Ordnungen, Lehr- und Prüfungsformen etc. vernetzt sind. Aus Sicht der Ressourcenplanung handelt es sich um (wiederkehrende) Veranstaltungen mit räumlichen, personellen und weiteren Anforderungen. Beide Sichten sind miteinander verbunden, ergänzen einander und erlauben zudem eine Vielzahl weiterer prozessspezifischer Projektionen. Somit erscheint es grundsätzlich angezeigt, auf der Datenintegrationsebene mit Graph-Technologien zu arbeiten. Davon abzugrenzen ist die Frage, ob diese Graph-Sicht auch auf die Endnutzerebene ausgedehnt werden kann und soll. Die bereits zitierte steile Lernkurve [Tu20] kann sich negativ auf die Akzeptanz auswirken, könnte aber zugleich einen wertvollen Beitrag zur Bekanntheit der Technologie als Ganzes leisten und damit die Hürden für einen breiteren praktischen Einsatz senken.

Aus dem Vorgesagten ergeben sich die folgenden Forschungsfragen:

1) Wie ist ein potenziell erweiterbares Graph-Schema für digitale Modulkataloge zu konzipieren?
2) Mit welchen Front-End-Technologien kann eine Graph-basierte Navigation implementiert werden?
3) Wie wirkt sich die Graph-Navigation auf die Nutzerakzeptanz der Fachanwendung für digitale Modulkataloge aus?

Zur Beantwortung der Forschungsfragen 1) und 2) wird die Methode des Prototyping angewandt. Die Beurteilung der Nutzerakzeptanz gemäß Forschungsfrage 3) erfolgt auf Basis einer qualitativen Untersuchung, die darüber hinaus der Ableitung von Anforderungen für die Weiterentwicklung der Fachanwendung dient.

3 Stand der Forschung

Entlang der drei Forschungsfragen sollen in diesem Abschnitt zunächst die relevanten Forschungsfelder adressiert und der Forschungs- und Entwicklungsstand mit Bezug zur jeweiligen Fragestellung dargestellt werden.

3.1 Graph-basierte Daten-Modellierung

Jede Fachanwendung basiert auf der Konzeption und der technischen Implementierung eines domänenspezifischen Datenmodells. Formal und logisch stringente, fachlich adäquate sowie potenziell erweiterbare Graph-basierte Datenmodelle zu entwickeln, erfordert neben einer tiefen Kenntnis der fachlichen Domäne zugleich ein ganzes Bündel weiterer Kompetenzen: Abstraktionsvermögen, Verständnis von Mustern und Antimustern, Orientierung in bestehenden Vokabularen und Spezifikationen, Beherrschung geeigneter Tools. Zu den letzten drei Feldern gibt es eine Fülle an neueren Forschungs- und Entwicklungsergebnissen.

Der Begriff Ontology Design Pattern (ODP) wurde 2005 von Aldo Gangemi geprägt [Ga05]. Die bis heute aktive Forschungs-Community stellt ihre Ergebnisse auf einem semantischen Wiki der breiten Nutzung zur Verfügung [OD20]. Für den vorliegenden Anwendungsfall sind insbesondere die Content-ODPs und im Hinblick auf Nachnutzung bzw. Erweiterung vorhandener Ontologien (Schemata) die Alignment-ODPs relevant. [Su12] thematisiert die Nutzung von Mustern für die visuelle Darstellung von Graph-Schemata. Eine umfassende Darstellung der Forschung zu ODPs findet sich in [Hi16].

Die Entwicklung von Vokabularen und Spezifikationen für die Graph-basierte Daten-Modellierung hat im letzten Jahrzehnt zahlreiche wertvolle Ergebnisse hervorgebracht. Für den Anwendungsfokus von besonderem Nutzen sind drei mächtige Spezifikationen des W3C: Turtle, SPARQL und SHACL[1]. Auch bei den Vokabularen in der Anwendungsdomäne sind die Ergebnisse zweier vom W3C kuratierter Community-Groups besonders relevant: zum einen schema.org mit Fokus auf `schema:Course` und `schema:CourseInstance` und zum anderen das Verifyable Credentials Data Model.

Bei den Tools für die praktischen Modellierung von Daten-Graphen, lassen sich drei Konstruktions-Paradigmen unterscheiden: (i) Code, (ii) Baum- und (iii) Graph-Struktur. Für alle drei Paradigmen gibt es Editoren bzw. auch komplexere Tools, allerdings folgt die Mehrzahl dem Baumstruktur-Paradigma. Zu diesen Werkzeugen gehört der klassische Open-Source-Editor Protégé [Mu15] ebenso wie das semantische Wiki (ebenfalls Open Source) OntoWiki [FAM16] sowie die leistungsfähigen kommerziellen Tools Top-BraidComposer [To20], Corporate Memory [Br19] und PoolParty [Po20b]. Alle Tools verfügen über Features oder PlugIns, um die Schema-Graphen zu visualisieren sowie über Export-Funktionen, um RDF-Dateien (also reinen Code) auszugeben. Dennoch besteht der größte Nachteil dieser Werkzeuge darin, dass ein sehr starker Fokus auf die Klassenhierarchie gelegt wird, während die Modellierung sämtlicher anderer Relationen zwischen Klassen nur mittelbar unterstützt wird.

Grafische Editoren sind weniger weit verbreitet. OWLGrEd [Ce19] ist ein Open-Source-Editor, der sich in der Visualisierung an UML-Klassendiagrammen orientiert, während Grafo [Da20] (ein kommerzielles Tool) die VOWL-Technologie [Lo16] implementiert. Bei beiden sind individuelle grafische Anpassungen nur bedingt möglich. Allerdings verfügen auch sie über Export-Funktionen. Der Support von Code-Editoren (z. B. rdfEditor

[1] https://www.w3.org/TR/turtle/, https://www.w3.org/TR/turtle/, https://www.w3.org/TR/shacl/

[Do20]) beschränkt sich auf Syntax-Highlighting und Syntaxprüfung für verschiedene Serialisierungsformate. Der größte Vorteil eines Code-Editors besteht darin, dass Struktur und Schlankheit des Codes in der Hand der Entwickler*innen verbleiben. Schließlich sei erwähnt, dass es eine Reihe von Werkzeugen für die verteilte Modellierung gibt. Genannt sei hier nur Quit Store [ARM16], basierend auf Git-Technologien. Dank der Import- und Export-Funktionen bzw. der Einbindung von PlugIns können Entwickler*innen nach Bedarf von einem Präsentationsparadigma zu einem beliebigen anderen wechseln.

3.2 Front-End-Technologien für die Graph-Navigation

Grafische Modelle dienen allgemein der Visualisierung komplexer Zusammenhänge. Sie erleichtern damit die Orientierung und verbessern die Aufnahme zusammenhängender Informationen. Dafür gibt es eine Vielfalt von Anwendungsgebieten: Ablaufdiagramme und Prozessmodelle, Strukturdiagramme, Organigramme, Netzpläne und nicht zuletzt semantische Graph-Schemata. Graphische Editoren verfügen zwangsläufig auch über Elemente einer Graph-Navigation, um Detailinformationen zu Graph-Elementen zu erfassen bzw. wiederzugeben. Aktuelle Beispiele dafür sind der Camunda Modeler [Ca20] für die Modellierung von Geschäftsprozessen und -entscheidungen sowie der bereits erwähnte Ontologie-Editor Grafo [Da20]. Aber auch Informationsangebote und Dokumentationen können von einer Graph-Navigation profitieren, wie z. B. die Process Instance View im Camunda Cockpit [Ca20] oder WebVOWL [Lo16] zur Graph-basierten Dokumentation von Ontologien.

Für die technische Umsetzung in Web-Anwendungen kommen dabei folgende Technologien zum Einsatz: XML, insbesondere SVG eingebettet in HTML5 für die Serialisierung der Graphen selbst, CSS für diverse Style-Definitionen, JavaScript-Bibliotheken für die Generierung und das Event-Handling von Graph-Elementen (z. B. D3.js) eingebettet in ein modernes Entwicklungs-Framework (z. B. Vue.js). Eine idealtypische Kombination dieser Technologien findet sich in einer neueren Arbeit zur multi-disziplinären Design-Optimierung [Al19]. Dort wird der Ansatz zur Automatisierung einer Kette von Design-Tools kritisch diskutiert und nachgewiesen, dass der Aufwand zur Vollautomatisierung zwischen 60-80% der gesamten Projektzeit in Anspruch nimmt. Erschwerend kommt hinzu, dass viele grafische Design-Tools (z. B. CmapTools [Cm20], Camunda Modeler [Ca20]) zwar eine Ausgabe in SVG unterstützen, diese jedoch weit hinter den Spezifikationsmöglichkeiten des Formats zurückbleiben. So werden z. T. Texte oder Kanten als Punktpfade serialisiert, keinerlei IDs für die Graph-Elemente übergeben etc. Letzteres ist für eine eindeutige Adressierung im Web besonders notwendig.

3.3 Graph-basierte Benutzerführung

Nach [Vo09, S. 126] ist die Mehrzahl der klassischen Web-Anwendungen hierarchisch organisiert, um den Nutzer*innen Zugang zu Inhalt und Funktionen der Anwendung bereitzustellen. Typische hierarchische Navigationsmuster sind horizontale und vertikale

Menüs, Listen und Baumstrukturen. Hier sind zwei semantische Ordnungsmuster erkennbar: (1) Differenzierung nach Kategorien und (2) Über- bzw. Unterordnung. Beide sind eher grob und geben nur eingeschränkt Auskunft über Zusammenhänge und Strukturen. Zugleich sind sie dadurch schnell zu erfassen und platzsparend. Letzteres ist insbesondere bei responsiven Webdesigns eine Kernanforderung (vgl. z. B. [Kr19]).

Diese Vorteile büßt eine hierarchische Navigation dann ein, wenn wesentliche semantische Charakteristika der betreffenden Anwendungsdomäne nicht oder nur unzureichend durch die Listen- oder Baumstruktur abbildbar sind. Das ist z. B. offensichtlich bei Anwendungen mit Geo-Lokation der Fall. Hier tritt die (Land-)karte als zentrales Navigationsfeature in den Mittelpunkt. Andere Beispiele finden sich z. B. im Geschäftsprozessmanagement, wenn Fachanwendungen über Prozesslandkarten navigierbar sind.

Aus der Kognitionsforschung ist bekannt, dass die Visualisierung von Verbindungen durch Graph-Kanten zu einer effektiveren Wahrnehmung struktureller Zusammenhänge führt [Wa13, S. 183]. Damit bietet Graph-basierte Navigation nicht nur einen semantisch reicheren Zugang zu Inhalten und Funktionen, sondern auch eine bessere Orientierung in der Domäne und respektive in der Anwendung.

Eine Reihe von Arbeiten beschäftigen sich mit der Gegenüberstellung des Nutzerverhaltens in Graph- und Hierarchie-geleiteten Informationssystemen [Sa16] sowie mit der Visualisierung von Graph-Strukturen generell [Po20a]. Zentrale Erkenntnisse sind hier, dass Graph-Visualisierungen die Orientierung in komplexen Wissensdomänen beschleunigen. Allerdings gehen diese positiven Effekte dann verloren, wenn die Anzahl der durch Kanten verbundenen Knoten zu groß wird.

4 Prototypenentwicklung

Um die Forschungsfragen beantworten zu können, musste der zu entwickelnde Prototyp eines Digitalen Modulkatalogs mit den gewünschten Fähigkeiten zur Datenintegration als vertikaler Prototyp über alle drei Schichten: Datenhaltung, Fachlogik und Benutzeroberfläche konzipiert werden. Bei der Entscheidung für einen „horizontalen“ Fokus des Prototyps fiel die Wahl auf folgende Fachanforderung, die in den Studiengängen zu einem hohen Arbeitsaufwand führt und für die in [MB18] eine mangelhafte Unterstützung durch bestehende IT-Systeme nachgewiesen wurde: „Lehrende sollen dedizierte Editierrechte für die von ihnen verantworteten Module erhalten.“ Als Durchstichlösung wurde zudem der PDF-Download von Modulbeschreibungen implementiert.

4.1 Passfähigkeit und Erweiterbarkeit des Graph-Schemas

In [MB18] wurde bereits ein auf Schema.org basierendes Graph-Schema für Digitale Modulkataloge entwickelt und begründet. [MHP19] berichtet u. a. von ersten Experimenten zur semi-

automatischen Population des Wissensgraphen. Als Schema-Referenz wurden die Modulbeschreibungen im Fachbereich Wirtschaft der Hochschule verwendet. Im Ergebnis eines Workshops mit allen Fachbereichen und zentralen Einrichtungen der Hochschule wurde eine Vielzahl weiterer Anforderungen erhoben und priorisiert. Daraus ergab sich die Notwendigkeit, das Graph-Schema auf Erweiterbarkeit in dreierlei Hinsicht zu prüfen: (i) systematische und klassifizierte Darstellung von Lernzielen statt einer einfachen Auflistung, (ii) Adaptation des semiautomatischen Populationsprozesses auf weitere Studiengänge, (iii) Ergänzung von Planungsgrößen für die Veranstaltungsplanung. Die Erkenntnisse daraus werden im Folgenden dargelegt.

Die kompetenzorientierte Definition von Lernzielen gehört zu den Kernpunkten der Bologna-Reform. Im initialen Schema wurden die Lernziele als ungeordnete, identifizierbare Liste (`schema:ItemList`) über die generische Relation `schema:about` codiert. Jedes einzelne Lernziel erschien ausschließlich als Literal und war somit nicht weiter spezfizierbar. Der Wertebereich der genutzten Property (`schema:itemListElement`) in Schema.org umfasst nicht nur die Datentypklasse `schema:Text`, sondern auch die Objektklasse `schema:ItemList`. Dank dieser Modifikation konnten die Lernziele als Entitäten codiert werden. Über `schema:position` werden jedem Lernziel eine Ordnungsnummer und über `schema:additionalType` die Klassifikation in Kompetenzarten (fachliche, soziale, personale) zugewiesen. Im Fall fachlicher Kompetenzen kann darüber hinaus eine Bloom'sche Taxonomiestufe (z. B. Verstehen) deklariert werden. Dieses partielle Refactoring des Schemas hatte keinen Einfluss auf den Rest der Datenbasis.

Die Modulbeschreibungen liegen als Texte mit schwach standardisierter Tabellenstruktur vor. Die Variationsbreite ist dabei selbst innerhalb eines einzigen Modulkatalogs so hoch, dass manuelle Vor- und Nachbereitungsschritte unabdingbar sind. Getestet wurden Populationsprozesse auf Basis von Python und Excel. Es hat sich gezeigt, dass die Toolchain für jeden Fachbereich und im Detail auch für jeden Modulkatalog etwas angepasst werden muss. Mit Hilfe von Python-Skripten können die Rohdaten geparst und vorstrukturiert werden. Die Transformation dieser tabellarischen Daten in RDF wurde in dieser ersten Forschungsphase mit Excel unter Verwendung komplexer Textfunktionen und Makros umgesetzt. Es ist geplant, bei der anstehenden Erweiterung auf den dritten Fachbereich der Hochschule mit OpenRefine [Op20] zu arbeiten und damit den Anteil manueller Aufgaben weiter zurückzudrängen. Alle verwendeten Skripte und Dokumente finden sich auf GitHub[2]. Die Erfahrung hat gezeigt, dass die Erweiterung der Datenbasis auch mit Schema-Anpassungen einher geht. Im aktuellen Fall betraf das insbesondere die Anbindung der Module an Studiengänge. Aus diesem Grund wurde die ohnehin etwas umständliche Codierung über `schema:AlignmentObject` zugunsten des flexibler nutzbaren `schema:PropertyValue` ersetzt. Die Änderungen an der bestehenden Datenbasis konnten mit Hilfe des SPARQL-Update-Protokolls umgesetzt werden.

Wie bereits erwähnt, ist ein Modul zum einen ein ideelles Konzept (`schema:Course`), zum anderen eine Veranstaltung in Raum und Zeit (`schema:CourseInstance`). Während die bisher betrachteten Refactoring-Schritte die konzeptionelle Seite betrafen,

[2] https://github.com/bmake/modcat-prototyp/wiki

betrifft Anforderung (iii) die Veranstaltungssicht. Änderungen an der Grundstruktur sind hierfür nicht erforderlich. Die notwendigen Ergänzungen beschränken sich auf quantitative Attribute sowie Relationen zu anderen Ressourcen (Räume, Personen). In einem aktuellen Konzeptworkshop wurde die Relevanz dieser Anforderung nochmals bestätigt.

In Summe zeigte sich das Graph-Schema als stabil in seiner Grundstruktur und mit vertretbarem Aufwand erweiterbar. Konkret wurden folgende Ergebnisse erzielt: Die Datenbasis umfasst jetzt sieben vollständige Modulkataloge aus zwei von drei Fachbereichen. Lernziele können nach Kompetenzkategorien differenziert und bei fachlichen Kompetenzen einer Taxonomiestufe nach Bloom zugeordnet werden. Die Datenbasis kann für weitere Planungstools des Studiengangmanagements, wie Lehrkapazitätsplanung und Veranstaltungsplanung, im Verbund genutzt werden.

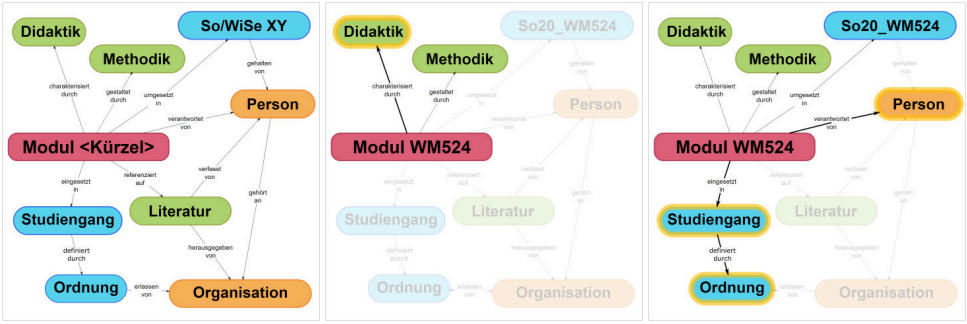
4.2 Implementierung der Graph-basierten Navigation

Die Fachanwendung selbst stellt eine komplette Neuentwicklung dar. Bisherige Vorstufen (vgl. [MHP19]) wurden mit Hilfe des Site-Generators Jekyll-RDF als rein statische Seiten implementiert. Da den Nutzer*innen ein unmittelbares Eingabefeedback und zugleich optimale Orientierung und Eingabesicherheit im Kontext komplexer Daten bereitgestellt werden sollten, fiel die Wahl auf folgende Front-End-Technologien: (i) die auf XML basierende Spezifikation des W3C zur Beschreibung zweidimensionaler Vektorgrafiken SVG in Verbindung mit D3.js, einer JavaScript-Bibliothek zur Erstellung dynamischer, interaktiver Datenvisualisierungen in Webbrowsern sowie (ii) das clientseitige JavaScript-Framework Vuetify mit vorgefertigten Oberflächenkomponenten[3].

Beim Entwurf des Navigationsgraphen war zunächst zu klären, welche Knoten und Kanten auszuwählen sind, um einen Kompromiss zwischen Orientierung in der Wissensdomäne und Übersichtlichkeit zu finden. Im Ergebnis wurden neun Knoten und zwölf Kanten ausgewählt (s. Abb.1a). Zwei der Knoten (Didaktik und Methodik) gruppieren mehrere Entitätstypen, sodass der Navigationsgraph zwölf Entitätsklassen und 15 Relationstypen repräsentiert. Bei der Farbcodierung wurden die folgenden Entscheidungen getroffen: der zentrale Knoten (Modul) hebt sich durch rote Färbung ab. Entitätsdaten, die von Modulverantwortlichen editiert werden können, sind grün eingefärbt. Dazu gehören Didaktik, Methodik und Literatur. Daten, die nur von Studiengangleitungen editiert werden können, sind blau eingefärbt. Für die Orientierung notwendige, in dieser Anwendung jedoch nicht editierbare Entitätstypen (Knoten) werden orange dargestellt.

Das Grundmodell für den Navigationsgraphen wurde mit dem Open-Source-Tool Inkscape erstellt, das ein hochwertiges SVG-Modell erzeugt und zudem spezifische Anpassungen über einen integrierten XML-Editor erlaubt. So sind alle Graph-Elemente mit sprechenden IDs versehen, Texte und grafische Objekte sind als solche spezifiziert. In Abhängigkeit von den Rechten und Interaktionen der Nutzer*innen weisen die Graph-Elemente orientierungsfördernde

[3] https://github.com/AKSW/jekyll-rdf, https://www.w3.org/XML/, https://d3js.org/, https://vuetifyjs.com/



(a) Basis Sicht (b) Modul-Verantwortlicher (c) Studiengang-Leitung

Abb. 1: Sichten des Navigationsgraphen

Effekte auf: Ghosting, Highlighting, Schatten (vgl. Abb.1b und Abb.1c). Diese Effekte werden durch den Einsatz der JavaScript-Bibliothek D3.js unterstützt. Der Navigationsgraph ist als eigenständige Vue-Komponente implementiert (`SVGGraph.vue`) und interagiert mit der übergeordneten Komponente `Starter.vue` und der benachbarten `FormBasisDaten.vue`. Die Auswahl des zu editierenden Moduls wird durch die Komponente `Select.vue` realisiert. `MainFooter.vue` beinhaltet Copyright-Informationen. Abb.2 zeigt das User Interface im Ganzen. Alle Ressourcen sind ebenfalls in GitHub[4] zugänglich.

Im Ergebnis bietet der Prototyp den Nutzer*innen für jedes zu editierende Modul eine auf ihre Bearbeitungsrechte angepasste Sicht auf die Graph-Navigation, die den Editierprozess konsistent unterstützt. Vue.js-Funktionen sorgen dafür, dass alle Eingaben während einer Sitzung auf dem Client zwischengespeichert werden. Der Bearbeitungsstand kann jederzeit in die Datenbank gespeichert werden. Das Rollen-Rechte-Konzept ist vorerst nur simuliert, einen Freigabeworkflow gibt es noch nicht.

5 Evaluation

Zur Ermittlung der Akzeptanz der neuen Fachanwendung für Digitale Modulkataloge durch die Nutzer*innen – in diesem Fall Lehrende verschiedener Studiengänge im Fachbereich Wirtschaft der Technischen Hochschule Brandenburg – wurde ein Softwaretest erarbeitet, durchgeführt und ausgewertet. Der Prototyp befindet sich in einer frühen Entwicklungsphase, daher hat die Evaluation durch Nutzer*innen einen großen Stellenwert für die Entwicklung der Fachanwendung im Hinblick auf User Interface (UI) Designs, Funktionalitäten und dahinter liegenden Technologien.

[4] https://github.com/bmake/modcat-prototyp

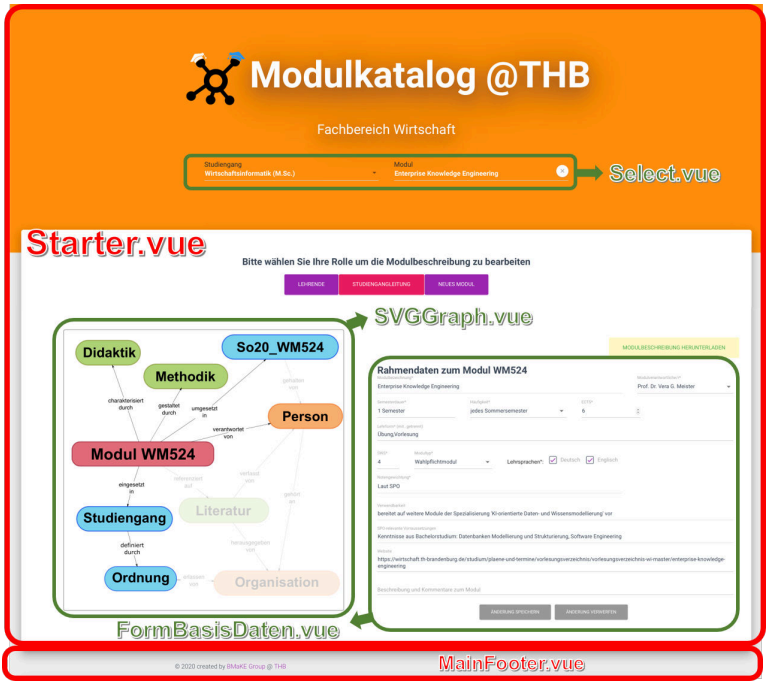


Abb. 2: User Interface mit Hervorhebungen zur Kennzeichnung der Vue-Komponenten

5.1 Untersuchungsdesign und -durchführung

Die Auswahl der Evaluationsmethoden [SB11] war von folgenden zielgebenden und begrenzenden Anforderungen geleitet: (i) Erkenntnisse zum Interaktionsfluss sammeln; (ii) neben positiven Aspekten auch potenzielle Nutzungsprobleme und Ideen zur Erweiterung aufdecken; (iii) den Tester*innen fachlich einschlägige Nutzungsszenarien anbieten; (iv) die Evaluation wegen der COVID-19-Pandemie remote durchführen. Ziel (i) erfordert eine Beobachtung des Nutzerverhaltens. Dabei sollen quantitative Daten zur Dauer von Interaktionen und zu Mausbewegungen gewonnen werden. Um diese Daten beurteilen zu können, sind zwei Teilnehmergruppen zu formieren: Noviz*innen und Expert*innen. Ziel (ii) kann mit der Methode des Lauten Denkens während der Beobachtung ergänzt um ein semistrukturiertes Interview (Online-Fragebogen) erreicht werden. Um Anforderung (iii) zu adressieren, wurden für jede/n Tester*in individuelle, fachlich einschlägige Use Cases entwickelt. Alle Beobachtungsdaten wurden von einer Desktop-Aufzeichnungssoftware gesammelt. Dieser Ansatz findet auch in Vor-Ort-Tests vielfache Anwendung. Aus Anforderung (iv) folgte die Notwendigkeit, den Teilnehmer*innen mediengestützte Anleitungen und eine Datenschutz-konforme Speicherumgebung bereitzustellen. Es wurde die frei nutzbare Software Active Presenter empfohlen und eine Bedienungsanleitung erstellt. Für den Upload der Aufzeichnungsvideos wurde die Lernplattform Moodle genutzt.

Es konnten 15 Lehrende des Fachbereichs Wirtschaft für die Teilnahme am Test gewonnen werden, welche die Gruppe der Noviz*innen bildeten. Als Expert*innen fungierten die fünf Mitglieder des Entwicklungsteams. Die Use Cases bildeten typische Aufgaben beim Editieren einer Modulbeschreibung ab: Lernergebnisse anpassen und spezifizieren, Inhaltselemente ändern, Lehr- und Lernmethoden strukturieren, weitere Daten ergänzen oder anpassen, wie z. B. Lehrsprachen, Prüfungsvoraussetzungen oder Kommentare. Dabei wurden keinen konkreten Änderungen vorgegeben, sondern auf die jeweilige Fachkompetenz der Tester*innen gesetzt, indem jede/r Editieraufgaben in einem ihr/ihm vertrauten Modul vornehmen sollte. Damit konnte der Test einer realen Nutzung der Software weitgehend angenähert werden. Die Teilnehmer*innen wurden Mitte April 2020 mit einer E-Mail zur Teilnahme aufgefordert. Mit der Mail wurden alle notwendigen Anweisungen, Links und Dokumente versandt. Wegen der Remote-Bedingungen musste der Testzeitraum über 4 Wochen ausgedehnt werden.

5.2 Auswertung der Untersuchungsergebnisse

Da die Erhebung der Rohdaten aus den übermittelten Aufzeichnungen und dem Fragebogen subjektiven Interpretationen unterliegt – das betrifft auch die quantitativen Daten – wurde sie durch zwei unabhängige Personen durchgeführt und im Anschluss konsolidiert. Die qualitativen Daten wurden einer thematischen Analyse nach [CBH15] unterzogen. Dafür wurden die Daten nach Themenbereichen gruppiert und nach konsistenten Beobachtungen (von mehr als 3 Teilnehmern bestätigt) und Besonderheiten strukturiert. Schließlich erfolgte eine Differenzierung in positive (Highlights) und negative (Probleme) Beobachtungen. Die aggregierten und konsolidierten Ergebnisse mit Fokus auf Forschungsfrage 3) zeigen Tab.1 (quantitative Daten) und Tab.2 (qualitative Daten).

Tab. 1: Quantitative Beobachtungsdaten zur Graph-Navigation bei Ausführung des Use Cases

Navigation zum Bereich (Knoten)	Zeit in sec.						Anzahl Klicks						Anzahl Scrollen					
	Experten			Novizen			Experten			Novizen			Experten			Novizen		
	$\overline{X}$	$\widetilde{X}$	σ	$\overline{X}$	$\widetilde{X}$	σ	$\overline{X}$	$\widetilde{X}$	σ	$\overline{X}$	$\widetilde{X}$	σ	$\overline{X}$	$\widetilde{X}$	σ	$\overline{X}$	$\widetilde{X}$	σ
Didaktik	4	5	2	31	13	40	1,0	1,0	0,0	2,9	1,0	6,3	0,0	0,0	0,0	2,6	1,0	3,4
Methodik	2	2	1	10	7	10	1,0	1,0	0,0	1,4	1,0	1,4	0,2	0,0	0,4	1,8	1,0	2,4
Rahmendaten	3	3	1	16	8	15	1,0	1,0	0,0	2,4	1,0	3,0	0,0	0,0	0,0	1,7	2,0	1,1

Die deutlichen Unterschiede zwischen der Bearbeitungszeit sowie der Anzahl von Klicks und Scrollen zwischen Noviz*innen und Expert*innen in Tab.1 weisen darauf hin, dass eine Graph-Navigation für die erste Gruppe eher ungewohnt ist. Auffällig ist, dass bei fast allen Items die Medianwerte der Noviz*innen sich wesentlich weniger von denen der Expert*innen unterscheiden als die Mittelwerte. Das lässt darauf schließen, dass nur wenige der Noviz*innen für diese Differenzen verantwortlich sind. Die Daten in Tab.2 machen deutlich, dass während der Graph-Navigation im Prototypen mehr Probleme als Highlights wahrgenommen wurden. Die ersten vier Probleme in Bereich konsistente Beobachtung lassen wiederum auf mangelnde Gewohnheit im Umgang schließen. Es werden Vorschläge in Richtung einer klassischen hierarchischen Navigation unterbreitet. Die anderen problematischen Beobachtungen stellen Verbesserungsvorschläge im Rahmen der Graph-Navigation dar. Einige davon wurden bereits

Tab. 2: Qualitative Beobachtungsdaten mit Fokus auf das Thema Graph-Navigation

Konsistente Beobachtung	Besonderheiten
Highlights	
-	- hilfreich, übersichtlich, optisch attraktiv - passende Anordnung
Probleme	
- nicht als Navigation erkannt - Abfolge unklar - zu viele Informationen - Drop-Down-Menüs verwenden - Zoomfunktion dysfunktional - Mouse-Over-Effekte nutzen	 - irreführendes Farbschema - maximal 3 Farben, mehr Kontrast

implementiert. In Vorbereitung der Evaluation gab es keinerlei Einführung oder Hinweise zum Thema Graph-Navigation. Es muss also konstatiert werden, dass auch in dieser Situation die Graph-Navigation nicht vollständig abgelehnt wurde. Allerdings sollten bei Einführung des Tools in den Regelbetrieb begleitende Einführungen und Tipps bereitgestellt werden. Um die

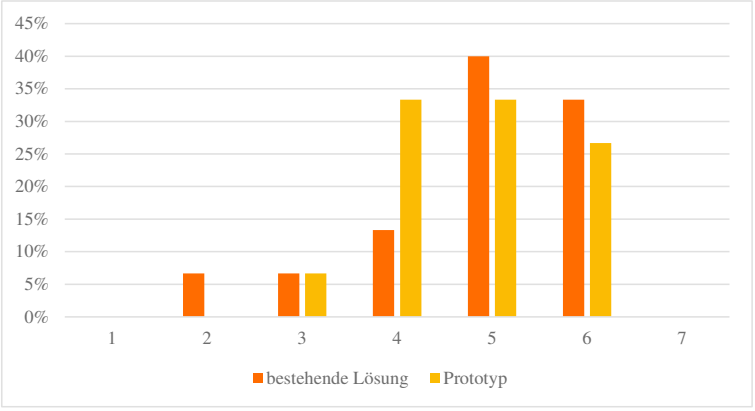


Abb. 3: Beurteilung der Zufriedenheit mit der bestehenden Lösung und dem getesteten Prototyen, N=15, 1 = sehr unzufrieden, 7 = außerordentlich zufrieden

Frage nach der Auswirkung der Graph-Navigation auf die Nutzerakzeptanz umfassend zu beantworten, sollen abschließend zwei strukturierte Items aus dem Online-Interview dargestellt werden. Wie Abb.3 zeigt, ist die Mehrheit der 15 Tester*innen zwar mit der bestehenden Lösung zur Bearbeitung von Modulkatalogen zufrieden, allerdings betrachteten sie zugleich den Prototypen als leistungsfähige Alternative. Die Bewertung sollte jeweils auf einer siebenstufigen Skala angegeben werden. Zusammengefasst kann somit festgestellt werden, dass eine Graph-Navigation als robuste Alternative zu herkömmlichen Navigationsmustern anzusehen ist.

6 Fazit und Ausblick

Aufbauend auf den Vorarbeiten [MB18; MHP19] stellt der hier präsentierte Prototyp einen wichtigen Schritt zu einer integrierten Fachanwendung für Digitale Modulkataloge dar. Die Graph-basierte Datenhaltung hat ihre Flexibilität und Erweiterbarkeit entlang anspruchsvoller Anforderungen bewiesen. Es ist gelungen, eine adaptive Graph-Navigation mit modernen Web-Technologien umzusetzen, die von den Tester*innen des Prototyps akzeptiert wurde. Im Rahmen der Evaluation sowie weiterer Meetings mit Stakeholdern wurden Anforderungen zu einem aufbauenden Entwicklungskonzept zusammengetragen. Neben der noch zu erweiternden Editierfunktion sollen das Browsen und die Dokumentation individueller Modulkataloge als Basisfunktionen implementiert werden. Für die Verknüpfung mit weiteren Fachanwendungen werden alternative Ansätze diskutiert. Eine leichtgewichtige Lösung könnte die Konfiguration einer REST API sein, auf die andere Anwendungen Zugriff haben. Aufwändiger und anspruchsvoller wäre eine Data-Hub-Lösung auf Basis hochwertiger Datenbanktechnologie. Die Kooperation mit anderen Hochschulen auf Open-Source-Basis wird angestrebt.

Literatur

[Al19] Alobaid, A.; Garijo, D.; Poveda-Villalón, M.; Santana-Pérez, I.; Fernández-Izquierdo, A.; Corcho, Ó.: Automating ontology engineering support activities with OnToology. J. Web Semant. 57/100472, 2019.

[ARM16] Arndt, N.; Radtke, N.; Martin, M.: Distributed Collaboration on RDF Datasets Using Git: Towards the Quit Store. In: SEMANTiCS 12th. S. 25–32, 2016.

[Br19] Brockmann, H.-C.: Wie Fair Data über den Erfolg der Digitalen Transformation entscheidet, 2019, URL: it-daily.net, Stand: 27. 11. 2019.

[Ca20] Camunda, 2020, URL: camunda.com/products, Stand: 01. 09. 2020.

[CBH15] Clarke, V.; Braun, V.; Hayfield, N.: Thematic Analysis. In: Qualitative Psychology: A Practical Guide to Research Methods. Sage, 2015.

[Ce19] Cerans, K.; Ovcinnikova, J.; Liepins, R.; Grasmanis, M.: Extensible Visualizations of Ontologies in OWLGrEd. In. Bd. 11762. LNCS, S. 191–196, 2019.

[Cm20] CmapTools: Institute for Human & Machine Cognition, 2020, URL: cmap.ihmc.us, Stand: 01. 09. 2020.

[Da20] Data.world, 2020, URL: gra.fo, Stand: 01. 09. 2020.

[Do20] DotNetRDF, 2020, URL: dotnetrdf.org, Stand: 01. 09. 2020.

[FAM16] Frischmuth, P.; Arndt, N.; Martin, M.: OntoWiki 1.0: 10 Years of Development – What's New in OntoWiki. In: SEMANTiCS 12th. Bd. 1695. CEUR, 2016.

[Ga05] Gangemi, A.: Ontology Design Patterns for Semantic Web Content. In: ISWC 5th. Bd. 3729. LNCS, S. 262–276, 2005.

[Hi16] Hitzler, P.; Gangemi, A.; Janowicz, K.; Krisnadhi, A.; Presutti, V.: Ontology Engineering with Ontology Design Patterns – Foundations and Applications. IOS Press, 2016.

[Kr19] Krijn, M.: Responsives Webdesign: 9 Do's and Don'ts bei der Navigation, 2019, URL: blog.amplexor.com, Stand: 15. 11. 2019.

[Kr20] Krupa, K.: Data Hub Guide for Architects. Marklogic, 2020.

[Ku10] Kultusministerkonferenz: Ländergemeinsame Strukturvorgaben für die Akkreditierung von Bachelor und Masterstudiengängen – Beschluss i. d. F. vom 04.02.2010, 2010.

[Lo16] Lohmann, S.; Negru, S.; Haag, F.; Ertl, T.: Visualizing ontologies with VOWL. Semantic Web 7/4, S. 399–419, 2016.

[MB18] Meister, V. G.; Becker, J.: Konzept und vergleichende Analyse eines Wissensgraph-basierten Modulkatalogs. In: LAiW co-located zur INFORMATIK. S. 14–28, 2018.

[MHP19] Meister, V. G.; Hu, W.; Pottenstein, P.: Wissensgraph-basierter Modulkatalog als Schnittstelle zwischen digitaler Lehre und digitalem Campusmanagement. In: Hochschulen in Zeiten der Digitalisierung. Springer Vieweg, S. 89–105, 2019.

[Mu15] Musen, M. A.: The protégé project: a look back and a look forward. AI Matters 1/4, S. 4–12, 2015.

[OD20] ODP: Semantic Web Portal to Ontology Design Patterns (ODPs), 2020, URL: ontologydesignpatterns.org, Stand: 01. 09. 2020.

[Op20] OpenRefine, 2020, URL: openrefine.org, Stand: 01. 09. 2020.

[Po20a] Po, L.; Bikakis, N.; Desimoni, F.; Papastefanatos, G.: Linked Data Visualization: Techniques, Tools, and Big Data. Morgan & Claypool, 2020.

[Po20b] Poolparty, 2020, URL: poolparty.biz, Stand: 01. 09. 2020.

[Sa16] Sarrafzadeh, B.; Vtyurina, A.; Lank, E.; Vechtomova, O.: Knowledge Graphs versus Hierarchies: An Analysis of User Behaviours and Perspectives in Information Seeking. In: CHIIR. S. 91–100, 2016.

[SB11] Sarodnick, F.; Brau, H.: Methoden der Usability Evaluation – Wissenschaftliche Grundlagen und praktische Anwendungen. Hans Huber, Bern, 2011.

[Su12] Suárez-Figueroa, M. C.; Gómez-Pérez, A.; Motta, E.; Gangemi, A.: Ontology Engineering in a Networked World. Springer, 2012.

[To20] TopQuadrant, 2020, URL: topquadrant.com, Stand: 01. 09. 2020.

[Tu20] Tudorache, T.: Ontology engineering: Current state, challenges, and future directions. Semantic Web 11/1, S. 125–138, 2020.

[Vo09] Vora, P.: Web Application Design Patterns. Morgan Kaufmann, 2009.

[Wa13] Ware, C.: Information Visualization – Perception for Design. Morgan Kaufmann, 2013.

Videoproduktion: Entwicklung eines adaptiven Wegweisers für Hochschullehrende

Linda Blömer[1] Christin Voigt[1] Alexander Piwowar[2]

Abstract: Spätestens seit Aufkommen der Corona-Pandemie spielt der Einsatz von Videos im Rahmen der Hochschullehre eine zentrale Rolle. Dabei werden Lehrende, die erstmalig ein Video produzieren, mit zahlreichen Herausforderungen konfrontiert. Serviceeinrichtungen der Hochschulen sind gefragt, um Lehrende mittels didaktischer Beratung und technischer Expertise zu unterstützen. Doch die zeitlichen Kapazitäten aller Beteiligten sind knapp. Mit dem Ziel, die Videoproduktion an Hochschulen zu fördern, hat ein interdisziplinäres Team der Universität Osnabrück einen Design Thinking Prozess mit zwei Fokusgruppeninterviews durchgeführt, um einen adaptiven Wegweiser für Hochschullehrende zu entwickeln. Der Wegweiser wird zukünftig in einen digitalen Konfigurator überführt, der Lehrenden als erste Anlaufstelle, zentrale Informationsplattform und Hilfe zur Selbsthilfe dienen soll. Zudem soll er die hochschulinternen Serviceeinrichtungen entlasten, indem er eine erste Orientierungshilfe bietet und einen Output liefert, der als Ausgangspunkt für persönliche Beratungsgespräche genutzt werden kann. Die Grundstruktur des Wegweisers kann auf andere Hochschulen übertragen werden, um weitere, hochschulspezifische Konfiguratoren zu entwickeln.

Keywords: Design Thinking; Digitale Hochschullehre; Konfigurator; Videoproduktion

1 Einleitung

In den vergangenen Jahren hat der Einsatz von Videos in der universitären Lehre stetig an Bedeutung gewonnen [FHS19]. Das Aufkommen der Corona-Pandemie hat ihre Relevanz noch weiter gesteigert. Universitäten wie die TU Graz setzen auf den Einsatz von Videos, um zu Zeiten der Pandemie ihre bisherigen Präsenzveranstaltung online anbieten zu können, was dort zu einem plötzlichen Anstieg der Serviceanfragen an den Video- und Didaktik-Support führte [ESBE20]. Für eine umfangreiche Produktion von Videos sollten eine geeignete Infrastruktur und ausreichend Personal vorhanden sein. Für Lehrende, die bisher keine Erfahrung mit digitalen Medien im Rahmen der Hochschullehre gesammelt haben, muss zudem eine professionelle Weiterbildung sichergestellt werden [Di19]. Da an den meisten Universitäten eine kurzfristige Aufstockung und umfangreiche Weiterbildung von Personal nicht realisierbar erscheint, stellt sich die Frage, wie trotz begrenzter Ressourcen und zum Teil geringer Vorerfahrung der Lehrenden die Erstellung von Videos unterstützt werden kann. Bislang gibt es zu diesem Thema nur wenig Forschung und viele, unter anderem didaktische und technische Fragestellungen bleiben ungeklärt [SLV00]. Lehrende, die erstmalig zu Videoproduzenten werden, müssen deshalb zumeist selbst in Erfahrung bringen, welche

[1] Universität Osnabrück, IMU/ BOW, Katharinenstr. 3, Osnabrück, 49074, <vorname.nachname>@uos.de
[2] Universität Osnabrück, virtUOS, Heger-Tor-Wall 12, Osnabrück, 49074, <vorname.nachname>@uos.de

doi:10.18420/inf2020_44

Arten von Videos es gibt, welche davon im jeweiligen Fall didaktisch sinnvoll erscheinen und welche technischen Werkzeuge sich für die Umsetzung eignen. Dabei wird der Austausch mit Serviceeinrichtungen der Universität möglicherweise dadurch erschwert, dass Begriffe, Methoden oder Tools bislang wechselseitig unbekannt sind oder uneinheitlich verwendet werden. Um Lehrenden zukünftig den Einstieg in die Videoproduktion zu erleichtern und den Austausch zwischen ihnen und den Serviceeinrichtungen der Hochschule reibungsloser und effizienter zu gestalten, adressiert die vorliegende Arbeit die skizzierte Forschungslücke. Zu diesem Zweck soll zu Beginn ermittelt werden, vor welchen Herausforderungen Lehrende im Rahmen der ersten Videoproduktion standen und welche Form von Unterstützung sie sich gewünscht hätten. Anhand der Ergebnisse soll ein geeignetes Instrument entwickelt werden, welches Lehrende im Rahmen der Videoproduktion unterstützt und gleichzeitig die Zusammenarbeit zwischen Lehrenden sowie Didaktik- und Video-Support begünstigt. Langfristig soll dadurch die Entstehung digitaler Lehrinhalte in Form von Videos an Hochschulen gefördert werden.

Die Forschungsfragen (FF) der Arbeit lauten deshalb wie folgt:

FF 1: *Über welche Herausforderungen während der ersten Videoproduktion berichten Hochschullehrende und welche Form von Unterstützung hätten sie sich gewünscht?*

FF 2: *Wie kann ein Prototyp zur Unterstützung von Hochschullehrenden im Rahmen der Videoproduktion basierend auf den Ergebnissen aus FF1 aussehen?*

FF 3: *Wie bewerten Lehrende ohne Vorerfahrung bezüglich Videoproduktion den Prototyp aus FF 2 und inwieweit sollte dieser angepasst werden?*

Das methodische Vorgehen dieser Arbeit wird in Kapitel 2 erläutert. Es basiert auf einem Design Thinking Prozess, in den zwei Fokusgruppeninterviews integriert wurden. In Kapitel 3 werden die Ergebnisse vorgestellt, die im Rahmen der einzelnen Design Thinking Stufen ermittelt werden konnten. Die Arbeit endet mit einem Fazit in Kapitel 4, in dem die Grenzen der Arbeit, zukünftige Forschungsfelder sowie Anwendungs-möglichkeiten des Wegweisers thematisiert werden.

2 Methodisches Vorgehen

Design Thinking lässt sich auf die Methode Design Science zurückführen, die in den 60er- und 70er-Jahren an Bekanntheit gewann [SSHF19]. Beide Ansätze werden dem übergeordneten Design-Grundsatz zugeordnet und verfolgen das Ziel, Relevanz für die Praxis zu schaffen. Doch während dieses Ziels im Design Science durch das Zusammentragen von vermeintlich objektivem, allgemeingültigem Wissen und der daraus resultierenden Entstehung von Artefakten erreicht werden soll, setzt die Methode des Design Thinking auf Wissen, welches aus dem Kontext oder dem Denker selbst resultiert. Dabei spielen unter anderem Kreativität und Einstellungen eine wichtige Rolle. Beim Design Thinking geht es demnach primär um den Denkprozess und weniger um die rein wissenschaftliche Behandlung

von Wissen [DR13]. Die Methode wird aktuell zur kreativen Lösung komplexer Probleme in unterschiedlichen Organisationen angewandt, darunter auch im Bildungsbereich [Ba18]. Sie eignet sich in besonderem Maße im Rahmen der Entwicklung menschenzentrierter Problemlösungen, bei denen Verständnis und Empathie gegenüber den Interessengruppen sowie deren Belangen im Vordergrund der Entwicklung steht [DR13]. Im vorliegenden Fall soll primär eine Lösung für Hochschullehrende entwickelt werden, um ihnen die Produktion von Videos zu erleichtern. Doch zu den Interessengruppen zählen auch hochschulinterne Serviceeinrichtungen, die zum Teil maßgeblich an der Entwicklung von Videos beteiligt sind. Um eine langfristige Lösung zu erhalten, die von allen Interessengruppen unterstützt wird, sollten alle betroffenen Personen an der Problemlösung beteiligt werden. Im vorliegenden Fall erscheint die Methode des Design Thinking als besonders geeignet, da Wissenschaftler, Didaktiker und Videoproduzenten mit spezifischem Fachwissen von der Problematik betroffen sind. Durch die Methode des Design Thinking können sie neben wissenschaftlichem Wissen insbesondere von kreativem Denken, gegenseitigem Verständnis und Empathie profitieren.

Design Thinking ist vielfältig interpretier- und anwendbar, was das Vorhandensein verschiedener Definitionen und unterschiedlichster Prozessmodelle verdeutlicht [SSHF19]. Ein Design Thinking Prozess besteht zumeist aus 3-5 Phasen [SSHF19] und wird häufig von interdisziplinären Teams angeleitet [SF13]. Als bekanntes Design Thinking Modell wird vielfach das 5-Stufen-Modell der Standford d.school erwähnt [DR13, SSHF19], welches aus den Phasen Emphazise, Define, Ideate, Prototype und Test besteht [DHKS18]. Die Phasen repräsentieren unterschiedliche Vorgehensweisen, die nicht zwangsläufig sequenziell sondern häufig iterativ durchgeführt werden [DR13]. Im vorliegenden Fall wurde ein fünfköpfiges, interdisziplinäres Team gebildet, bestehend aus zwei wissenschaftlichen Mitarbeiterinnen, die an der Hochschullehre beteiligt sind, einem Mitarbeiter des Video-Supports, der die Videoproduktion an der Universität bei Bedarf begleitet und durchführt sowie zwei Mitarbeitern des Zentrums für Digitale Lehre, Campus Management und Hochschuldidaktik, die Lehrende an der Universität Osnabrück (UOS) beratend unterstützen. Das schrittweise Vorgehen des Design Thinking Prozesses orientiert sich am 5-Stufen-Modell der Stanford d.school [DHKS18], welches aufgrund seiner praxisorientieren Anwendbarkeit ausgewählt wurde. Da die Stanford d.school jede Stufe sowie das Vorgehen je Stufe prägnant und anschaulich beschreibt, konnten sich alle Teammitglieder unabhängig ihres methodischen Vorwissens schnell und problemlos in den Prozess einfinden. Die Stufen des Modells werden in Abbildung 1 dargestellt. Darüber hinaus wird in Abbildung 1 deutlich, in welchen Stufen Interviews durchgeführt und Forschungsfragen beantwortet werden. Anhand des in Stufe 1 durchgeführten Fokusgruppeninterviews wird das Problem genauer definiert. Daraufhin werden Ideen zur Problemlösung entwickelt, die in der Erstellung von Prototypen münden. Durch ein weiteres Fokusgruppeninterview wird der Prototyp abschließend in Stufe 5 getestet und weiter angepasst.

Um eine gemeinsame Ausgangslage zu schaffen und das vorhandene Problem zu skizzieren, traf sich das Forschungsteam im Dezember 2019 für einen ersten Erfahrungsaustausch.

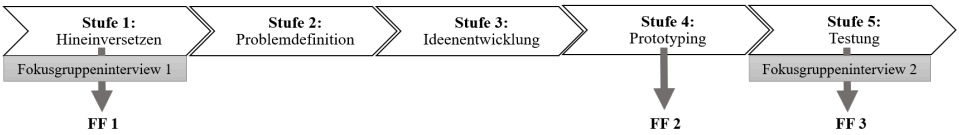


Abb. 1: Methodisches Vorgehen entlang des Design Thinking Prozesses [DHKS18]

Dabei wurde festgelegt, welche Aspekte evaluiert werden sollen und wie methodisch vorgegangen wird. Im Januar 2020 fand im Rahmen von Stufe 1 das erste Fokusgruppeninterview statt, an dem sechs Lehrende beziehungsweise Teammitglieder von Lehrenden mit Videoproduktionserfahrung teilnahmen. Im Anschluss an die Durchführung wurde das Interview einer qualitativen Inhaltsanalyse nach Mayring [Ma07] unterzogen. Im Zuge der Auswertung wurde induktiv vorgegangen, sodass sich die Kategorien aus den Aussagen der Teilnehmenden ergaben. Im ersten Schritt „Paraphrasieren“ wurden zunächst alle Aussagen gemäß der Leitfragen in relevante Abschnitte unterteilt. Beispielsweise wurden Abschweifungen sowie alle nicht zielführenden Textpassagen entfernt. Im zweiten Schritt „Generalisierung“ wurde der Kern dieser Aussagen in wenigen Worten zusammengefasst. Anschließend erfolgte die „Reduktion und Subsumtion“ im dritten Schritt, indem aus allen generalisierten Aussagen Kategorien gebildet wurden [Ma88]. Die daraus ermittelten Ergebnisse beantworten FF 1 und bilden die Grundlage von Stufe 2, 3 und 4 des Design Thinking Prozesses. Im März 2020 fand sich das Forschungsteam zu einem weiteren Treffen zusammen, um das Problem genauer zu definieren (Stufe 2) und erste Ideen für die Problemlösung zu sammeln (Stufe 3). In Stufe 4 entwickelte das Team einen Prototyp, wodurch FF2 beantwortet wurde. Die Entwicklung des Prototyps fand zwischen Juni und Juli 2020 aufgrund der Corona-Pandemie im virtuellen Raum statt. Dabei wurde an drei aufeinander folgenden Terminen mit Hilfe eines Online-Whiteboards gemeinsam am Prototyp gearbeitet. Die Testung des Prototyps (Schritt 5) erfolgte im August 2020 mithilfe eines virtuellen Fokusgruppeninterview, in dem fünf Lehrende beziehungsweise Teammitglieder von Lehrenden ohne Videoproduktionserfahrung Hinweise zum Prototyp gaben, die in eine anschließende Überarbeitung einflossen.

3 Ergebnisse des Design Thinking Prozesses

Das Forschungsteam traf initial im Rahmen eines Drittmittelprojektes zusammen, dessen Ziel es unter anderem war, digitale Lehrvideos für verschiedene Veranstaltungen zu erstellen. Im Rahmen ihrer Tätigkeit standen die Beteiligten vor unterschiedlichen Herausforderungen. Gleichzeitig betrachteten sie die Aufgabe aus unterschiedlichen Blickwinkeln. Während beispielsweise der Video-Support hauptsächlich eine technische Sichtweise einnahm, standen für die wissenschaftlichen Mitarbeiter primär didaktische Fragestellungen im Vordergrund. Auch eine einheitliche Definition der verschiedenen Videoarten konnte nur schwer gefunden wurden. Gliedert man diese nach den technischen Werkzeugen oder didaktischen Möglichkeiten? Betrachtete man das komplette Video oder einzelne Elemente? Für die erfolgreiche

Zusammenarbeit im Team war daher das Zusammenbringen aller Perspektiven und das Finden einer gemeinsamen Sprache notwendig. Darüber hinaus stellten die wissenschaftlichen Mitarbeiter fest, dass sie im Rahmen einer Internetrecherche zwar viele unterschiedliche Beispiele für Lehrvideos finden konnten, sich ihnen die spezifischen Möglichkeiten der Videoproduktion an der eigenen Universität allerdings erst nach mehreren Gesprächen mit unterschiedlichen Ansprechpartnern des Didaktik- und Video-Supports erschlossen. Der Didaktik- und Video-Support berichtete hingegen, dass sich Lehrende in ersten Beratungsgesprächen zum Teil mit sehr unkonkreten Vorstellungen über die Video-produktion an sie wandten. In einigen Fällen stellten Lehrende erst während des Beratungsprozesses fest, dass sie eine andere didaktische Möglichkeit als die Videoproduktion vorziehen. Alle Beteiligten beklagten daher einen zu hohen Zeitaufwand und ein ineffizientes Verfahren. Da zudem an der Universität Osnabrück ein steigender Bedarf an digitaler Lehre verzeichnet wurde, wünschten sich die Beteiligten eine effizientere Gestaltung der Informationswege und Beratungsgespräche. Hieraus entstand die Idee, die Videoproduktion sowohl für Lehrende an der Universität Osnabrück als auch für den Didaktik- und Video-Support durch einen Konfigurator zu erleichtern. Die Ergebnisse wurden entlang der fünf Stufen des Design Science Prozesses erhoben, die nacheinander aber teilweise iterativ stattfanden. So wurden beispielweise die in Stufe 3 entwickelten Ideen während der Erstellung des Prototyps in Schritt 4 hinterfragt und bei Bedarf angepasst.

3.1 Stufe 1: Hineinversetzen

Um sich in Hochschullehrende hineinzuversetzen und Herausforderungen und Unterstützungsmöglichkeiten im Rahmen der Videoproduktion an der Universität Osnabrück zu identifizieren, wurde im Dezember 2019 ein Fokusgruppeninterview durchgeführt. Zu den Teilnehmern zählten Lehrende und deren Teammitglieder, die bereits Erfahrung mit der Erstellung universitärer Videos gesammelt haben. Um diesbezüglich eine einheitliche Terminologie zu nutzen, wurde die Videoproduktion in Anlehnung an Klimsa in die drei Phasen *Vorproduktion*, (*mediale) Produktion* und *Postproduktion* [Kl06] unterteilt. Die *Vorproduktion* beinhaltet alle planenden, strukturierenden Aufgaben wie beispielsweise das Abwägen technischer und didaktischer Möglichkeiten sowie das Erstellen eines Storyboards. Während der *medialen Produktion* wird das Bild- und Tonmaterial produziert und die *Postproduktion* umfasst alle nachträglichen Arbeiten wie Schnitt und Nachbearbeitung. Als Voraussetzung für die Teilnahme am Fokusgruppeninterview mussten die Teilnehmer an mindestens einer dieser drei Phasen aktiv mitgewirkt haben. Dieser Fokus wurde bewusst gelegt, um bereits entstandene und nicht prognostizierte Herausforderungen zu erheben. Da der Konfigurator in erster Linie unerfahrene Lehrende während der erstmaligen Videoproduktion unterstützen soll, wurde im Fokusgruppeninterview explizit nach der gesammelten Erfahrung während der ersten durchgeführten Videoproduktion an einer Universität gefragt. Das Fokusgruppeninterview wurde mit zwei Professoren, drei wissenschaftlichen Mitarbeitern und einer studentischen Hilfskraft durchgeführt. Es startete mit einer Vorstellungsrunde aller Beteiligten, woraufhin eine quantitative Befragung zur Beschreibung der Stichprobe

durchgeführt wurde. Vier der Befragten hatten zum Zeitpunkt des Interviews erst eine Videoproduktion innerhalb der universitären Lehre durchgeführt. Davon verfügte jedoch eine Person aufgrund ihres außeruniversitären Werdegangs über umfangreiche Vorerfahrungen. Ein Befragter hatte zum Zeitpunkt des Interviews 16 Videos produziert, während eine weitere Person angab, bereits mehr als 180 universitäre Lehrvideos produziert zu haben. Abbildung 2 verdeutlicht die Selbsteinschätzungen aller Fokusgruppenteilnehmer/innen bezüglich ihrer Erfahrung in der Videoproduktion auf einer Skala von 1 (Unerfahren) bis 10 (Experte).

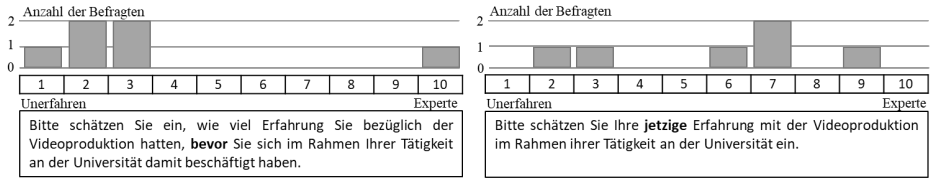


Abb. 2: Vorerfahrung der Teilnehmer des Fokusgruppeninterviews 1

Ausgenommen der Person mit beruflicher Vorerfahrung schätzten sich alle Befragten vor ihrer ersten universitären Videoproduktion als eher unerfahren ein. Deutlich wird zudem, dass sich zwei Befragte auch nach erster gesammelter Erfahrung als eher unerfahren einschätzen. Während ihrer ersten universitären Videoproduktion erhielten darüber hinaus fünf der Befragten Unterstützung durch den Video-Support. Zwei Befragte erhielten Unterstützung von weiteren wissenschaftlichen Mitarbeitern und ein Befragter wurde von einer studentischen Hilfskraft unterstützt. Nach der quantitativen Befragung wurde die qualitative Diskussionsrunde durchgeführt. Sie orientiert sich an zwei Leitfragen, die im folgenden Kapitel näher beschrieben werden. Die Dauer des Interviews betrug eine Stunde. Allerdings verblieb die Mehrheit der Befragten ca. 30 weitere Minuten und führte die anregende Diskussion auf freiwilliger Basis fort.

3.2 Stufe 2: Definieren des Problems

Die Auswertung des Fokusgruppeninterviews orientiert sich an den zwei Leitfragen „*Vor welchen Herausforderungen standen Sie während Ihrer ersten Videoproduktion?*“ und „*Welche Form von Unterstützung hätten Sie sich rückwirkend gewünscht?* Für jede Leitfrage wurden nach der oben beschriebenen Inhaltsanalyse nach Mayring Kategorien gebildet, die in Tabelle 1 und 2 dargestellt werden. Unvorhergesehene Komplikationen beim Dreh reichten dabei über benachbarte Bauarbeiten, einem knacksenden Boden bis hin zu technischen Störungen. Die Folge waren Frustrationen und der Versuch, sich an die Bedingungen anzupassen. Ein Unwohlsein vor der Kamera trat bei der Mehrheit der Befragten auf und wurde vor allem auf die veränderten Rahmenbedingungen zurückgeführt. So sei die Umgebung im Studio oder während der Aufnahme im Büro nicht dem Hörsaal zu vergleichen. Die fehlende Kommunikation mit den Studierenden während des Vortrags und der Wunsch nach Perfektionismus habe vielen Lehrenden die Videoproduktion erschwert.

Allerdings würde dieses Unwohlsein vor der Kamera nur bei der ersten Videoaufnahme auftreten. Ein Befragter sagte aus „*Irgendwann kommt man in den Flow [...]. Je mehr Videos man aufgenommen hat, desto einfacher wird das auch.*“ Darüber hinaus wurde von einem Befragten die Sinnhaftigkeit der Videos insgesamt hinterfragt und der Wunsch nach weiterer Forschung über die Auswirkungen von Lehrvideos auf den Lernerfolg und die Zufriedenheit der Studierenden geäußert.

Tab. 1: Ermittelte Herausforderungen beim Videodreh

Unterkategorien	Oberkategorien	Ankerbeispiel
Frustration beim Videodreh Störungen beim Dreh durch Geräusche Technische Störung bei der Aufnahme	**Unvorhergesehene Komplikationen beim Dreh**	*„Ich finde es ganz schön, wie es hier anklingt, wie Technik unser Verhalten [beim Videodreh] eigentlich beeinflusst.“*
Unwohlsein mit Mimik und Gestik Umgang mit Versprechern Perfektionismus beim Videodreh Multitasking vor der Kamera Fehlende Interaktion	**Unwohlsein vor der Kamera**	*„Ich kann mich nicht hören und ich kann mich nicht sehen. Ich bin Perfektionist. Solche Versprecher machen mich wahnsinnig.“*
Akzeptanz der Studierenden unklar Fehlender Mehrwert befürchtet	**Frage nach Sinnhaftigkeit von Videos**	*„Ich bin mir aber nicht sicher, ob die Studierenden das überhaupt wollen.“*
Fehlende finanzielle Mittel Mangelnde technische Ausstattung Fehlende Zeit Zu hoher Zeitaufwand	**Fehlende Ressourcen**	*„Denn dann hatten wir auf einmal festgestellt, dass die Zeit gar nicht reicht, die wir eingeplant hatten.“*
Didaktischer Aufbau Auswahl der Videoart Auswahl der Videolänge	**Didaktische Fragestellungen**	*„Wie macht es auch didaktisch Sinn? Wie baue ich dort ein ansprechendes artikulares Schema auf?“*
Fehlendes technisches Know How Rechtliche Fragen Fehlende Experten	**Fehlendes Expertenwissen**	*„Und ich habe dann auch irgendwann gemerkt, ich bin nicht vom Fach.“*
Aufwandseinschätzung kompliziert Planung im Vorfeld Begründung der Auswahl vor dem Vorgesetzten Fehlende Übersichten	**Fehlende Übersichten**	*„[...] weil ich mir viele Gedanken gemacht habe [und] weil man das Gefühl hatte, man muss so viele Dinge gleichzeitig beachten.“*

Als fehlende Ressource wurde hauptsächlich mangelnde Zeit genannt. Ein Befragter kritisierte auch die technische Ausstattung der Universität und demzufolge fehlende finanzielle Mittel. Im Vordergrund der Videoproduktion standen jedoch die didaktischen Fragestellungen. So

wurden mangelnde Erfahrungswerte als zentrale Herausforderung im Entscheidungsprozess genannt, *„weil es auch weniger darum geht, ein wirklich qualitativ hochwertiges Video zu produzieren, sondern es geht eher darum, den didaktischen Mehrwert auszuprobieren“*. Zuletzt bestand bei fast allen Teilnehmenden des Fokusgruppeninterviews die Herausforderung, eine aussagekräftige Checkliste und Regelwerke für die Videoproduktion zu finden. Eine erfahrenere Person berichtete: „*Wenn ich jetzt gerade im Zusammenhang mit Studierenden daran arbeite, dass da dann ein wahnsinniges Grundwissen erst einmal aufgebaut werden muss.* [...] *Was mir aber auch hier an der Uni deutlich auffällt, wenn E-Mails verschickt werden [...] Das ist dann so ein Waking up, so ein Unverständnis im Umgang mit digitalen Medien*.“ Die Ergebnisse der Leitfrage 2 werden in Tabelle 2 dargestellt.

Tab. 2: Ermittelte Unterstützungen beim Videodreh

Unterkategorien	Oberkategorien	Ankerbeispiel
Wissensweitergabe von Experten in den eigenen Reihen Bessere Tools bereitstellen Eigenständige Aufnahme ermöglichen	**Hilfe zur Selbsthilfe**	*„Ich finde da ist das mit dem One Button Studio jetzt schon eine tolle Entwicklung.“*
Mehr Expertenunterstützung Experten als Support bei der Aufnahme Unterstützung durch Experten aus den eigenen Reihen Support bei der Nachbearbeitung	**Personelle Unterstützung**	*„Wo gibt es denn die Kollegen mit einer fachlichen oder technischen Expertise?“*
Technische Ausstattung an der UOS Vorhandene Strukturen an der UOS Regelwerke und Checklisten Vor- und Nachteile der Formate Didaktische Fragestellungen unklar	**Leitfäden und Übersichten für die UOS**	*„Ich habe zum Beispiel Checklisten gefunden, [...] die bestanden hauptsächlich aus Fragen. Mich hätten die Antworten interessiert.“*

Nach der zweiten Leitfrage wurde im Gespräch zumeist zwischen gewünschter /erhaltener sowie gewünschter/ nicht erhaltener Unterstützung unterschieden. So wurde einerseits angegeben, dass die Universität zufriedenstellend Hilfe zur Selbsthilfe anbietet, anderseits universitätsspezifische Übersichten und Leitlinien während der Recherche in der Vorproduktion hilfreich gewesen wären. Gelobt wurde insbesondere das neu eingerichtete One Button Greenscreen Studio, das Lehrenden nach einer Einführung die selbstständige Aufnahme von Videos ermöglicht. Zudem berichtete die Mehrheit der Befragten, Unterstützung von Experten der Universität erhalten zu haben.

3.3 Stufe 3: Entwicklung von Ideen

Um erste Ideen für einen Wegweiser als Grundlage für den Konfigurator zu entwickeln, ist das Forschungsteam in einem dreistündigen Meeting zusammengekommen. Dabei

wurden die Ergebnisse aus dem Fokusgruppeninterview vorgestellt, diskutiert und erste Lösungsvorschläge aus den jeweiligen Blickwinkeln der verschiedenen Beteiligten entwickelt. Basierend auf dem Fokusgruppeninterview und auf den Erfahrungen des Forschungsteams wurden verschiedene Perspektiven eingenommen, woraufhin alle Ideen zusammengetragen und diskutiert wurden. Anschließend wurden die Ergebnisse in einen ersten Entwurf der Wegweiser-Struktur überführt. Sofern möglich sollten für jede der in Tabelle 1 und 2 genannten Kategorien Lösungsvorschläge entwickelt werden. Primär wurde die Unterteilung der Struktur aufgrund der Kernfragen: *„Was möchte ich mit dem Video erreichen?“* und *„Wie kann ich mein Ziel umsetzen?“* vorgenommen. Die erste Frage bezieht sich auf das allgemeine Ziel, den Zweck und die Zielgruppe, die mit dem Video erreicht werden soll. Die zweite Frage adressiert den spezifischen Plan zur Erreichung des definierten Ziels und deckt die konkrete Operationalisierung im Sinne von Videoarten und Video-Stilmitteln beziehungsweise Video-Bausteinen ab. Dazu wurde eine Unterteilung der Videoarten erarbeitet, die die praktische Erfahrung der Teammitglieder berücksichtigt. Schritt 2 bietet darüber hinaus einen Überblick über die technischen Möglichkeiten, den Service an der Universität Osnabrück und rechtliche Rahmenbedingungen. Videoarten, Stilmittel/Bausteine sowie technische Ressourcen werden dabei klar voneinander abgegrenzt, wodurch ein gemeinsames Verständnis und die Verwendung einer gemeinsamen Sprache sichergestellt wird. Zur Veranschaulichung wurde der Einsatz zahlreicher Beispiele und nach Bedarf aufrufbarer Infoboxen mit konkreten Optionen und Instruktionen vorgesehen. Darüber hinaus wurde beschlossen, den Aufbau des Konfigurators *„von abstrakt zu konkret“* zu gestalten.

3.4 Stufe 4: Erstellung eines Prototyps

Das Ziel der vierten Stufe bestand darin, sowohl die Erkenntnisse aus Schritt 1-3, als auch die Sichtweisen der betroffenen Stakeholder (Lehrende, Didaktik- und Video-Support - vertreten durch die Teammitglieder) bestmöglich in den prototypischen Wegweiser zu integrieren. Nach Aufkommen der Pandemie hatte sich dieser Anspruch, mit der Entstehung des Konfigurators zur Entlastung aller an der Videoproduktion Beteiligten beizutragen, in Stufe 4 zu einem zentralen Anliegen der Teammitglieder entwickelt. Aufgrund der Corona-bedingten Einschränkungen musste der weitere Design Thinking Prozess in den virtuellen Raum verlegt werden. Im Zeitraum vom 19.06.2020 - 03.07.2020 fand ein wöchentlicher Austausch statt. Während der 2- bis 3-stündigen Treffen wurde ein Online-Whiteboard zur Erstellung und Überarbeitung des Wegweisers verwendet. Während intensive Diskussionen einigte man sich auf die Visualisierung eines Wegweisers in Form eines Baumdiagramms, der nach der Evaluation in einen adaptiven, digitalen Konfigurator übertragen werden sollte. Der entstandene Prototyp des Wegweisers wird in Abbildung 3 durch die schwarzen und weißen Elemente dargestellt. Alle grauen Elemente aus Abbildung 3 wurden erst in späteren Schritten ergänzt beziehungsweise überarbeitet. Das Baumdiagramm besteht aus einer sequenziellen Abfolge von Fragen, sodass die angezeigten Inhalte im Konfigurator an die vorher getroffene Auswahl anpasst wird. Gibt ein Nutzer beispielweise an, dass er

seinen Bildschirm aufzeichnen lassen will, würden im weiteren Prozess keine ungeeigneten Stilmittel oder technischen Ressourcen (wie die Aufnahme im One Button Studio) angezeigt. Tiefergehende Informationen erhalten die Nutzer immer dort, wo in Abbildung 3 ein *i* im Kreis erscheint. Hier werden bei Bedarf eine einleitende Erklärung sowie rechtliche, didaktische und technische Hinweise gegeben. Dazu zählen beispielsweise Informationen über Nutzungsrechte verwendeter Inhalte oder Hinweise zu Schulungen bzgl. Mimik und Gestik beim Videodreh. Auch Ansprechpartner der Universität werden aufgeführt. Nachdem ein Nutzer den Konfigurator durchlaufen hat, wird eine Datei generiert, die die getroffene Auswahl zusammenfasst. Diese Übersicht kann bei Bedarf unmittelbar an benötigte Ansprechpartner wie den Video- oder Didaktik-Support weitergeleitet werden, um einen Gesprächstermin oder einen Termin für die Videoaufnahme zu vereinbaren. Andernfalls kann sich der Nutzer die Übersicht abspeichern, um anhand der Auswahl die Videoaufnahme eigenständig vorzunehmen.

3.5 Stufe 5: Testung des Prototyps

Um die Testung des Prototyps durch potenzielle Nutzer vorzunehmen, fand im August 2020 ein weiteres Fokusgruppeninterview statt. Obwohl der Wegweiser auch von erfahrenen Lehrenden als Informationsquelle genutzt werden kann und soll, richtet er sich als Orientierungshilfe primär an noch unerfahrene Lehrende. Daher wurden zu dem Interview Lehrende und deren Teammitglieder eingeladen, die bisher noch keine Erfahrung bezüglich der Entwicklung von Videos im Rahmen der Hochschullehre gesammelt haben, allerdings im Rahmen ihrer Tätigkeit zukünftig mit digitaler Lehre in Berührung kommen werden. Insgesamt konnten drei wissenschaftliche Mitarbeiter/innen und zwei studentische Hilfskräfte für das Fokusgruppeninterview 2 gewonnen werden, die dem Auswahlkriterium entsprachen. Vier der fünf Befragten standen zum Zeitpunkt des Interviews kurz vor ihrer ersten Videoproduktion, da sie die Lehrinhalte für das Wintersemester 2020 aufgrund der anhaltenden Pandemie digitalisieren mussten. Um weiterhin persönliche Kontakte innerhalb der Universität zu vermeiden, wurde das Interview virtuell durchgeführt. Der Ablauf des Fokusgruppeninterview 2 gliedert sich in die Vorstellung des Themas und der Personen, Erleben des Konfigurators, Vorstellen des Wegweisers und Diskussion. Das Interview wurde von den zwei wissenschaftlichen Mitarbeitern des Forschungsteams angeleitet, die den Termin mit einer Vorstellung des Themas und einer anschließenden Vorstellungsrunde aller Teilnehmer eröffneten. Im Anschluss wurde den Befragten ein Prototyp des Konfigurators zur Verfügung gestellt, den die Teilnehmenden im eigenen Tempo selbst durchlaufen konnten. Dafür wurde zu Anschauungszwecken eine Teilstruktur des Wegweisers in eine animierte Power-Point-Präsentation übertragen, in der sich die Befragten interaktiv bewegen konnten, um die Funktionsweise des geplanten Konfigurators zu testen. Nachdem alle Teilnehmer angaben, sich ausreichend mit dem Konfigurator befasst zu haben, wurde die zugrundeliegende Struktur des Wegweisers (Abfolge der schwarzen und weißen Felder in Abbildung 3) vorgestellt, woraufhin verschiedene Fragen diskutiert wurden. Die Ergebnisse dieser Diskussion dienten als Grundlage der Prototyp-Überarbeitung.

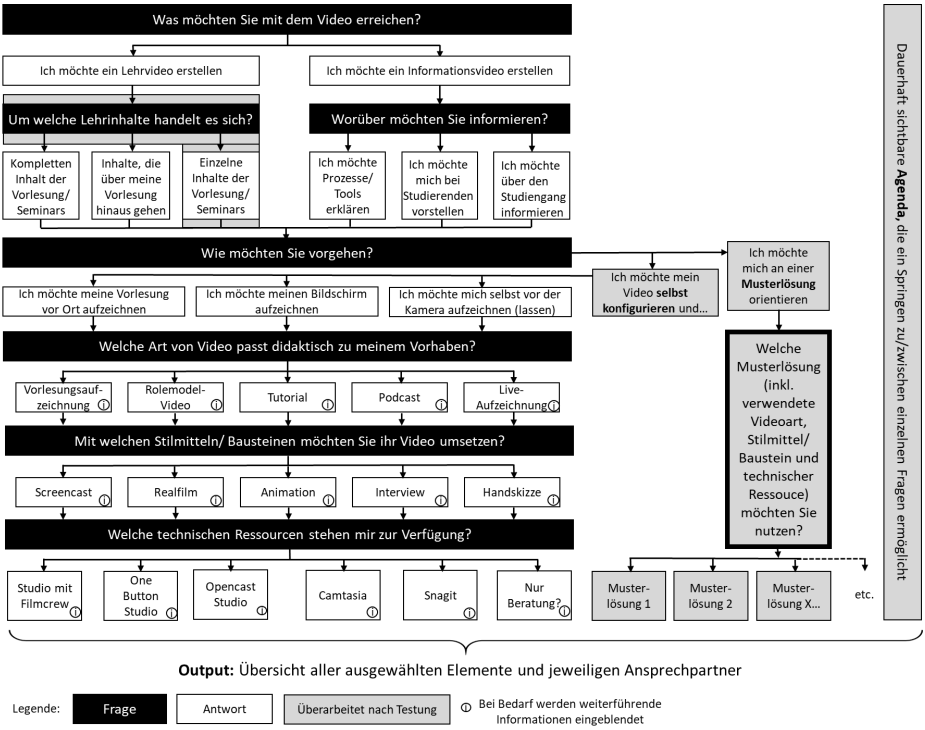


Abb. 3: Überarbeiteter Prototyp des Wegweisers als Ausgangspunkt für den Konfigurator

Mit der ersten Diskussionsfrage sollte ermittelt werden, ob der Zweck des Wegweisers deutlich wurde. Die Befragten gaben an, dass der Zweck insbesondere in Form der Unterstützung von Lehrenden erkennbar sei. Indem der Wegweiser Lehrenden einen „*Überblick über die Möglichkeiten*“ bietet, könnten sich Lehrende bereits vor einem persönlichen Beratungsgespräch über wichtige Faktoren wie verfügbare Tools informieren. Dies würde die Anfragen nach persönlichen Beratungsgesprächen verringern, so vermutete ein Teilnehmer. Zudem könnte der Wegweiser Lehrenden die Angst nehmen, zum ersten Mal ein Video zu erstellen und gleichzeitig dazu beitragen, initiale eigene Ideen zu hinterfragen. Die zweite Diskussionsfrage befasst sich mit dem größten Mehrwert aus Sicht der Befragten sowie spezifischen Änderungsvorschlägen, die sich sowohl auf die zentralen Fragen im Wegweiser, als auch auf die dargebotenen Inhalte beziehen. Als größter Mehrwert werden insbesondere die Übersichten und Informationen am Ende des Wegweisers bezüglich möglicher Videoarten, Stilmittel/Bausteinen und technischer Ressourcen gesehen. Zudem wird erwartet, dass die Anpassungsfähigkeit des Wegweisers die Konfiguration eines Videos erleichtert, da je nach Vorauswahl in den darauffolgenden Schritten ungeeignete Bausteine oder Ressourcen nicht angezeigt oder ausgegraut werden. Bezüglich der Reihenfolge der zentralen Fragen sind sich die Befragten uneinig. Während sich einige Personen bereits zu

Beginn eine Übersicht der technischen Ressourcen gewünscht hätten, vertreten andere die Meinung, dass eine derartige Übersicht unbekannter Technik wenig zielführend ist. Eine Person erklärte, dass er die bisherige Auflistung der Fragen im Wegweiser „*von abstrakt zu konkret*“ für unerfahrene Lehre zielführend findet, während Erfahrene gegebenenfalls direkt bei den Ressourcen beginnen wollen. Des Weiteren wurde von den Befragten angemerkt, dass Best Practice Beispiele und fertige Musterlösungen hilfreich wären, um sich das Ergebnis des Zusammenspiels bestimmter Videoarten, Stilmittel/Bausteine und technischer Ressourcen vorstellen zu können. Zu diesem Zweck könnten 10- bis 15-sekündige Videosequenzen präsentiert werden, unter denen die verwendeten Arten, Bausteine und Ressourcen aufgelistet werden. Zudem wäre auch ein Verweis auf die Möglichkeiten der didaktisch zielführenden Einbindung von Videos in eigene Hochschullehre sinnvoll. Eine Teilnehmerin wies darauf hin, dass der Wortlaut einer Frage und einer Antwort unklar ist und beide deutlicher formuliert werden sollten. Diskussionsfrage 3 sollte ermitteln, ob die Teilnehmer der Ansicht sind, dass ein Konfigurator für die Videoproduktion ein persönliches Beratungsgespräch teilweise oder vollständig ersetzen kann. Vier der fünf Befragten sagten aus, dass Sie den Konfigurator als erste Anlaufstelle sehen, um sich eigenständig grundlegende Informationen zu beschaffen. Eine Hilfskraft gab an, dass er dieses Vorgehen als einfacher und zeitsparender betrachtet als ein unmittelbares Beratungsgespräch. Daraufhin betonte ein wissenschaftlicher Mitarbeiter, dass die Wahl zwischen Konfigurator und Beratung für ihn primär davon abhängt, für welche Variante er am wenigsten Zeit aufwenden muss. Sollte er in beiden Fällen gleichviel Zeit investieren müssen, würde er die Beratung vorziehen. Dies gilt auch für die Vermittlung komplexer und sehr umfangreicher Inhalte, wie ein weiterer Teilnehmer betonte. Das folgende Statement eines Teilnehmers fasst die Aussagen treffend zusammen: Der Konfigurator dient „*als Vorbereitung aber nicht als Ersatz für die Beratung. Das sollte Hand in Hand gehen.*“

Anhand des Interviews wurden Überarbeitungen am Prototyp vorgenommen, die in Abbildung 3 durch graue Felder erkenntlich gemacht werden. Zum einen wird es dem Nutzer bereits frühzeitig ermöglicht, Mustervideos inkl. einer Kurzinformation über die verwendete Videoart, Stilmittel/Baustein und technische Ressource einsehen zu können. Wer sein Video zeitsparend anhand eines Musterbeispiels erstellen will, erhält hier ausschließlich die Informationen, die er/sie dafür benötigt. Sollte die Lehrperson bei der Aufnahme Hilfe benötigen, kann Sie sich mit diesen zentralen Eckpfeilern des geplanten Videos an die Video-Beratung der Universität wenden. Eine weitere Änderung betrifft die Umformulierung einer Frage und einer Antwort. Die Überarbeitungen wurden in Abbildung 3 bereits vorgenommen und durch einen grauen Rahmen ober- und unterhalb der entsprechenden Felder gekennzeichnet. Zudem wurde eine Agenda im Wegweiser implementiert, die es den Nutzern zu jeder Zeit ermöglicht, zwischen den Ebenen des Wegweisers zu springen. Dadurch können sich sowohl unerfahrene als auch erfahrene Nutzer direkt und individuell mit spezifischen Inhalten wie den technischen Ressourcen befassen.

4 Fazit

In diesem Beitrag wurde der Prototyp eines adaptiven Wegweisers entwickelt, der zukünftig Hochschullehrende bei der Videoproduktion unterstützen und gleichzeitig den Austausch zwischen Lehrenden und hochschulinternen Serviceeinrichtungen effizient gestalten soll. Der Prototyp entstand im Rahmen eines Design Thinking Prozesses, der zwei Fokusgruppeninterviews beinhaltet, die mit Lehrenden und deren Teammitgliedern durchgeführt wurden. Der Prozess begann im Dezember 2019. Sein Verlauf wurde ab März 2020 durch den Ausbruch der Corona-Pandemie beeinflusst, da zum einen Stufe 4 und 5 virtuell durchgeführt werden mussten und zum anderen der Bedarf an Informationen und Unterstützung bezüglich Videoproduktion an der Universität Osnabrück plötzlich stark anstieg. Da viele Hochschullehrende auch im Wintersemester 2020 ihre Präsenzveranstaltungen ganz oder teilweise durch digitale Medien wie Videos ersetzen müssen, gilt es in kurzer Zeit digitale Lehrinhalte zu produzieren. Der Leitfaden des Konfigurators hatte jedoch ursprünglich den Anspruch, alle Möglichkeiten der Videoproduktion schrittweise zu adressieren. Im August 2020 wurde der Prototyp anhand einer Zielgruppe getestet, die zu Zeiten der Corona-Pandemie kurz vor der ersten Videoproduktion unter starkem Zeitdruck stand. Die Bedürfnisse dieser Lehrenden sind in die Überarbeitung des Wegweisers eingeflossen, um die situative Gegebenheit zu berücksichtigen und Lehrende sowie Serviceeinrichtungen auch während der Pandemie bestmöglich zu unterstützen.

Die Struktur des Wegweisers soll zeitnah in einen adaptiven Online-Konfigurator überführt werden, um Lehrenden auch zu Zeiten der Corona-Pandemie eine Orientierungshilfe zu bieten, was zu einer effizienteren Produktion von Videos und damit zu einer Entlastung der Serviceeinrichtungen führen soll. Dabei gilt es den Konfigurator so zu gestalten, dass er sich individuell entlang der getroffenen Auswahl anpasst und gleichzeitig eine flexible Bewegung im Prozess ermöglicht. Sobald der Konfigurator fertiggestellt ist, sollte seine Nutzerfreundlichkeit überprüft werden. Wie die Testung des Prototyps ergab, wurde insbesondere die zeitsparende Informationsvermittlung als Mehrwert des Konfigurators gesehen. Daher gilt es die Inhalte, Erklärungen und Beispiele so anschaulich und kurz wie möglich zu halten. Die Grundstruktur des Wegweisers kann nicht nur von der Universität Osnabrück, sondern auch von anderen Hochschulen genutzt werden, indem sie mit hochschulspezifischen Inhalten gefüllt wird.

Limitationen dieses Beitrags sind zum einen in den teils kleinen Teilnehmerzahlen der Fokusgruppen sowie in der Corona-bedingten Verzögerung und Veränderung des Design Thinking Prozesses zu finden. Zum anderen ist der Wegweiser nicht universell übertragbar. Stattdessen muss die Grundstruktur mit universitätsinternen Informationen bestückt werden, bevor Betroffene an anderen Hochschulen von der Implementierung eines hochschulspezifischen Konfigurators profitieren können. Zudem müssen für die Entwicklung des Konfigurators ausreichende, personelle Ressourcen zur Verfügung stehen. Dabei stehen den Betroffenen, die den Konfigurator entwickeln und von ihm profitieren könnten, insbesondere zu Zeiten der Corona-Pandemie kaum zeitliche Ressourcen zur Verfügung, was die Entwicklung des Konfigurators verzögern oder gegebenenfalls sogar verhindern könnte.

Wir bedanken uns an dieser Stelle ausdrücklich bei unseren Kollegen und Teammitgliedern des Design Thinking Prozesses Henrik Jürgens und Axel Wolpert, die neben den Autoren dieses Beitrags maßgeblich an der Entstehung des Wegweisers mitgewirkt haben.

Literaturverzeichnis

[Ba18] Baran, G.: The role of design thinking in education management as a design science. In: Leading and Managing for development, Jagiellonian University Institute of Public Affairs, Krakow, S. 25-37, 2018.

[DR13] Devitt, F.; Robbins, P.: Design, Thinking and Science. In (Helfert, M.; Donnellan, B. Hrsg.): Design Science: Perspectives from Europe, Communications in Computer and Information Science. Springer International Publishing, Cham, S. 38-48, 2013.

[Di19] Dinmore, S.: Beyond lecture capture: Creating digital video content for online learning – a case study. Journal of University Teaching and Learning Practice**,** S. 1-10, 2019.

[DHKS18] Doorley, S.; Holcomb, S.; Klebahn, P.; Segovia, K.; Utley, J.: Design Thinking Bootleg. d.school at Stanford University, Hasso Plattner, Stanford, 2018.

[ESBE20] Ebner, M.; Schön, S.; Braun, C.; Ebner, M.; Grigoriadis, Y.; Haas, M.; Leitner, P.; Taraghi, B.: COVID-19 Epidemic as E-Learning Boost? Chronological Development and Effects at an Austrian University against the Background of the Concept of "E-Learning Readiness". In: Future Internet, Bd. 12, Multidisciplinary Digital Publishing Institute, S. 1-20, 2020.

[FHS19] Findeisen, S.; Horn, S.; Seifried, J.: Lernen durch Videos – Empirische Befunde zur Gestaltung von Erklärvideos. In: MedienPädagogik: Zeitschrift für Theorie und Praxis der Medienbildung, S. 16–36, 2019.

[Kl06] Klimsa, P.: Produktionssteuerung—Grundlagen der Medienproduktion. In: Handbuch Medienmanagement, Springer, Berlin, Heidelberg, S. 601–617, 2006.

[Ma07] Mayring, P.: Qualitative Inhaltsanalyse: Grundlagen und Techniken. Beltz, 2007.

[Mayr88] Mayring, P.: Die qualitative Wende. In: Deutsche Gesellschaft für Psychologie, Bd. . 36, S. 306-313, 1988.

[SF13] Seidel, V.P., Fixson, S.K.: Adopting Design Thinking in Novice Multidisciplinary Teams: The Application and Limits of Design Methods and Reflexive Practices. In: Journal of Product Innovation Management, Bd. 30, S. 19–33, 2013.

[SLV00] Soares, F.; Lopes, A.P.; Vieira, I.: Designing Video Lectures for MOOC. Proceedings of ICERI2015 Conference, S. 1873-1878, 2015.

[SSHF19] Sarooghi, H.; Sunny, S.; Hornsby, J.; Fernhaber, S.: Design Thinking and Entrepreneurship Education: Where Are We, and What Are the Possibilities? In: Journal of Small Business Management, Bd. 57, S. 78–93, 2019.

Persönliche Lernumgebungen von Studierenden im Corona-Semester

Ulrike Lucke [1]

Abstract: Das Lernen in der heimischen Umgebung birgt eine Reihe von Herausforderungen und Chancen, die in diesem Beitrag anhand der Einreichungen zu einem Fotowettbewerb an der Universität Potsdam sichtbar gemacht werden. Die Fotos und Texte der Studierenden veranschaulichen schlaglichtartig, wie diese Mischung aus physischen und virtuellen Lernorten aussehen kann, welche Schwierigkeiten damit verbunden sind und welche Auswege gefunden wurden. Dabei sind organisatorische und psychologische Fragen ebenso betroffen wie pädagogische und technische Aspekte. Eine Interpretation dieser direkt dokumentierten Erfahrungen der Studierenden entlang verschiedener Kriterien liefert wertvolle Hinweise für analoge und digitale Ansätze zur Gestaltung von Blended Learning auch nach der Pandemie.

Keywords: Personal Learning Environment; Online Learning; Home Office

1 Studium und Lehre während der Pandemie

Die Covid-19-Pandemie hatte erhebliche Auswirkungen auf die Hochschullehre. Während der Austausch von Angesicht zu Angesicht fast vollständig zum Erliegen kam, erfuhren digitale Formate eine neue (wenn auch schon lange proklamierte [Ra20]) Bedeutung. Eine Reihe von Analysen widmen sich den Folgen der Pandemie für unser Bildungssystem, wobei eine institutionelle Sicht auf Lehrformate und unterstützende Infrastrukturen dominiert [Dr20, Sc20, Za20]. Der vorliegende Beitrag nimmt dagegen studentische Lernumgebungen [KH17] in den Fokus, die sich unserer Kontrolle und normalerweise auch unserem Blick entziehen. Anders als bei längsschnittlich angelegten Studien [Bö20a, Lo20, Nö20] erfolgt hier eine Momentaufnahme in der Mitte des Sommersemesters 2020.

Das „Corona-Semester“ hat nicht nur Lehrende, Support-Personal und Entscheider an Hochschulen unvermittelt in eine bis dahin unbekannte Situation gebracht, sondern macht auch Studierende zu unfreiwilligen Teilnehmern eines umfassenden, gesellschaftlichen Experiments [Bö20b] – wobei sie die Situation kaum aktiv gestalten können. Damit erfährt die persönliche Lernumgebung in den eigenen vier Wänden eine immense Bedeutung, wird zu einer Aggregation individueller Vorstellungen von Lernen. Der Zusammenhang von Artefakten, Handlungen und Situationen ist bildlich gut erfassbar, erst recht dank dieser engen räumlichen Fokussierung. Um dies zu veranschaulichen und gleichzeitig

[1] Universität Potsdam, Institut für Informatik & Computational Science, A.-Bebel-St. 89, 14482 Potsdam, ulrike.lucke@uni-potsdam.de, https://orcid.org/0000-0003-4049-8088

doi:10.18420/inf2020_45

die damit verbundenen Herausforderungen und Chancen wissenschaftlich ausdeuten zu können, wurde an der Universität Potsdam im Sommersemester 2020 ein Fotowettbewerb[2] durchgeführt. Von dem Wettbewerbscharakter und der Aussicht auf ein Preisgeld bei gleichzeitig recht einfacher Teilnahme erwarteten wir sowohl eine höhere Beteiligung als auch bewusstere Arrangements mit dem Ziel, die Charakteristika dieser Situation möglichst treffend auszudrücken. Die Studierenden waren aufgefordert, ein Foto ihrer persönlichen Lernumgebung mit einem Titel und optional einem beschreibenden Text einzureichen. Dafür wurde in Mails an die Studierenden (in deutscher sowie etwas später noch einmal in englischer Sprache), über Twitter, auf der Website der Hochschule sowie verschiedener hochschulweiter Projekte und über die lokale Presse um Einsendungen geworben. Mit der Unterstützung eines Sponsors wurden Preisgelder für die besten Bilder ausgelobt, die von einer interdisziplinären Jury ausgewählt wurden.

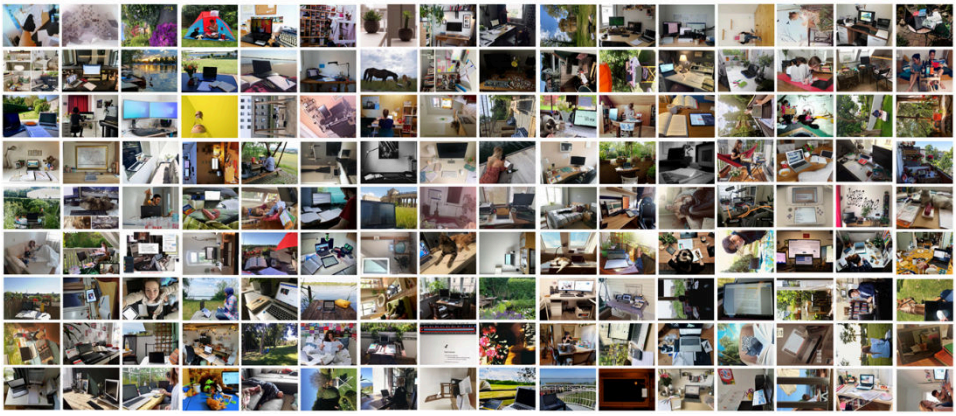


Abb. 1: Galerie der eingereichten und zur Analyse freigegebenen Fotos

Die Auswertung dieser Einreichungen richtet sich hier v.a. auf die Fotos. Abb. 1 zeigt einen Überblick. Die beschreibenden Texte werden nur ergänzend mit betrachtet. Die Ergebnisse der Analyse sind z.T. über das Corona-Semester hinaus verallgemeinerbar.

2 Auswertung der eingereichten Fotos

Vom 01.05. bis zum 30.06.2020 wurden insgesamt 148 Fotos für den Wettbewerb eingereicht; das entspricht einem Anteil von 0,7% an der derzeitigen Studierendenschaft. Männliche Einreicher waren mit insgesamt 34 (23%) gegenüber ihrem Anteil in der Studierendenschaft (43%) unterrepräsentiert. Es gab 19 Beiträge (13%) in englischer Sprache oder mit verbalem Mobilitätsbezug, was knapp dem Anteil internationaler Studierender entspricht. Alle Fotos wurden mit einem Titel versehen; lediglich 2 hatten keinen ergänzenden Text. Alle Studierenden haben eine E-Mail-Adresse hinterlassen. 21 haben nur den Vornamen angegeben,

[2] https://www.uni-potsdam.de/de/elis/fotowettbewerb

1 nur den Nachnamen und 2 ein Pseudonym; die restlichen 124 (84%) haben den vollen Klarnamen hinterlassen. 2 Einreichende haben der weiteren Auswertung widersprochen und werden nachfolgend nicht weiter betrachtet.

Die restlichen 146 Einreichungen werden in einer Bildanalyse [PW14, LD18] ausgewertet. Während Texte eine sequentielle Struktur aufweisen und so über (verbale) Argumentation Sinn zu konstruieren vermögen, weisen Bilder parallele Darstellungs- und Sinnstrukturen auf, die ikonisch/symbolisch aus Sicht von Urheber oder Betrachter interpretierbar sind und somit die nachträgliche Rekonstruktion dahinterstehender Logiken möglich bzw. nötig machen. Die von den Studierenden für den Fotowettbewerb festgehaltenen, als festhaltenswert dokumentierten, inszenierten Lernumgebungen sind folglich Ausdruck der jeweiligen Handlungspraxis und des dahinterstehenden Orientierungswissens [PW14]. Für die Rekonstruktion dieser Logiken werden quantitative Auswertungen des objektiv Wahrnehmbaren und qualitative Analysen der möglichen Interpretationen vorgenommen. Dies erfolgt entlang der folgenden Kriterien:

1. Wie ist das Bild aufgebaut?
2. Welche Hilfsmittel für das Lernen werden in den Bildern präsentiert?
3. Welche Sozialstrukturen werden durch das Bild sichtbar?
4. Welche Interpretationen ermöglichen die eingereichten Bilder?

Während die Punkte 3 und 4 vor allem die Besonderheiten des Corona-Semesters zeigen, sind die Ergebnisse zu 2 wohl auch bei einer Rückkehr in den Präsenzbetrieb noch für die persönlichen Lernumgebungen der Studierenden verallgemeinerbar; daher liegt hierauf der Schwerpunkt der Analyse. Ergänzend liefern die Ergebnisse zu 1 Indizien für den zielführenden Umgang der Studierenden mit den vorhandenen technischen Hilfsmitteln.

Zu beachten ist dabei, dass es sich um keine objektiven Aussagen handelt, sondern nur die subjektiv und teil unbewusst von den Studierenden in Szene gesetzten Elemente ihrer persönlichen Lernumgebungen. So mögen gewiss Zeitpläne im Rechner oder Stifte in einer Federtasche vorhanden gewesen sein, doch wenn sie nicht explizit auf den Fotos sichtbar waren, wurden sie nicht gezählt – sondern nur die inszenierten Elemente.

2.1 Wie ist das Bild aufgebaut?

In diesem Abschnitt werden die eher technisch-gestalterischen Aspekte betrachtet. Die meisten Einreichungen (72,6%) waren im Querformat gehalten; 22,6% waren im Hochformat und nur 4,8% quadratisch. Nur 2 Fotos (1,4%) wurden in schwarz-weiß gemacht. Bei 4 Bildern (2,7%) wurde mit Beschriften sowie mit Overlays und/oder Collagen gearbeitet. Handwerkliche Schwächen gab es nur bei wenigen Bildern; 1,4% zeigten Unschärfe, 3,4%

einen zu geringen Kontrast und 7,5% mangelnde Ausleuchtung, wobei letzteres z.T. ein Stilmittel war. Die Studierenden haben offenbar den Fokus auf die Dokumentation der Lage gelegt und waren weitestgehend zur Bedienung der Kamera in der Lage. Eine manuelle Bildbearbeitung erfolgte offenbar nur in wenigen Fällen.

2.2 Welche Hilfsmittel für das Lernen werden in den Bildern präsentiert?

Hier werden die unterstützenden oder auch mehr oder weniger bewusst ablenkenden Artefakte ausgewertet. Dazu zählen:

- analoge Lernmedien
- Einrichtungsgegenstände
- passive und aktive Rahmungen
- elektronische Geräte
- erkennbare Software/Dienste

Nicht überraschend bilden Bücher sowie handschriftliche Notizen und Stifte bei etwa der Hälfte aller Bilder die dargestellten analogen Hilfsmittel zum Lernen; siehe Abb. 2. Dagegen sind Ausdrucke oder Kopien nur bei 18,5% zu sehen. Knapp 22% der Fotos zeigen verbale oder bildliche Motivationshilfen; dagegen nur 11% Visualisierungen des Lerninhalts und 9% Zeitpläne. Auf knapp 5% der Bilder ist eine Uhr als separates Objekt zu sehen; nicht gezählt wurden Zeitanzeigen auf Bildschirmen von Notebooks etc.

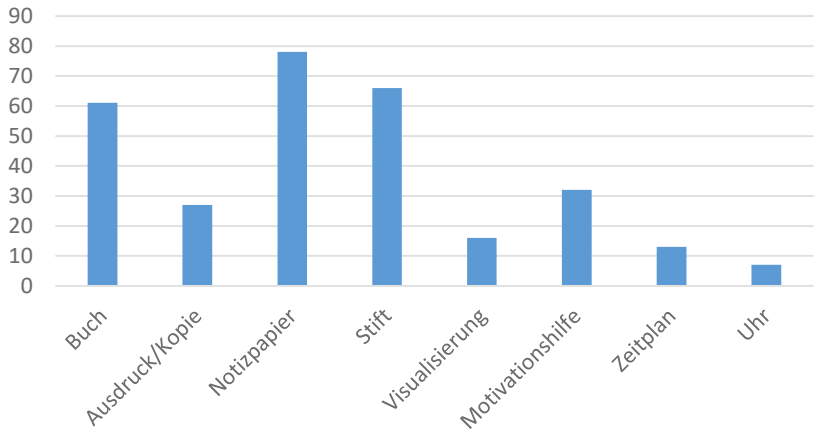


Abb. 2: Absolute Häufigkeit der abgebildeten analogen Hilfsmittel zum Lernen

Die Möblierung (siehe Abb. 3) zeigt bei den Sitzgelegenheiten mehr Variationen als bei den Tischen. Reichlich 67% der Bilder präsentieren einen für das dauerhafte Arbeiten geeigneten Schreib- oder Esstisch; auf etwa 23% der Fotos bilden die Studierenden gar keinen Tisch ab, sondern verwenden den eigenen Schoß oder den Rasen vor sich. 31,5% der Fotos beinhalten die Sitzgelegenheiten nicht; weitere 35,6% zeigen Stühle oder Bänke die eine aufrechte Sitzposition ermöglichen. Insgesamt 33,5% der Fotos zeigen lümmelnde oder liegende Lernpositionen. Nur 2 mal werden Stehplätze gezeigt.

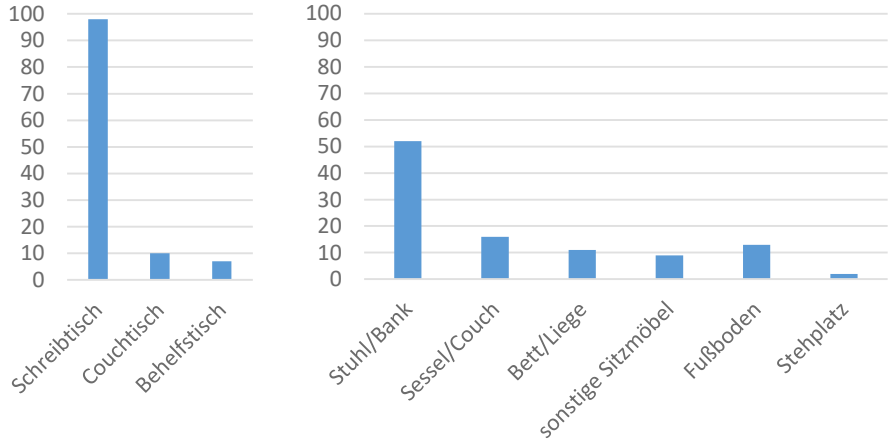


Abb. 3: Absolute Häufigkeit der abgebildeten Tischformen und Sitzgelegenheiten

Reichlich 23% der Fotos zeigen Lampen an den Arbeitsplätzen, knapp 29% Ablagen oder Staufächer. Auf knapp 10% ist ein Sonnenschutz zu sehen; Gardinen oder Jalousien an den Fenstern oder Schirme im Freien. Pflanzen (55%) und Bilder (34%) werden häufig mit abgebildet, Kuscheltiere (5,5%) dagegen selten. Vollständige Mahlzeiten sind nur an 2 der abgebildeten Arbeitsplätze zu sehen, gelegentlich Snacks an 10% und sehr häufig Getränke mit 53%. Alkohol ist auf 4 Bildern (2,7%) erkennbar.

Die Vielfalt der dargestellten elektronischen Geräte zeigt Abb. 4. Dabei dominieren Notebooks (72%), während PCs und Tablets eine untergeordnete Rolle spielen. Smartphones sind nur auf 11,6% der Fotos zu sehen – was aber wohl daran liegt, dass die Fotos damit aufgenommen wurden, wie diverse offen liegende USB-Ladekabel vermuten lassen. Nur auf knapp 11% der Fotos sind keine elektronischen Lernhilfen abgebildet.

Die Fotos zeigen nicht viele Peripherie-Geräte. Von den Fotos mit Notebooks, PCs oder Tablets zeigen nur 24% eine Maus und nur 15% Kopfhörer. Dedizierte Mikrofone oder Kameras sind kaum zu sehen. Die geringe Zahl abgebildeter Drucker passt zur seltenen Darstellung Ausdrucken oder Kopien.

Auf 67 der dargestellten Bildschirme war der Inhalt erkennbar; Abb. 5 zeigt eine Auswertung der angezeigten Programme/Dienste. Es dominieren statische Lerninhalte mit knapp 36%.

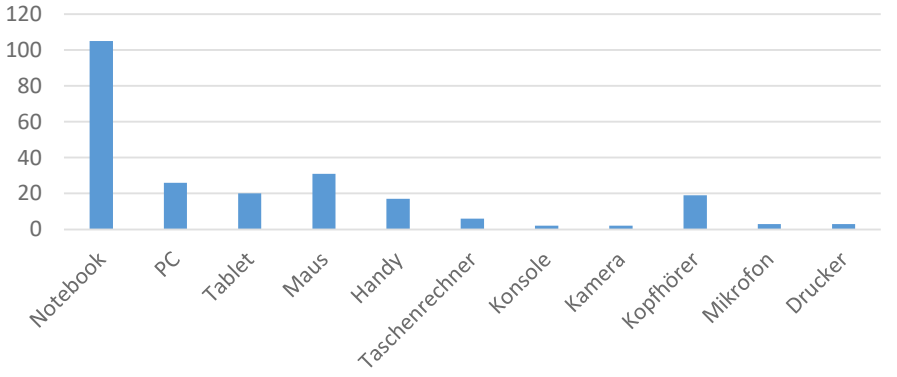


Abb. 4: Absolute Häufigkeit der abgebildeten elektronischen Hilfsmittel zum Lernen

Die Website der Hochschule, eine Videokonferenz oder ein Video(portal) sind auf jeweils 13,4% der Bildschirme zu sehen, die Lernplattform der Hochschule auf 9%. 4,5% zeigen den Web-Mailer der Hochschule, jeweils 3% die Einwahlseite des Single SignOn bzw. das lokale Campus Management System.

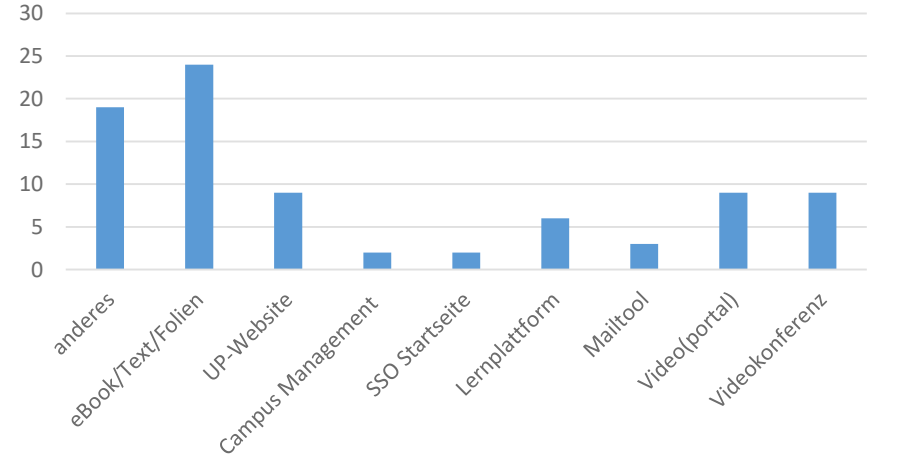


Abb. 5: Absolute Häufigkeit der abgebildeten Software/Dienste auf den Bildschirmen

Generell gilt es zu bedenken, dass die quantitative Auswertung keine Rückschlüsse auf die Nutzung der abgebildeten Hilfsmittel in der Breite der Studierendenschaft erlaubt. Bedingt übertragbar sind die Typen der abgebildeten Lernhilfen oder deren Kombinationsformen, nicht jedoch ihre Häufigkeiten.

2.3 Welche Sozialstrukturen werden durch das Bild sichtbar?

In diesem Abschnitt werden die inszenierten Lebensumgebungen genauer betrachtet:

- der gewählte Ort
- der Lernende selbst
- Lernpartner
- anwesende Personen oder Tiere

Die abgebildeten Lernorte (siehe Abb. 6) umfassen gemäß den eingesetzten analogen und digitalen Lernhilfen überwiegend geschützte Orte in oder nahe an Gebäuden. Im Inneren der Wohnung oder unmittelbar an einem Fenster liegen zusammen 69% der abgebildeten Arbeitsplätze. Nur knapp 7% der Fotos zeigen Lernorte im Garten, knapp 10% andere Orte im Grünen. Insgesamt bieten 55% der Arbeitsplätze einen Blick ins Freie.

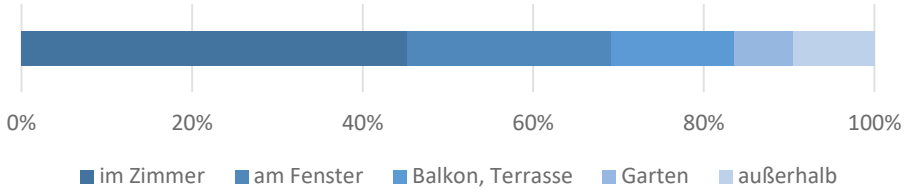


Abb. 6: Relative Häufigkeit der Lage der abgebildeten Lernorte

Die in den eingereichten Fotos erkennbaren sozialen Strukturen spiegeln die Isolation im Corona-Semester eindrücklich wider, wie Abb. 7 illustriert. 25% der Fotos zeigen die Lernenden in ihren Lernumgebungen und wurden offenbar von Dritten aufgenommen. 7% sind Selfies. In 92% der Fotos ist kein Lernpartner erkennbar. Vor Ort sind auf 4 Bildern (2,7%) andere Menschen (teils Erwachsene, teils Kinder) beim Lernen zu sehen, und auf 8 Bildern (5,5%) sind Lernpartner oder Dozenten online abgebildet.

Eine Auswertung der auf den Fotos sichtbaren anderen Personen (lernend oder nicht) und Tiere zeigt Abb. 8. Auf 6 Fotos sind die Studierenden mit ihren Kindern zu sehen. Andere Erwachsene finden sich auf nur 3 anderen Fotos. Deutlich häufiger sind Tiere zu finden, und zwar überwiegend Katzen. Die Bilder mit Vögeln und einem Schmetterling zeigen offenbar zufällige wilde Bekanntschaften.

Über 80% der Fotos zeigen niemand anderes. Jenseits der abgebildeten Einsamkeit können die ebenfalls eingereichten Texte einen tieferen Aufschluss über die Selbstbilder und Rollenerwartungen der Studierenden liefern; das ist aber nicht Gegenstand dieses Beitrags.

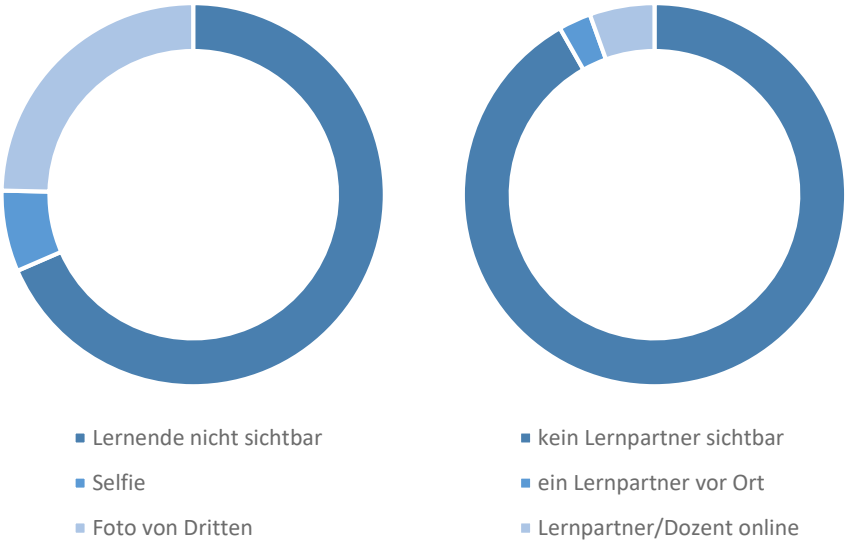


Abb. 7: Relative Häufigkeit der sichtbaren sozialen Einbindungen für die Aufnahme des Fotos (links) und für den Lernprozess (rechts)

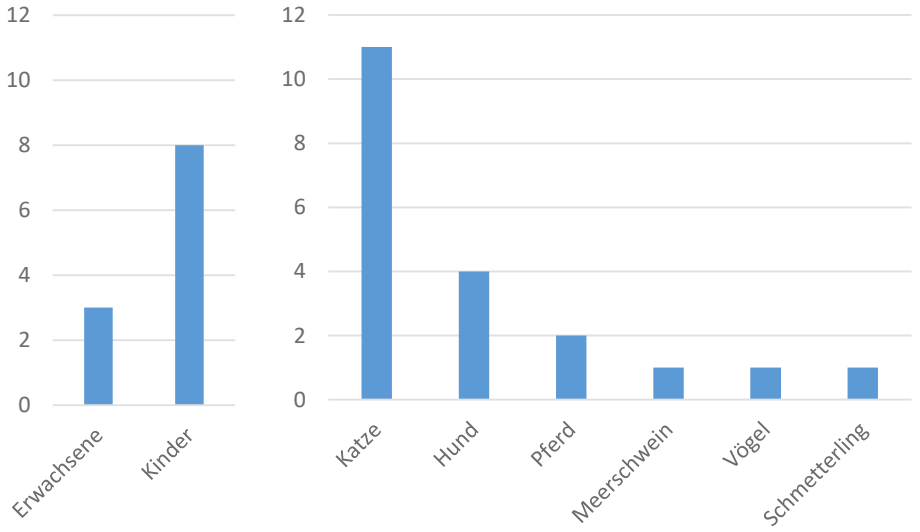


Abb. 8: Absolute Häufigkeit der abgebildeten anderen Personen (links) und Tiere (rechts)

2.4 Welche Interpretationen ermöglichen die eingereichten Bilder?

Eine manuelle Gruppierung der Einreichungen hat als mögliche Kategorien für eine zusammenfassende Interpretation die folgenden Phänotypen ergeben, wobei die Bezeichnungen noch als vorläufig zu verstehen sind:

- die Aufgeräumten
- die Eingerichteten
- die Verbundenen
- die Ausgezogenen
- die Geflüchteten
- die Kreativen

Darüber hinaus gab es Fotos, die bewusst auf künstlerische Übertreibung setzen oder sehr spezifische Details fokussieren; diese wurden hier zunächst außer Acht gelassen. Nachfolgend wird jeweils ein besonders markantes Foto aus jeder der oben genannten Kategorien vorgestellt.

Abb. 9: „Lernen allein zu Haus“ und „Morgens halb 10 im Online-Semester“

Einen typischen Vertreter der „aufgeräumten“ Kategorie zeigt Abb. 9 (links). Nicht nur die Einrichtung des Schreibtisches ist auf das Wesentliche reduziert: Notebook und Kopfhörer als Zugang zu Inhalten und zu Kommunikation, Papier und Stift als Instrumente des eigenen Lernprozesses. Auch die Bildgestaltung in schwarz/weiß unterstreicht diesen spartanischen Eindruck. – Dagegen steht in Abb. 9 (rechts) ein deutlich intensiver „eingerichtetes“ Beispiel. Der Schreibtisch zeigt ein Sammelsurium von vielerlei Dingen, die mehr oder weniger zum Lernen beitragen können. Das schließt neben verschiedenen Geräten und Unterlagen auch Beiträge für das körperliche und seelische Wohl mit ein. Dementsprechend bunt ist das Foto.

Einen „Verbund“ von Lernen und Leben zeigt das Foto in Abb. 10 (links). Hier wachsen nicht nur verschiedene Tätigkeiten zusammen: Studium, Yoga und Zuwendung für den

Abb. 10: „Lernen mit Lumpi“ und „Fernweh von der Balkon-Oase“

Hund. Auch die dafür vorgesehenen Orte verschmelzen dafür; die Isomatte wird zum Lernort, das Notebook zum einzigen Objekt des (nun gezwungenermaßen eher passiven) Lernens. – Auch in Abb. 10 (rechts) verändern sich Lernorte, hier ist ein „Auszug“ aus der Wohnung auf dem Balkon erfolgt, der eine Reduktion auf das Notwendige erfordert: Notebook, Smartphone, Wasser. Der Rest ist Grün und Sonne. Diese beiden Phänotypen rücken das Lernen ein wenig zurück und das Leben stärker in den Fokus.

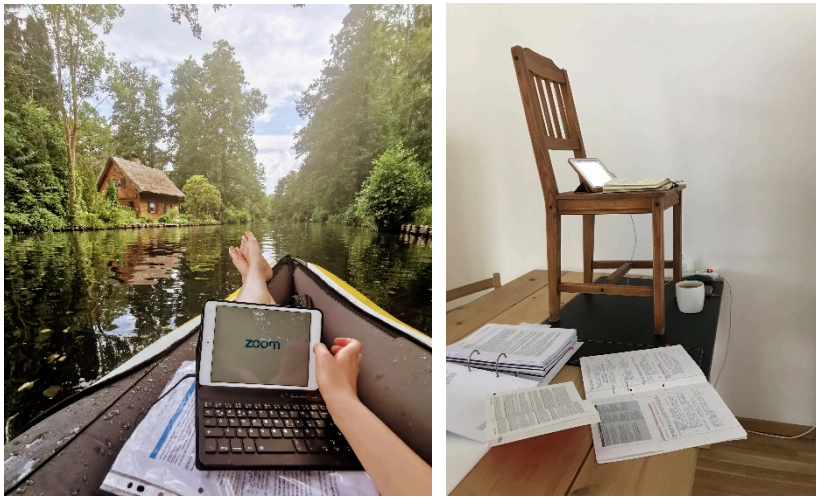


Abb. 11: „Onlineseminar mit einer Extraportion Vitamin D“ und „Gesunder Rücken trotz Arbeit am Küchentisch“

Noch radikaler ist das in Abb. 11 (links) gezeigte Beispiel für eine „Flucht“, hier auf ein Boot im Spreewald. Die Bedingungen sind nicht wirklich für das Lernen geeignet, weshalb die Papiere und das Notebook wasserfest eingeschlagen sein müssen, was die Möglichkeiten zum aktiven Lernen noch weiter einschränkt. – Zurück in die Wohnung führt Abb. 11 (rechts) mit einem Beispiel für „Kreativität“. Aus der Not eine Tugend gemacht, wird der

Esstisch samt Stuhl zum Stehplatz. Die entstehenden zwei Ebenen führen zur Einführung von Hierarchien, von primär (oben) und nur sekundär (unten) angeordneten Elementen.

Diese sechs Phänotypen von präsentierten Lernumgebungen bzw. von damit verbundenen Studierenden können beim Design mediengestützter Lernumgebungen als Orientierung (z.B. für die Formulierung von Personas) dienen.

3 Zusammenfassung und Ausblick

Die vorgestellte Analyse der Einreichungen zum Fotowettbewerb liefert nicht nur (teils sehr persönliche) Einsichten in die physischen und digitalen Lernumgebungen von Studierenden, sondern erlaubt es auch Schlussfolgerungen zu Gestaltungsprinzipien hybrider Lernumgebungen abzuleiten. Hierfür können die abgeleiteten Phänotypen als Ausgangspunkt dienen. Eine Verfeinerung der hier erfolgten manuellen Auswertung könnte durch Einsatz von automatisierten Clustering-Algorithmen erfolgen, die u.U. nochmals andere Gruppierungen finden.

Offen bleibt hier zudem eine Analyse der Fotos jenseits des Sichtbaren in ihren nonverbalen Sinnstrukturen. Das kann ikonographische Analysen und ikonologische Interpretationen umfassen [Pa79]. Auf dieser Basis können die inszenierten Rollenerwartungen und -normen [Go79] sowie der dahinterstehende Ausdruck vorhandenen Orientierungswissens tiefergehend analysiert werden. Dafür sollten ergänzend auch die eingereichten Beschreibungstexte der Bilder mit herangezogen werden.

Diese empirisch gewonnenen Erkenntnisse sind von besonderem Wert für künftige Arbeiten in Mediendidaktik und Bildungstechnologie, die von einer gezielten Synthese digitaler und analoger Formate geprägt sein werden [Ra20]. Wenn auch manche Schilderung in den eingereichten Texten eher anekdotische Evidenz als quantitative Belege für die Auswirkungen der Corona-Krise auf unsere Studierenden liefern mag, gestattet es doch der Blick über die Menge der Einreichungen hinweg fundierte Thesen über persönliche Lernumgebungen zu formulieren. Diese können von Hochschulen zur Ausgestaltung studentischer Lernorte genutzt werden [Ba16, LD18].

Danksagung

Für die Durchführung des Fotowettbewerbs gebührt Dank Sophia Rost, Maren Schulze und Christian Kröll. Für die großartige Unterstützung bei der Prämierung geht zudem ein herzliches Dankeschön an die Universitätsgesellschaft Potsdam.

Literaturverzeichnis

[Ba16] Bachmann, G.: Kollaboration auf dem Weg zum Campus von morgen. *Campus Innovation*, 2016.

[Bö20a] Böschen, S.: *Corona Tagebuch*. RWTH Aachen. https://www.soscisurvey.de/coronatb2020/

[Bö20b] Böschen, S.: Das Virus experimentiert mit uns. *Aachener Zeitung*, S. 20, 17.05.2020.

[Dr20] Dreyer, M.: *Umfrage zu den Angeboten für Digitale Lehre an Hochschulen*. ZKI/AMH, 2020. https://doi.org/10.5281/zenodo.3911092

[Go79] Goffman, E.: *Gender Advertisements – Communications and Culture*. Macmillan, 1979.

[KH17] Kiy, A.; Hofhues, S.: Persönliche Lernumgebungen im Spannungsfeld der institutionalisierten Lehre der Hochschule. In: *Trendy, hip und cool? Auf dem Weg zu einer innovativen Hochschule*. Bertelsmann, S. 117-126, 2017.

[LD18] Ladwig, T.; Duerkop, A.: *open[learning]spaces*. Exploring Learning Communities Worldwide. https://openspaces.rz.tuhh.de/

[Lo20] Logge, T.: *Coronarchiv*. Uni Gießen. https://www.coronarchiv.de/

[Nö20] Nölting, B.: *Logbuch der Veränderungen*. HNE Eberswalde. https://logbuch-der-veraenderungen.org/

[Pa79] Panofsky, E.: Ikonographie und Ikonologie. In: E. Kaemmerling (Hrsg.): *Bildende Kunst als Zeichensystem. Ikonographie und Ikonologie.* Band 1. DuMont, S. 207–225, 1979.

[PW14] Przyborski, A.; Wohlrab-Sahr, M.: *Qualitative Sozialforschung*. De Gruyter, 2014.

[Ra20] Rachel, T.: Die Digitalisierung der Bildung ist kein Selbstzweck. In: Freimuth, A.; Frenz, B. (Hrsg.) *Zukunft der universitären Lehre*. Hanns Martin Schleyer-Stiftung, Band 96, S. 27-37, 2020.

[Sc20] Schmermund, K.: Hochschulen schalten auf Notbetrieb – Ausbreitung des Coronavirus fordert digitale Lösungen. In: *Forschung & Lehre* 04/20, S. 294-295, 2020.

[Za20] Zacher, H.: Aus den Augen, aus dem Sinn? Forschen und Lehren im Homeoffice. In: *Forschung & Lehre* 05/20, S. 438-439, 2020.

Voneinander lernen – miteinander gestalten

Hochschulübergreifende Netzwerke für die Digitalisierung der Lehre

Matthias Bandtel[1]

Abstract: In diesem Beitrag werden Strategien, Strukturen und Handlungsfelder von Hochschulen zur kooperativen Weiterentwicklung digitaler Lehre diskutiert. Die Erfahrungen aus der Umstellung auf Onlineangebote im Sommersemester 2020 im Zuge der Corona-Pandemie haben die Potentiale standortübergreifender Vernetzung für die nachhaltige Entwicklung von Lehre und Lernen an Hochschulen deutlich werden lassen. Am exemplarischen Fall des Hochschulnetzwerks Digitalisierung der Lehre Baden-Württemberg (HND-BW), zu dem sich alle Landesuniversitäten zusammengeschlossen haben, um kooperative Infrastrukturen für die Digitalisierung der Lehre aufzubauen und gemeinschaftliche Dienste zu betreiben, werden Governancemodelle, Kommunikationswege und Themen hochschulübergreifender Konsortien beleuchtet. Als Ergebnis einer solchen strategischen Zusammenarbeit wird die systematische Sammlung von Handlungsfeldern der Digitalisierung der Lehre vorgestellt, in denen gemeinschaftliches Agieren als besonders wirkungsvoll erachtet wird. Anhand von vier Beispielen werden abschließend mögliche Umsetzungen eines kooperativen Vorgehens bei der gemeinsamen Gestaltung der Zukunft der Hochschulen vorgestellt.

Keywords: Hochschulnetzwerk; Digitalisierung; Lehre & Lernen; kooperative Infrastruktur

1 Einleitung

Die Zukunft der Hochschulen wird maßgeblich durch standortübergreifende Vernetzung geprägt sein. Insbesondere die Entwicklung informationstechnischer Infrastrukturen, der Betrieb lehrunterstützender Dienste und die Bereitstellung von Supportangeboten für die digitale Lehre können in Kooperation besonders effizient und nachhaltig umgesetzt werden. In diesem Beitrag werden unterschiedliche thematische Schwerpunktsetzungen und mögliche Ausgestaltungen der Governance für die institutionelle Integration der mit der Querschnittsaufgabe eLearning befassten Einheiten diskutiert. Als Beispiel wird das Hochschulnetzwerk Digitalisierung der Lehre Baden-Württemberg (HND-BW) näher beleuchtet: In diesem Verbund sind jeweils die Leitungen aller neun Landesuniversitäten sowie Vertreter*innen von Rechenzentren, Hochschuldidaktikabteilungen, eLearning-Servicestellen und Bibliotheken eingebunden. Dadurch sind alle relevanten Akteursgruppen, die zur Gestaltung der Digitalisierung der Lehre an Hochschulen beitragen, eng miteinander verzahnt. Diese Multiperspektivität ermöglicht eine ganzheitliche Betrachtung zukünftiger Entwicklungen

[1] Hochschulnetzwerk Digitalisierung der Lehre Baden-Württemberg (HND-BW), Geschäftsstelle HND-BW, KIT – Karlsruher Institut für Technologie, Karl-Friedrich-Str. 17, 76133 Karlsruhe, matthias.bandtel@kit.edu

doi:10.18420/inf2020_46

in der Digitalisierung der Lehre. Am Beispiel von vier Handlungsfeldern werden die synergetischen Potentiale der hochschulübergreifenden Zusammenarbeit bei der kooperativen Weiterentwicklung des digitalen Lehrens & Lernens illustriert.

2 Voneinander lernen: Systematisierung der Erfahrungen aus dem Online-Sommersemester 2020

Die pandemiebedingte Aussetzung der Präsenzlehre im Sommersemester 2020 und die dadurch notwendig gewordene Umstellung auf Online-Angebote hat an vielen Hochschulen zu einem wahren Entwicklungssprung der Unterstützungsstrukturen für digitale Lehre geführt. In einem gemeinschaftlichen Kraftakt von Hochschulleitungen, Rechtsabteilungen, Rechenzentren, Bibliotheken, lehrunterstützenden Einheiten wie eLearning-Services und Hochschuldidaktikstellen sowie Lehrenden & Lernenden aller Fächer konnten technische Infrastrukturen ertüchtigt, Beratungs- und Supportangebote ausgeweitet und Hochschulstrukturen auf die Querschnittsaufgabe abgestimmt werden. Einige Schlaglichter auf die Maßnahmen, die exemplarisch an baden-württembergischen Universitäten getroffen wurden, verdeutlichen das Ausmaß der erfolgten Umstellungen.

Im Bereich der technischen Infrastrukturen wurden die Kapazitäten vorhandener Dienste – allen voran Lernplattformen wie ILIAS oder moodle – aufgestockt und an die gesteigerte Nachfrage und Nutzung angepasst. Am Karlsruher Institut für Technologie (KIT) beispielsweise hat sich die Nutzungsintensität von ILIAS verzehnfacht. Damit dieser Entwicklung Schritt gehalten werden konnte, mussten die Ressourcen und Kapazitäten der ILIAS-Systemarchitektur am KIT um den Faktor drei erweitert werden [Hoye20]. Um der sprunghaft angestiegenen Zahl externer Zugriffe auf die Universitätsnetze gerecht zu werden, mussten die VPN-Dienste massiv aufgerüstet werden. Im Falle des KIT verzeichnete das Steinbuch Centre for Computing (SCC) vor Beginn des ersten „Corona-Semesters“ maximal 500 gleichzeitige Nutzungen des VPN-Dienstes. Durch die Umstellung auf Online-Lehre und die Verlagerung auf das Home-Office stieg die Auslastung immens auf 2.500 parallele Nutzungen an. Um die CPU-Last angesichts dieses Ansturms im grünen Bereich halten zu können, wurden die ursprünglich vier virtuellen Maschinen, auf denen der OpenVPN-Dienst am KIT betrieben wird, um sechs weitere ergänzt [MaNS20]. An allen Universitäten wurden Videokonferenzwerkzeuge in Betrieb genommen. Dabei setzte ein Großteil der baden-württembergischen Universitäten aus Verfügbarkeitsgründen (zunächst) auf Clouddienste wie Cisco Webex, Zoom oder MS Teams [HaSe20, MeSe20]. An einzelnen Standorten wurden alternativ oder parallel On-Premise- respektive Open Source Lösungen auf Basis von Jitsi und Big Blue Button aufgebaut. Letztere wurden nicht zuletzt benötigt, um sichere und datenschutzkonforme Lösungen für Online-Prüfungen bereit zu stellen [BKWW20]. Zur Verwaltung der rapide angestiegenen Menge an Lehr-Lernvideos mussten Opencast-Server oder Video-Content-Management-Systeme installiert werden [Dier20]. Die neu aufgesetzten Dienste wurden in die Gesamtinfrastruktur integriert, um eine zentrale Administration zu gewährleisten und die Nutzung für Lehrende & Lernende intuitiv zu gestalten, insbesondere

durch Integration in die jeweiligen Lern-Management-Systeme. Hardwareseitig wurde darüber hinaus an einzelnen Universitäten in Baden-Württemberg die Hörsaalausstattung für Aufzeichnung und Streaming von Veranstaltungen ausgebaut, um Lehrenden sogenannte „Geistervorlesungen“ zu ermöglichen.

Mit Blick auf Support und Beratung wurden analog zu den Anpassungen der IT-Versorgung entsprechende Unterstützungsangebote für Lehrende erheblich ausgeweitet respektive neu implementiert. Viele Rechenzentren haben ihre Servicedesks verstärkt, um die Nachfrage nach Support bei der Einrichtung und Nutzung neuer Dienste zu bedienen [West20]. Über rein technische Fragen hinaus sind an den baden-württembergischen Universitäten Informationen zur Online-Lehre jeweils auf einer zentralen Plattform gebündelt worden. Dort wurden Szenarien für den Online-Lehrbetrieb, Kursvorlagen und Qualifizierungsangebote bereitgestellt. Teilweise wurden Schulungen in Form von Online-Seminaren angeboten [BaSe20]. Die Beratungskapazitäten wurden zum einen durch die Verstärkung kollegialer Netzwerke und zum anderen durch Umverteilung vorhandener Ressourcen kurzfristig aufgestockt. Peer-to-Peer-Beratungen und Diskussionsforen wurden ausgeweitet. An einigen Universitäten konnten Multiplikatorinnen und Multiplikatoren zur Dissemination zentraler Angebote in Fakultäten installiert werden.

Nicht zuletzt haben Universitäten und Hochschulen ihre Strukturen an die Herausforderungen des Online-Semesters angepasst. Die mit verschiedenen Aspekten digitaler Lehre befassten Einheiten (insbesondere Rektorate respektive Präsidien, Rechenzentren, Bibliotheken, Rechtsabteilungen, Didaktikstellen und eLearning-Services) haben sich noch enger verzahnt [FNNS20, SeNu20]. Innerhalb praktisch aller Hochschulen sind Task-Forces gebildet worden. Zudem wurden hochschulübergreifende Initiativen und Netzwerke gestärkt oder neu eingerichtet. Diese strukturelle Dimension der Folgen des ersten Corona-Semesters ist in ihrer Relevanz für die künftige Entwicklung der Digitalisierung von Lehre & Lernen nicht zu unterschätzen. Denn gerade in der Krise haben die Universitäten und Hochschulen unter Beweis gestellt, dass sie in der Lage sind, ihre Strukturen agil an sich verändernde Umweltbedingungen anzupassen. Aus der Bereitschaft, zentrale Herausforderungen kooperativ anzugehen, erwächst die Stärke, die Zukunft gemeinsam zu gestalten (vgl. Kap. 5).

3 Miteinander Gestalten: Entwicklungsperspektiven

Die positiven Erfahrungen mit der Umstellung auf Online-Lehre haben die Erwartungshaltung von Lehrenden und Lernenden nachhaltig verändert. Welche Elemente digitaler Lehre werden die zukünftige Entwicklung an Hochschulen in besonderem Maße prägen?

Auf operativer Ebene setzen sich eLearning-Expert*innen im Hochschulnetzwerk Digitalisierung der Lehre Baden-Württemberg regelmäßig mit dieser Frage auseinander. Ihre aktuelle systematische Bestandsaufnahme ist vor dem Hintergrund der Erfahrungen aus dem Corona-Semester noch einmal bestätigt worden [Hoch19]. Vier Feldern kommt dabei eine besondere Relevanz zu.

Lernplattformen wie moodle oder ILIAS nehmen als virtueller Veranstaltungsraum an Wichtigkeit zu. Hier steht zu erwarten, dass die Nutzung fortgeschrittener Funktionen steigen wird, um komplexere Lehr-Lernsettings zu realisieren. Daraus resultiert eine erhöhte Nachfrage nach didaktischem und technischem Support.

Auch bei qualitätsvollen Lehr-Lernvideos und Vorlesungsaufzeichnungen wird davon ausgegangen, dass die Nachfrage weiter steigen wird. Hierdurch erhöhen sich die Bedarfe nach sowohl Serverkapazitäten als auch didaktischer Beratung. Zudem wird die Bedeutung von Videokonferenzen für flexible und standortübergreifende Lehr-Lernangebote weiter wachsen. Die vielerorts bislang verwendeten Cloud-Dienste sind teilweise aus datenschutzrechtlichen und ethischen Gesichtspunkten unbefriedigend [Berl20, John20, Zent20]. Für den Dauereinsatz müssen gegebenenfalls kooperative Lösungsansätze im Verbund mehrerer Hochschulen entwickelt und betrieben werden.

Zudem gewinnt das Thema Open Educational Resources (OER) weiter an Relevanz [Mini19]. Wenn allerorts vermehrt digitale Lehr-Lernmaterialien produziert werden, steigen auch Synergiepotentiale, die durch eine geteilte Nutzung realisiert werden könnten. Um die Bekanntheit und Akzeptanz von OER in der Breite zu erhöhen, müssen geeignete Anreizsysteme für Lehrende implementiert werden. Zudem müssen Regime zur Qualitätssicherung und -entwicklung installiert werden. Und nach wie vor gibt es eine große Nachfrage nach Beratung in (urheber-)rechtlichen Fragen im Zusammenhang mit OER, der durch entsprechende lokale oder zentrale Angebote gedeckt werden muss [Bund19].

Nicht zuletzt werden elektronische Prüfungen nachhaltig das Spektrum an Prüfungsformen erweitern. Dabei geht es um weit mehr als nur um das aktuell heiß diskutierte Online-Proctoring [Baum20, Mcgr20]. Mit Open- oder Closed Book Klausuren, fernmündlichen Prüfungen, formativen Varianten wie Portfolios, Gamification-Elementen wie Quizze oder innovativen Formaten, die an forschendes bzw. problembasiertes Lernen angelehnt sind, stehen neue (didaktische) Möglichkeiten zur Verfügung, Leistungskontrollen kompetenzorientiert und lernendenzentriert durchzuführen und Studierenden individuelles Feedback zum Lernerfolg zu geben [HoSc20]. Diese Potentiale sind längst nicht ausgeschöpft. In diesem Kontext sind noch grundlegende technische sowie (datenschutz- und prüfungs-)rechtliche Fragen zu klären [Elan14, Hoch15a, Hoch15b]. Außerdem müssen hochschul- und mediendidaktische Aufklärungsarbeit geleistet und entsprechende Unterstützungsangebote für Lehrende aufgesetzt werden [HoSc20, MmMi15].

4 Chancen: Gelingensbedingungen für die erfolgreiche Weiterentwicklung digitaler Lehre

Vor dem Hintergrund der Erfahrungen aus dem Online-Semester 2020 eröffnet die gegenwärtige Situation eine besondere Chance für die erfolgreiche Weiterentwicklung des digitalen Lehrens & Lernens. Um positive Ansätze zu konsolidieren und nachhaltig fortzuführen, ist

es essentiell, einige Rahmenbedingungen, die die schnelle Umstellung auf Online-Lehre geprägt haben, auch langfristig zu erhalten.

Zunächst einmal hat eine allgemeine Atmosphäre der Offenheit, Fehlertoleranz und Nachsicht wesentlich dazu beigetragen, Berührungsängste bei der Produktion, Distribution und Nutzung digitaler Angebote zu minimieren. Darüber hinaus konnten Lehrende und Servicestellen von einer (temporären) Entlastung von rechtlichen Unsicherheiten profitieren. Um das geweckte Interesse weiterzutragen sowie verbliebene Skeptikerinnen und Skeptiker zu überzeugen, empfiehlt es sich, Kapazitäten in den Supportstrukturen auszubauen.

Die Institutionalisierung der Zusammenarbeit aller mit dieser Querschnittsaufgabe befassten Einheiten in Rechenzentren, Bibliotheken, Hochschuldidaktik, Personalentwicklung, Rechtsabteilungen und eLearning-Zentren ist für die nachhaltige Gestaltung digitaler Lehre zentral. Multiplikatorinnen und Multiplikatoren in den Fakultäten fördern die Akzeptanz digitaler Lehrformen im Kollegium und vermitteln fachspezifische Bedarfe an lehrunterstützende Einheiten.

Nicht zuletzt fungieren hochschulübergreifende Netzwerke, Initiativen und Kooperationen als Plattformen für den Erfahrungsaustausch und Motoren für künftige Entwicklungen. Sie bilden gleichsam „politische Rahmenbedingungen“ [Schm00] für die Digitalisierung der Lehre. In einer aktuellen Studie des HIS-Instituts für Hochschulentwicklung haben Bosse und Würmseer (2020) Hochschulverbünde systematisiert [BoWü20]. Insbesondere mit Blick auf die Ausgestaltung der Kooperationsgovernance sind vorliegende Studien zur „Kooperation von Rechenzentren“ [SHWS16] ausgesprochen instruktiv. Aus Umsetzungsperspektive thematisieren die Diskussionspapiere zum *Future Lab* des Stifterverbandes inhaltliche und strukturelle Fragen von Hochschulkooperationen [WaHo20, WaNK20, WCEW20, WKNW19, WWND19].

Einige Beispiele – freilich bei Weitem ohne Anspruch auf Vollständigkeit – veranschaulichen im Folgenden unterschiedliche Schwerpunktsetzungen sowie die verschiedenen Beiträge zu einer ganzheitlichen Gestaltung digitalen Lehrens & Lernens.

Der *Arbeitskreis der Leiterinnen und Leiter der wissenschaftlichen Rechenzentren in Baden-Württemberg (ALWR)* formiert sich aus den Leiterinnen und Leitern der Rechenzentren bzw. Informationszentren der baden-württembergischen Universitäten. Im Auftrag der Landesrektorenkonferenz und in Abstimmung mit dem Wissenschaftsministerium entwickelt und gestaltet der ALWR landesweite Strategien für IT-Dienste und Infrastrukturen in Forschung, Lehre und Verwaltung, setzt Konzepte und gemeinsame Lösungen um [Arbe00, Mini14].

Die *Zentrale Datenschutzstelle der baden-württembergischen Universitäten (ZENDAS)* ist eine Serviceeinrichtung, die bei Rechtsfragen berät und juristisch fundierte Informationsmaterialien zur Verfügung stellt. Die Bündelung von Ressourcen in der Bearbeitung komplexer werdender Rechtsfragen im Zusammenhang mit digitaler Lehre wird bereits seit längerer Zeit von Universitäten und Hochschulen gefordert [Mini15, S.57].

Das *eLearning Academic Network e.V. (ELAN)* versteht sich als „Impulsgeber zur stetigen Qualitätsverbesserung der medienbasierten Lehre an niedersächsischen Hochschulen“ [Elan00]. Durch Unterstützungsmaßnahmen in verschiedenen Kompetenzbereichen (audiovisuelle Medien und Medientechnik, E-Assessment, Mediendidaktik, Rechtsfragen des eLearnings sowie Software für Lehre, Studium und deren Management) verfolgt ELAN e.V. das Ziel, die Kooperation der Mitgliedshochschulen und weiterer Mitglieder im Bereich standortübergreifender Lehre und eLearning weiterzuentwickeln.

Zum *Hochschulnetzwerk Digitalisierung der Lehre Baden-Württemberg (HND-BW)* haben sich die neun Landesuniversitäten Freiburg, Heidelberg, Hohenheim, das Karlsruher Institut für Technologie (KIT), Konstanz, Mannheim, Stuttgart, Tübingen und Ulm zusammengeschlossen, um die digitale Lehre gemeinsam weiterzuentwickeln. Das HND-BW setzt sich aus den Prorektorinnen und Prorektoren respektive Vizepräsidentinnen und Vizepräsidenten für Lehre, einer Vertreterin oder einem Vertreter des Kreises der Prorektorinnen und Prorektoren für Digitalisierung (bwCIO) sowie eLearning-Expertinnen und Experten der Netzwerkuniversitäten aus Rechenzentren, Bibliotheken, Didaktikstellen und Lehr-Lernzentren zusammen. Hochschulartenübergreifend besetzte Special Interest Groups arbeiten themenspezifisch zu relevanten Handlungsfeldern der Kooperation (vgl. Kap. 5). Diese Konstellation verzahnt zum einen Leitungs- und operative Ebene hochschulübergreifend miteinander. Zum anderen eröffnet die unterschiedliche Verortung der Mitglieder einen ganzheitlichen Blick auf das Querschnittsthema Digitalisierung der Lehre. So werden miteinander die Ziele verfolgt, Herausforderungen gemeinsam zu meistern und geschlossen die politisch, rechtlichen und finanziellen Rahmenbedingungen für digitales Lehren & Lernen zu gestalten.

5 Gemeinsam handeln: Themenfelder der Kooperation

Anhand von vier exemplarischen Handlungsfeldern der hochschulübergreifenden Kooperation im Bereich digitaler Lehre werden die Potentiale einer gemeinschaftlichen Lösungsfindung deutlich. Im Folgenden werden konkrete Projekte vorgestellt, die aus der Zusammenarbeit der Landesuniversitäten im HND-BW hervorgegangen sind.

Kooperativer Aufbau und Betrieb eines Repositoriums für Open Educational Resources. Aus dem HND-BW ist mit dem ZOERR ein Zentrales Repositorium für Open Educational Resources aufgebaut worden. Es wird als zentraler Dienst unter der Federführung der Universität Tübingen von mehreren Hochschulrechenzentren und -bibliotheken dauerhaft betrieben.

Arbeitsumgebungen für ePrüfungen. Die Universitäten im HND-BW sind der Überzeugung, dass elektronische Prüfungen künftig eine größer werdende Rolle spielen werden, weil sie flexible und ortsunabhängige Leistungskontrollen ermöglichen (vgl. Kap. 3). Allerdings erfordert ihr flächendeckender Einsatz grundlegende Vorarbeiten, die die Kapazitäten einzelner Standorte übersteigen (z.B. Aufbau der technischen Infrastruktur, Klärung rechtlicher Rahmenbedingungen, Regelung der organisatorischen Abwicklung und didaktische Einbindung). Das HND-BW hat zu diesen Fragen eine gemeinsame Arbeitsgruppe eingesetzt,

die Anforderungen identifiziert, Umsetzungsmodelle bewertet und den Ressourcenbedarf verschiedener ePrüfungslösungen bemisst.

Zentrale Beratungsstelle für Rechtsfragen digitaler Lehre. Durch die Ausweitung digitaler Lehre nimmt die Komplexität datenschutz-, urheber- und hochschulrechtlicher Problemstellungen zu. Hochschulen wünschen sich eine zentrale Beratungsstelle, bei der rechtliche Fragen gesammelt und verbindlich Auskunft eingeholt werden kann. Auf regionaler Ebene bestehen in Deutschland bereits einige Initiativen in diesem Feld (vgl. Kap. 4). Für Baden-Württemberg erörtert das HND-BW aktuell im Rahmen einer Special Interest Group, wie die Reichweite dieser Angebote ausgeweitet, institutionell gestärkt und gegebenenfalls ergänzt werden kann.

Data Literacy Education. Hochschulen stehen in der besonderen Verantwortung, Lehrenden und Lernenden aller Fächer einen planvollen, verantwortlichen und kritischen Umgang mit Daten zu ermöglichen. Über die Entwicklung von Anwendungskompetenzen zur Erhebung, Kuratierung, Auswertung und Visualisierung von Daten hinaus geht es dabei auch um die Sensibilisierung für datenethische Gesichtspunkte. Bei der Konzeption entsprechender Kompetenzrahmen, der Entwicklung geeigneter Lehr-Lernangebote und der institutionellen Integration von Data Literacy Education in die Curricula aller Fächer können Hochschulen von der standortübergreifenden Vernetzung immens profitieren [BaKW21]. Positive Beispiele, wie das durch den Stifterverband initiierte deutschlandweite Data Literacy Education Netzwerk, verdeutlichen das in der praktischen Umsetzung.

6 Fazit

Die Erfahrungen aus dem Sommersemester 2020 haben die Potentiale aber auch die Herausforderungen von Online-Lehr-Lernangeboten in der Breite sichtbar werden lassen. In der Ausnahmesituation der Aussetzung der Präsenzlehre ist vor allem deutlich geworden, dass für die nachhaltige Digitalisierung von Lehre & Lernen substantielle Entwicklungen der technischen Infrastrukturen, auf Ebene der Support-, Beratungs- und Qualifizierungsangebote für Lehrende sowie mit Blick auf Organisationsstrukturen innerhalb und zwischen Hochschulen erforderlich sind.

Auf zentralen Handlungsfeldern der Digitalisierung von Lehre & Lernen eröffnet gerade die hochschulübergreifende Zusammenarbeit wertvolle Potentiale. Dabei reicht das Spektrum vom niedrigschwelligen Erfahrungsaustausch unter Kolleginnen und Kollegen in der Lehre oder in lehrunterstützenden Einrichtungen über den Betrieb gemeinsamer Dienste bis zur kooperativen Entwicklung von Digitalisierungsstrategien.

Die strategische Stärke von Hochschulkooperationen wird mit Blick auf die großen Herausforderungen der Digitalisierung von Lehre & Lernen weiter an Bedeutung zunehmen. Insbesondere in der Gewinnung der erforderlichen Ressourcen sowie bei der Ausgestaltung rechtlicher Rahmenbedingungen können Hochschulen gemeinsam in einen konstruktiven Dialog mit der Politik auf Bundes- und Landesebene treten.

Literaturverzeichnis

[Arbe00] Arbeitskreis der Leiterinnen und Leiter der wissenschaftlichen Rechenzentren in Baden-Württemberg: *Leitbild des ALWR*. URL https://www.alwr-bw.de/leitbild-des-alwr/. - abgerufen am 2020-08-24.

[BaKW21] Bandtel, Matthias ; Kauz, Leonie ; Weissker, Natalia: Data Literacy Education für Studierende aller Fächer. Kompetenzziele, curriculare Integration und didaktische Ausgestaltung interdisziplinärer Lehr-Lernangebote. In: Hochschulforum Digitalisierung beim Stifterverband (Hrsg.): *Digitalisierung in Studium und Lehre gemeinsam gestalten*. Wiesbaden : Springer, 2021 (i.E.)

[BaSe20] Ballach, Janina ; Sexauer, Andreas: Vorbereitung und Unterstützung der Lehrenden. In: *SCCnews - Corona-Sonderteil: Online-Lehre und Homeoffice am KIT* Bd. 2020 (2020), Nr. 01, S. 13

[Baum20] Baume, Matthias: Beaufsichtigung von digitalen Prüfungsformaten (Online-Proctoring) - Interview mit Matthias Baume. In: Gino Krüger: *Hochschulforum Digitalisierung Blog* (24.04.2020). URL https://hochschulforumdigitalisierung.de/de/blog/online-proctoring/. - abgerufen am 2020-08-20.

[Berl20] Berliner Beauftragte für Datenschutz und Informationsfreiheit: *Hinweise für Berliner Verantwortliche zu Anbietern von Videokonferenz-Diensten*. Berlin : BlnBDI, 2020

[BKWW20] Braun, Samuel ; Krauss, Peter ; Weiss, Ulrich ; Wenzel, Alvar: Online-Prüfungen in Zeiten der Pandemie. In: *SCCnews - Corona-Sonderteil: Online-Lehre und Homeoffice am KIT* Bd. 2020 (2020), Nr. 01, S. 20–22

[BoWü20] Bosse, Elke ; Würmseer, Grit: *Hochschulverbünde. Ein aktueller Überblick zu Rahmenbedingungen, Organisation, Herausforderungen und Erfolgsfaktoren lehrbezogener Zusammenarbeit*, *HIS-HE: Medium* (Nr. 4, 2020). Hannover : HIS-Institut für Hochschulentwicklung e. V., 2020

[Bund19] Bundesministerium für Bildung und Forschung: *Urheberrecht in der Wissenschaft. Ein Überblick für Forschung, Lehre und Bibliotheken*, *Fachinformation des Bundesministeriums für Bildung und Forschung*. Berlin : BMBF, 2019

[Dier20] Dierolf, Uwe: Wissenswertes zum Opencast-Projekt im Corona-Sommersemester 2020. In: *SCCnews - Corona-Sonderteil: Online-Lehre und Homeoffice am KIT* Bd. 2020 (2020), Nr. 01, S. 16–18

[Elan00] ELAN e.V.: *Über den ELAN e.V.* URL https://elan-ev.de/ueber_elan_ev.php. - abgerufen am 2020-08-24

[Elan14] ELAN e.V.: *E-Assessments & E-Klausuren. E-Prüfungen an Hochschulen*. URL https://ep.elan-ev.de/wiki/Hauptseite. - abgerufen am 2020-08-20.

[FNNS20] Frank, Martin ; Neumair, Bernhard ; Nussbaumer, Martin ; Streit, Achim: Editorial. In: *SCCnews - Corona-Sonderteil: Online-Lehre und Homeoffice am KIT* Bd. 2020 (2020), Nr. 01, S. 2

[HaSe20] VON DER HAGEN, PATRICK ; SEXAUER, ANDREAS: Einführung von Zoom. In: *SCCnews - Corona-Sonderteil: Online-Lehre und Homeoffice am KIT* Bd. 2020 (2020), Nr. 01, S. 14–15

[Hoch15a] HOCHSCHULFORUM DIGITALISIERUNG: *E-Assessment als Herausforderung – Handlungsempfehlungen für Hochschulen*, *Themengruppe Change Management & Organisationsentwicklung* (Arbeitspapier Nr. 2). Berlin : Hochschulforum Digitalisierung, 2015

[Hoch15b] HOCHSCHULFORUM DIGITALISIERUNG: *E-Asessment als Herausforderung: Handlungsempfehlungen für die Hochschulpolitik*, *Themengruppe „Innovationen in Lern- und Prüfungsszenarien" koordiniert vom CHE im Hochschulforum Digitalisierung*. Berlin : Hochschulforum Digitalisierung, 2015

[Hoch19] HOCHSCHULNETZWERK DIGITALISIERUNG DER LEHRE BADEN-WÜRTTEMBERG: *Positionspapier zur Digitalisierung in der Lehre an baden-württembergischen Hochschulen*. Baden-Württemberg : Hochschulartenübergreifender Arbeitskreis (HÜA) des Hochschulnetzwerks Digitalisierung (HND-BW), 2019

[HoSc20] HORN, JANINE ; SCHMEES, MARKUS: *Online-Pruefungen*, *ELAN e.V. Handouts*. Oldenburg : ELAN e.V., 2020

[Hoye20] HOYER, PHILIP: Mit ILIAS fit für die Online-Lehre. In: *SCCnews - Corona-Sonderteil: Online-Lehre und Homeoffice am KIT* Bd. 2020 (2020), Nr. 01, S. 19–20

[John20] JOHN, NICOLAS: Corona is calling. Datenschutzrechtliche Probleme bei der Auswahl und Benutzung von Videokonferenzprogrammen für den Arbeits- und Hochschulalltag. In: *DFN-Infobrief Recht* Bd. 2020 (2020), Sonderausgabe Corona, S. 6–8

[MaNS20] MALL, KLARA ; NEUFFER, BENEDIKT ; SCHUH, JULIAN: Ausbau der Netzinfrastruktur. In: *SCCnews - Corona-Sonderteil: Online-Lehre und Homeoffice am KIT* Bd. 2020 (2020), Nr. 01, S. 7–8

[Mcgr20] MC GRATH, OWEN: Mensch gegen Maschine. Zur datenschutzrechtlichen Relevanz von automatisierten Proctoringdiensten bei digitalen Prüfungen. In: *DFN-Infobrief Recht* Bd. 2020 (2020), Nr. 9, S. 2–4

[MeSe20] MEIER, MICHAEL ; SEXAUER, ANDREAS: Microsoft Teams in der Lehre. In: *SCCnews - Corona-Sonderteil: Online-Lehre und Homeoffice am KIT* Bd. 2020 (2020), Nr. 01, S. 12

[Mini14] MINISTERIUM FÜR WISSENSCHAFT, FORSCHUNG UND KUNST BADEN-WÜRTTEMBERG: *E-Science: Wissenschaft unter neuen Rahmenbedingungen. Fachkonzept zur Weiterentwicklung der wissenschaftlichen Infrastruktur in Baden-Württemberg*. Stuttgart : Ministerium für Wissenschaft, Forschung und Kunst Baden-Württemberg, 2014

[Mini15] MINISTERIUM FÜR WISSENSCHAFT, FORSCHUNG UND KUNST BADEN-WÜRTTEMBERG: *E-Learning. Strategische Handlungsfelder der Hochschulen des Landes Baden-Württemberg zur Digitalisierung in der Hochschullehre*. Stuttgart : Ministerium für Wissenschaft, Forschung und Kunst Baden-Württemberg, 2015

[Mini19] Ministerium für Wissenschaft, Forschung und Kunst Baden-Württemberg: *Open Educational Resources (OER). Informationen für Hochschullehrende zur Nutzung und Veröffentlichung von OER*. Stuttgart : Ministerium für Wissenschaft, Forschung und Kunst Baden-Württemberg, 2019

[MmMi15] MMB-Institut für Medien- und Kompetenzforschung ; Michel, Lutz P.: *Digitales Prüfen und Bewerten im Hochschulbereich*, *CHE Themengruppe „Innovationen in Lern- und Prüfungsszenarien“ im Hochschulforum Digitalisierung* (Arbeitspapier Nr. 1). Berlin : Hochschulforum Digitalisierung, 2015

[Schm00] Schmidt, Markus: *Ländereinrichtungen und (Landes-)Initiativen*. URL https://www.e-teaching.org/projekt/politik/laenderzentren. - abgerufen am 2020-08-24.

[SeNu20] Sexauer, Andreas ; Nussbaumer, Martin: Von der Präsenzuniversität zur temporären Online-Universität. In: *SCCnews - Corona-Sonderteil: Online-Lehre und Homeoffice am KIT* Bd. 2020 (2020), Nr. 01, S. 9–11

[SHWS16] Schulz, J. Chr. ; Hotzel, H. ; Wimmer, M. ; Schneider, G. ; Schulz, J. (Hrsg.): *Kooperation von Rechenzentren: Governance und Steuerung - Organisation, Rechtsgrundlagen, Politik*. Berlin : De Gruyter Oldenbourg, 2016

[WaHo20] Wagner, Nick ; Hornbostel, Stefan: *Wirkung messen - Die Evaluationsphase in Hochschulkooperationen*, *Future Lab Diskussionspapier* (Diskussionspapier Nr. 5). Essen : Stifterverband für die Deutsche Wissenschaft e.V., 2020

[WaNK20] Wagner, Nick ; Nieveler, Sebastian ; Kessler, Marte Sybil: *Clever konfigurieren: Hochschulkooperationen die geeignete Form geben*, *Future Lab Diskussionspapier* (Diskussionspapier Nr. 3). Essen : Stifterverband für die Deutsche Wissenschaft e.V., 2020

[WCEW20] Wagner, Nick ; Conradi, Silja ; Ebeling, Johanna ; Walzik, Sebastian: *Zusammenarbeit agil und kollaborativ gestalten*, *Future Lab Diskussionspapier* (Diskussionspapier Nr. 4). Essen : Stifterverband für die Deutsche Wissenschaft e.V., 2020

[West20] Westergom, Horst: Service Support zum Semesterstart. In: *SCCnews - Corona-Sonderteil: Online-Lehre und Homeoffice am KIT* Bd. 2020 (2020), Nr. 01, S. 13

[WKNW19] Wagner, Nick ; Kopp, Janina ; Nieveler, Sebastian ; Winde, Mathias: *Smart starten: Wie der Einstieg in Hochschulkooperationen gelingt*, *Future Lab Diskussionspapier* (Diskussionspapier Nr. 2). Essen : Stifterverband für die Deutsche Wissenschaft e.V., 2019

[WWND19] Winde, Mathias ; Wagner, Nick ; Nieveler, Sebastian ; Dauchert, Annett ; Kleimann, Bernd: *Kooperationsgovernance*, *Future Lab Diskussionspapier* (Diskussionspapier Nr. 1). Essen : Stifterverband für die Deutsche Wissenschaft e.V., 2019

[Zent20] Zentrale Datenschutzstelle der baden-württembergischen Universitäten: *Allgemeine datenschutzrechtliche Überlegungen zu Videokonferenzsystemen*. Stuttgart : ZENDAS, 2020.

Applying COBIT 2019 to IT Governance in Higher Education

Establishing IT governance for the collaboration of all universities and universities of applied sciences in Bavaria

Armin Gerl [1], Markus von der Heyde [2], Rainer Groß [3], Rainer Seck [4], Laura Watkowski [5]

Abstract: The Bavarian higher education environment is aiming to renew its IT strategy. The overall objective is to find an organisational solution which allows both local independence and collaborative solutions in those areas which are either commodities or which are too complex or costly to be solved several times in parallel. All Bavarian CIOs are engaged in the development of a model which respects local governance to be included into the overarching IT strategy of Bavaria, potentially Germany and Europe. As a common framework, COBIT was chosen to structure the process and guarantee completeness. The application of COBIT was started by the agreement to a common model. Further, the description of the respective responsibilities and competencies for all stakeholders was defined. We share this current state of the discussion with the broader community of higher education to promote further discussions about methods and objectives.

Keywords: IT governance; COBIT; higher education

1 Introduction

In industry, the Chief Information Officer (CIO) belongs to the top management level (C-level) of a company. He or she is responsible for strategy in the areas of information technology (IT) and computer systems in order to support the company's goals.

At the beginning of the 2000s, the debate on the need for overall responsibility for strategy in the areas of information technology (IT) and computer systems reached German universities and universities of applied sciences (jointly referred to as universities) [Bo02]. This was followed by the recommendations of the *German Research Foundation* (DFG) [De06; De10].

[1] University of Passau, Department of Distributed Information Systems, 94032 Passau, Germany, armin.gerl@uni-passau.de, https://orcid.org/0000-0001-9991-4539
[2] vdH-IT, Weimar, Germany, info@vdh-it.de, https://orcid.org/0000-0002-6026-082X
[3] Nuremberg Institute of Technology, Computer Science Department, 90121 Nürnberg, Germany, rainer.gross@th-nuernberg.de, https://orcid.org/0000-0002-6876-060X
[4] CIO CISO, University of Applied Sciences Munich, Department of Electrical Engineering and Information Technology, 80335 München, Germany, rainer.seck@hm.edu, https://orcid.org/0000-0001-7019-489X
[5] University of Bayreuth, Department of Information Systems Management, 95440 Bayreuth, Germany, laura.watkowski@uni-bayreuth.de, https://orcid.org/0000-0003-3706-0923

doi:10.18420/inf2020_47

The result was the introduction of the designation CIO at many universities, in an analogy to industry; this move was evaluated systematically in [He14; Ho15; Sc09].

[He18] aptly characterises the fact that although the designation 'CIO' is uniform, it says nothing about the actual role of the CIO at a university. As in industry, the CIO of a university is formally assigned the task of overall responsibility for strategy in the areas of information technology (IT) and computer systems for the university. However, an accepted uniform definition and description of competences of the role of the CIO at a university is still missing, despite the recommendations of the DFG.

As [He18] shows, this has led to the development of various CIO role models to date. The individually chosen form often is a mixture of factors from among the following options:

1. Strategic CIO with management functions (often included in the board of directors)
2. Strategic CIO as a staff position reporting to the board of directors
3. Operational CIO with direct responsibility for central IT supply
4. Collective CIO as a small decision-making group, in contrast to large Senate advisory commissions

In the past, the model of the CIO role mainly included local aspects of the individual university. The focus remained on the strategic alignment of the university's decision-making processes with the IT services [He18]. There was hardly any cross-university aspects included, since most of the collaboration was driven by the IT leadership on the operational level. Furthermore, the advantages of joining forces to further discussions at the political and ministerial levels has not yet been addressed by CIOs in general.

It is difficult for a single university CIO to achieve both, to both create synergies and exert influence. It stands to reason that this could probably be better organised collectively. How can such cooperation be successfully organised? Does a uniform model of a university CIO role have to be found first? And what about university central IT managers who have to translate strategies into operational business? How do you measure achieved synergies and influence?

Sustainable collaboration requires a definition of rules for collaboration, identification of the roles of all participants, and allocation of competencies and responsibilities [Su16]. It also should be guided by a shared common vision, an IT governance model for all participating universities.

In order to establish stable, efficient and effective IT services, cooperation and collaboration between universities is not just an option but a requirement [GPT11], [De06; De10; De16], [DS15], [He17]. To coordinate collaboration not only between scientists, but at the level of guaranteed services described ideally by service level agreements or similar contracts, universities also need to collaborate in the decision making for and steering of joint services.

Previous publications have called for IT governance to be applied at two levels: first at the level of the individual university, and second at the level of cooperation between universities [He16]. It had also been suggested that a model should be chosen that would ensure both individual freedom to organise local needs and the exchange of services across institutional boundaries. In order to reflect necessary decisions at the institutional level before adding the level between universities, we have to take into account the constitution of decision-making rights within universities. A statistical analysis of decision rights revealed typical constellations observed at German universities [HB18].

In the literature, various reference models and standards for the implementation of IT governance and IT compliance can be found. Some of these reference models deal with the performance aspect of IT governance. Other reference models primarily consider the compliance aspect of IT governance. The reference model *Control Objectives for Information and Related Technology* (COBIT 2019) offers a methodical approach by which, using a process reference model, governance processes within and between universities can be considered and recommendations for action for organisations can be derived. An agreement was reached on this approach for the derivation of recommendations for the responsibilities and tasks of those involved in IT governance within and between Bavarian universities. Using both approaches described in the previous paragraph in addition to the COBIT framework has led the Bavarian universities towards a collective IT service supply and demand model. It should be noted here that the same approach would not necessarily lead to the same result in another federal state.

Following this argumentation, we focus on the core research questions:

- Which method can be utilised to structure the definition of a collaborative IT governance model for Higher Education in Bavaria?
- How can the initial review completely cover the existing services, demand and collaborative structures?
- How can an overarching model of the CIO role be established? How can the common understanding of competencies and responsibilities be moderated between all stakeholders of the collaboration?

To answer these questions, the remaining of the paper is structured as follows: Section 2 introduces the current state of IT governance in Bavaria. Section 3 details how the reference model COBIT 2019 is used as a guideline to determine a common understanding of the Bavarian IT governance Model. This model is then detailed in Section 4, where roles with their responsibilities and competencies are defined. Section 5 takes the key findings from a specific view to a more generally applicable level before the work is concluded in Section 6.

2 State of the Art within Bavaria

As of 2020, IT governance structures with a CIO or a CIO board have been introduced at all universities in Bavaria in accordance with their statutes. However, the CIO function has not yet been fully anchored as a department within the university management. In some cases, the office of CIO is linked to the department of a vice president; in others, the CIO is a permanent member of the university management (but not linked to a vice president's office), or the CIO is anchored in the university administration or as a staff unit and reports directly to the university management. An analysis of Germany-wide CIO models by Von der Heyde [He18] finds a total of 7 models: operative CIO, collective CIO (CIO board), three versions of a strategic CIO (with and without membership on the board of directors), Chief Digital Officer (CDO), or a mix of these models. In this context, the analysis in [He18] advocates a consistent development of responsibilities and tasks within the IT governance structures with regard to technical developments, digital challenges and the demands of society on universities.

Since 2010, the universities have devoted themselves to defining the responsibilities and tasks within their IT governance structures at varying speeds. At all Bavarian universities, the CIO or CIO board has the task of determining the fundamental issues of IT deployment and the digitisation goals at the university. The CIO proposes how to and what resources are needed to achieve these goals. In this context, Hechler and Pasternack [HP17] discuss possible digitisation strategies and policies at universities.

Together with the StMWK (Bavarian State Ministry of Sciences and Arts), the Bavarian universities have defined a series of digitisation targets in the 2022 university development plan, which are implemented in the current individual target agreements with the universities. The Bavarian State Ministry is an important stakeholder in the IT governance framework. The 2010 IT strategy of the Bavarian universities [CI10] also laid the foundation for cooperation across university borders. The institutional exchange among CIOs has been successfully established in the respective CIO boards of the University of Bavaria e.V. and the Bavarian University of Applied Sciences e.V., as recommended in the 2010 strategy.

3 Methodology

In four meetings, the editorial focus group, set up by the CIO boards of the universities and universities of applied sciences in Bavaria, collaboratively discussed the need for a Bavarian-wide IT governance model. The aim of the editorial focus group is to institutionalise the existing informal exchange between CIOs in order to improve adequate digitisation across Bavaria and to create added value for all stakeholders through synergies. In order to achieve this, the editorial focus group proposes a IT governance model, which describes the tasks and the interaction of those responsible. In addition, the cooperation will ultimately lead to Bavaria-wide IT service centres. The improvement of the Bavaria-wide digitisation, fostering of synergies, and engagement of stakeholders is in progress and already partially

implemented. The synergies will ultimately be created by cooperation with Bavaria-wide *Cooperative IT Service Providers*, which are already partly set up.

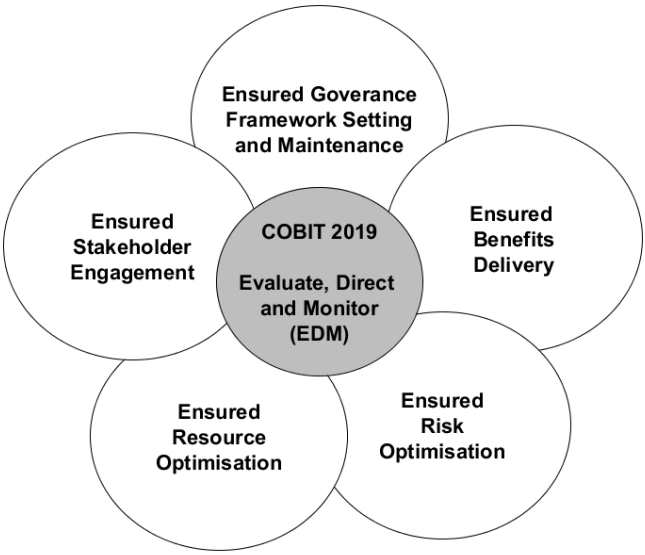


Fig. 1: Focused topics derived from COBIT 2019 were used as guidelines in the interview.

In the literature, diverse reference models and standards for the implementation of IT governance and IT compliance exist [JG07]. But in higher education only a few implementations of frameworks are known [FL09]. Implementations in Germany are rare, an example is [KIT17]. Overall, some reference models deal with the performance aspect of IT governance, while other reference models primarily consider the compliance aspect of IT governance. The reference model *Control Objectives for Information and Related Technology (COBIT 2019)* offers a guideline in which governance processes have to be addressed. These guidelines were utilised to conduct a qualitative survey, i.e., structured interviews (n=6) with a total of 85 distinct tasks, with CIOs and Heads of IT in Bavaria. The interviews were structured according to the governance *Domain* of 'evaluate, direct and monitor' (EDM; see Fig. 1). The COBIT 2019 *Objectives* used for the interviews were *Ensured governance Framework Setting and Maintenance*, *Ensured Benefits Delivery*, *Ensured Risk Optimisation*, *Ensured Resource Optimisation* and *Ensured Stakeholder Engagement*, for which each the corresponding *Practices* was discussed in the context of the Bavarian IT governance by the interviewees. Based on the interviews, recommendations for the responsibilities and tasks of those involved in IT governance within and between Bavarian universities have been derived. Thus, the diverse perspectives of IT governance were structured by using COBIT 2019.

4 A Proposed Bavarian Higher Education IT Governance Model

The results of the survey were summarised, structured and discussed within the editorial focus group, which resulted in a proposed IT governance model for the Bavarian higher education system (see Fig. 2). Along with the use of COBIT 2019, the principles of *Demand*- and *Supply*-IT are considered Wulf et al. [WWB12]. Accordingly, all strategic tasks of the university's IT organisation would fall within the scope of *Demand*-IT, while *Supply*-IT encompasses the management tasks within the university and *Cooperative IT Service Providers*.

The next section shows a detailed list of tasks for each role, e.g., *Universities* or *Stakeholders*, and their interaction and communication with each other. The survey showed that in particular, the integration of the university CIOs of Bavaria, *CIO-Boards*, *Demand*- and *Supply-Boards* and *Cooperative IT Service Providers* can offer sustainable added value and synergy effects for Bavarian IT governance.

4.1 Governance Interaction

To ensure a consistent and efficient implementation of IT requirements and their optimisation, governance relationships between and within the clusters *Stakeholders*, *CIOs Bavaria*, *Universities* and *Cooperative IT Service Providers* were defined.

In order to address higher-level strategic IT requirements by the *Stakeholders*, there are governance relationships with the *CIOs Bavaria* ❶ on the one hand and with the *Universities* ❸ on the other hand . The former ❶ also serves to inform *CIOs Bavaria* about strategic IT requirements in a timely manner so that they can develop meaningful and cooperative proposed solutions by using the respective boards of *CIOs Bavaria*.

The governance relationship between *CIOs Bavaria* and *Universities* ❷ intends to identify cross-university synergy effects and efficiently solve common challenges. The *Demand Board* of every university regulates the exchange between the *Board of Directors* and the respective university organisation ❹. The goal is to identify synergies within the universities and to prioritise IT projects.

On the basis of the directives from the governance relationships with *Stakeholders* and *Universities* ❸, *CIOs Bavaria* regulates the demand-oriented development of cross-university IT services in close exchange with *Cooperative IT Service Providers* ❽.

The governance relationship between the *Board of Directors* and the university's *Central IT* ❺ is intended to ensure the adequate development of internal IT services and the efficient use of the university's IT resources. In addition to the governance relationships, there are operational relationships in *Supply*-IT between *Cooperative IT Service Providers*, the universities' *Central IT* ❼ and individual decentralised IT organisations ❻ in order to fulfil

day-to-day business. This includes operational coordination between *t*he Demand Board and *Supply Board* ⑩.

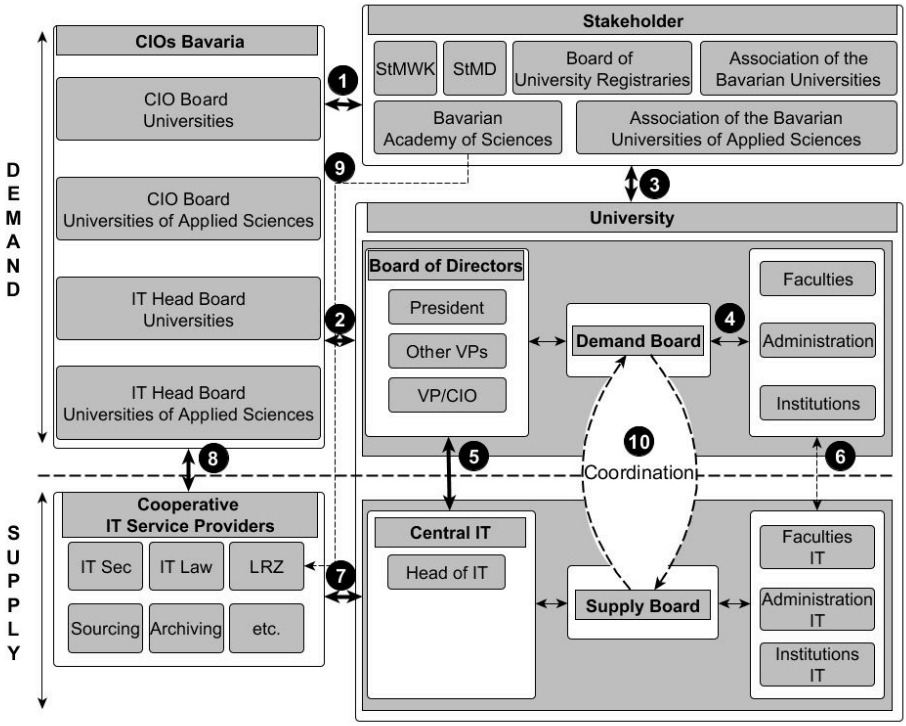


Fig. 2: Proposed IT governance model for the Bavarian higher education system. The numbered governance interaction levels are referred to within the text.

4.2 Governance Role Descriptions

Within the proposed IT governance model, the *Stakeholders* are comprised of Ministries, e.g., *StMWK* and *StMD*, and strategic institutions for higher education in Bavaria, i.e., *Association of Bavarian Universities* and *Association of Bavarian Universities of Applied Sciences*, *Board of University Registraries*, and *Bavarian Academy of Sciences*. The *Stakeholders* define and communicate the framework conditions for the *Universities* ❸ and *CIOs Bavaria* ❶, e.g., announcing new laws and regulations, or applying for funds from the Bavarian state parliament.

The *CIOs Bavaria* association is comprised of the *CIO and IT Head Boards of Universities and Universities of Applied Sciences*, which is envisioned to be established. The *CIOs Bavaria* has an advisory function for *Stakeholders*, *Universities* and *Cooperative IT Service*

Providers. The association acts as an information body, an advisory body, and a focus group for CIO concerns. Besides these roles, the *CIOs Bavaria* acts as a driving force for innovative IT and digitisation ideas and projects by collecting the Bavarian universities' *Demands*, evaluating them, and proposing appropriate *Cooperative IT Service Providers* for implementation, e.g., by applying for starting funds from the *Stakeholders* ❶.

Each *University* as an institution makes its own IT governance decisions. It defines the strategic orientation of IT in its university development plan and establishes a continuous improvement process.

Within each *University*, the *President* ultimately is the decision-making authority at the university. He or she commissions and decides on the creation of IT guidelines and also issues a set of basic rules that define the tasks of the CIO. The university *President* decides on strategic IT and digitisation goals and provides sufficient resources to achieve them. In particular, he or she bears responsibility for IT risks (IT security, data protection, etc.) and decides on the procedure for dealing with them, e.g., by establishing IT risk management systems. Furthermore, the *President* comes to agreements to achieve certain strategic objectives with the science ministry, i.e., *StMWK*, is decided ❸ and the implementation of the agreed measures, e.g., with internal target agreements, are controlled.

The CIO acts as an intermediary between university stakeholders (*Faculties*, *Administration*, *Institutions*, and *Central IT*). The CIO, in any form [He18], must be represented on the *Board of Directors*[6] and has the exclusive right to propose all IT related initiatives. In addition, the CIO coordinates the university's *Demand* management of the *Demand Board*, which is also comprised of representatives of the university stakeholders ❹. Part of the CIO's tasks is to represent the university on relevant committees (*CIO Bavaria* and *CIO Board*) ❷, such that *Demand* requirements are shared Bavaria-wide. The CIO proposes and further develops IT relevant organisational structures. Furthermore, the CIO is responsible for updating the university's IT strategy and participates in the design of internal and external target agreements.

The *Demand Board* develops and recommends strategic IT and digitisation goals of the university. All IT requirements throughout the university are collected, evaluated and prioritised based on the relevant university-internal strategies. The *Demand Board* is in close coordination with the *Supply Board* ❿.

The *Head of IT* is responsible for the *Central IT* of the university. The *Central IT* ensures the secure and sustainable operation and provision of the university's IT infrastructure. The central IT department is responsible for the use and procurement of IT and IT services; it can delegate parts of the tasks to decentralised IT departments of the universities' stakeholders.

[6] Putting the CIO in close proximity to the board of directors within each individual university refers back to the recommendation of [De06] and following publications. However, in reality, a multitude of CIO forms have been established, as shown by [He18]. The discussion and adoption of the agreed model of the role needs to be implemented over the upcoming months/years.

Furthermore, the *Head of IT* is a member of the Bavaria-wide *IT Head Board* ❷, which fosters an exchange on the IT management level.

The *Supply Board* controls the implementation of the actions decided upon by the university and the operational coordination of IT. It consists of the *Head of IT* as well as representatives of the universities' stakeholders IT ❻, such that both centralised and decentralised IT is managed. A close coordination with the *Demand Board* ❿ is established.

Cooperative IT Service Providers aggregate standardised tasks, e.g., those which do not require a *University* specific specialisation. They provide cross-university IT services ❼. IT services can be of both an operational and a conceptual nature. The specifications are defined and updated by *CIOs Bavaria* in coordination with the *Stakeholders* ❶. The implementation of the *Cooperative IT Service Providers* is recommended throughout Bavaria with the involvement of the Ministry (StMWK). Established *Cooperative IT Service Providers* regularly provide information on the use of their inter-university services to the *CIOs Bavaria* ❽. Examples of already established or planned *Cooperative IT Service Providers* are:

- Staff Unit Information Security of Bavarian Universities of Applied Sciences and Universities [Ho20]
- Staff Unit IT Law, License Management, E-Procurement [Ju20a]
- Cloud Storage Services (Sync and Share (LRZ) [Le20] and FAUbox (RRZE) [Re20])
- Maintenance, expansion and digitisation of cooperative IT procurement [Ju20b]
- Digital Archiving by the DIMAG Coordination Office [Un19]
- Coordination office and IT service centre for data processing in the university administration for the universities of applied sciences in Bavaria (KDV) [Ko13]

In Bavaria, the Leibniz-Rechenzentrum (LRZ) offers cross-university services like other IT service centres. However, the LRZ has its own management and is subject to the control structure of the *Bavarian Academy of Science* ❾.

5 Discussion

The findings derived from the detailed process support the previously mentioned findings of [HB18], and further develop them by not only considering the institutional level of the university, but the higher education environment of the state, i.e., Bavaria. The federal state level has been omitted in this model, because the critical decision-making boundary for IT governance is on the level of the state, like Bavaria. On the *Supply*-level, the definition and establishment of *Cooperative IT Service Providers* is a desired outcome to establish synergy

effects and and sustainable IT across university boundaries. Primarily, these services, which are negotiated and recommended by the *CIO Boards*, should be provided by university *Central IT* on a 'One for all' principle.

Visible results of the initiative in the application of COBIT are:

- The intense discussion within the established CIO boards of the universities and universities of applied sciences in Bavaria led to an increased level of trust between the collaborating partners with their commitment to a joint collaboration.
- As other federal states (e.g., Baden-Wuerttemberg, Thuringia) have suggested with their models, integrating the ministry and the *Cooperative IT Service Providers* into the discussion seems a crucial success factor.
- A consistent model of the CIO role (based on COBIT, refined by interviews) was finally agreed upon and constitutes the baseline of mutual understanding of competencies and responsibilities on the individual and shared levels.
- Both extant and newly established *Cooperative IT Service Providers* started to provide services based on the mutual agreement of a joint need.
- The application of the model has so far guaranteed a constant, structured and visible progress.

Bootstrapping collaboration required a joint initiative, which is commonly a reaction to external influence. The project gained momentum at the stage where the visibility across the majority of the participating institutions was high enough to focus attention, even at the political level, on the ongoing process.

This model is neither complete nor validated to its final implementation. Due to the nature of step-wise implementation of collaborative approaches, it might provide insights and document a good practice.

6 Conclusion

The participation of all stakeholders in the development of the proposed IT governance model together with the application of the COBIT 2019 approach is a prerequisite for the Bavaria-wide implementation of the model. However, the proposed model lacks aspects related to the dynamics between *Demand and Supply* on one hand and IT projects on the other. Often, sudden insights from the business operations drive the strategic discussions in the short term and vice versa. The strategic decision-making process is rather dynamically interwoven with the insights gained from the business operations. A further aspect that has been neglected is the importance of IT projects, which has been defined in [He16], and their organisational embedding and visibility in the overall IT governance. IT projects as an

organisational form are best practice in the implementation and testing of new IT strategies, and thus have a direct influence on the demand and supply levels in terms of content and personnel. A core question remains: how are project-driven activities embedded into the implementation of the IT governance model? As Ross et al. [RWR06] pointed out, this also depends on the enterprise architecture model of the Bavarian universities as a group, which in fact maybe more similar to the structure of a holding company.

The application of COBIT 2019 to the higher education sector seems promising. Central aspects of IT governance are able to couple autonomy with central aspects of collaborative IT service management. How the currently proposed model reacts to overall change remains open. The requirement for risk management, change management and other frameworks will be more obvious once the effect of the coupling between universities is visible. Changing the core - the culture of universities - towards a professional management will prove to be a key asset.

7 Acknowledgements

We thank all participating CIOs, CIO Boards, the distinct editorial team of the Bavarian IT Strategy 2025. This research is supported by the funding of the *Bayerisches Staatsministerium für Wissenschaft und Kunst (StMWK)*. We also thank the reviewers for valuable input to the manuscript.

References

[Bo02] Bode, A.: Universitat im Wandel: Die Rolle des CIO bei der Erneuerung der Prozesse. Information Consulting & Management 17/, pp. 43–47, 2002, URL: http://gcc.uni-paderborn.de/www/WI/WI2/wi2_lit.nsf/35ae96bebc983d53c12573e70058bbb2/458d4f298d370fe4c1256c8e002e5914/%5C$FILE/CIO-unis.pdf.

[CI10] CIOs of Bavaria: IT-Strategie der bayerischen Hochschulen - Leitlinien für die Weiterentwicklung der IT-Infrastruktur, Last accessed: 13.07.2020, 2010, URL: https://w3-mediapool.hm.edu/mediapool/media/zak_2/lokal_zak/formulare_4/it_strategie_hochschulen_bayern_2010.pdf.

[De06] Deutsche Forschungsgemeinschaft: Informationsverarbeitung an Hochschulen - Organisation, Dienste und Systeme: Empfehlungen der Kommission für Rechenanlagen für 2006-2010, Bonn, Germany: Deutsche Forschungsgemeinschaft, Oct. 2006, URL: http://www.dfg.de/download/pdf/foerderung/programme/wgi/wgi_kfr_empf.pdf.

[De10] Deutsche Forschungsgemeinschaft: Informationsverarbeitung an Hochschulen - Organisation, Dienste und Systeme: Empfehlungen der Kommission für IT-Infrastruktur für 2011-2015, Bonn, Germany: Deutsche Forschungsgemeinschaft, 2010, URL: http://www.dfg.de/download/pdf/foerderung/programme/wgi/empfehlungen_kfr_2011_2015.pdf.

[De16] Deutsche Forschungsgemeinschaft: Informationsverarbeitung an Hochschulen - Organisation, Dienste und Systeme: Empfehlungen der Kommission für IT-Infrastruktur für 2016-2020, Bonn, Germany: Deutsche Forschungsgemeinschaft, Mar. 2016, URL: http://www.dfg.de/download/pdf/foerderung/programme/wgi/kfr_stellungnahme_2016_2020.pdf.

[DS15] Degkwitz, A.; Schirmbacher, P., eds.: Informationsinfrastrukturen im Wandel. Informationsmanagement an deutschen Universitäten. BOCK + HERCHEN Verlag, 2015, 390 pp., ISBN: 978-3-88347-254-6, URL: http://www.dini.de/fileadmin/docs/DINI_Informationsinfrastrukturen.pdf, visited on: 08/20/2016.

[FL09] Fernández, A.; Llorens, F.: An IT Governance Framework for Universities in Spain. In. EUNIS 2009. 2009, URL: http://www.gti4u.es/pdf/an_it_governance_framework_for_universiti%20es_in_spain.pdf, visited on: 08/20/2016.

[GPT11] Görl, S.; Puhl, J.; Thaller, M.: Empfehlungen für die weitere Entwicklung der Wissenschaftlichen Informationsversorgung des Landes NRW. epubli GmbH, Berlin, 2011, ISBN: 978-3-8442-0694-4.

[HB18] von der Heyde, M.; Breiter, A.: Factorial analyses in IT governance reveal constellations of decision shares and their consequences on IT in higher education institutions. EUNIS Journal of Higher Education/, 2018, URL: https://www.eunis.org/download/2018/EUNIS_2018_paper_35.pdf.

[He14] von der Heyde, M.: CIOs und IT-Governance an deutschen Hochschulen, Dec. 2014, URL: https://www.zki.de/fileadmin/zki/Publikationen/ZKI_CIO-Studie_final.pdf, visited on: 08/20/2016.

[He16] von der Heyde, M.: Replikations- und Diversifikationsmodelle für IT-Governance in Hochschulverbünden. In: Kooperation von Rechenzentren Governance und Steuerung - Organisation, Rechtsgrundlagen, Politik. De Gruyter, Berlin, Boston, pp. 303–309, 2016, ISBN: 978-3-11-045975-3, URL: https://doi.org/10.1515/9783110459753-025, visited on: 11/01/2016.

[He17] von der Heyde, M.; Auth, G.; Hartmann, A.; Erfurth, C.: Hochschulentwicklung im Kontext der Digitalisierung - Bestandsaufnahme, Perspektiven, Thesen. In: INFORMATIK 2017. GI Jahrestagung 2017. Vol. 275. Lecture Notes in Informatics (LNI) - Proceedings, ISSN 1617-5468, Chemnitz, Germany, pp. 1757–1772, Sept. 28, 2017, ISBN: 978-3-88579-669-5.

[He18] von der Heyde, M.: Mehr oder weniger CIO - Überblick zu CIO-Formen an deutschen Hochschulen. In: INFORMATIK 2018. GI Jahrestagung 2018. Vol. Lernen und Arbeiten im Wandel 2018. Lecture Notes in Informatics (LNI) - Proceedings, ISSN 1617-5468, Springer, Berlin, Germany, Sept. 24, 2018, ISBN: 978-3-88579-669-5.

[Ho15] Hotzel, H.; Lang, U.; Wimmer, M.; von der Heyde, M.: CIOs at German Universities–a Survey by ZKI. European Journal of Higher Education IT 3/2015, Oct. 2015, ISSN: 2519-1764, URL: http://www.eunis.org/download/2015/papers/EUNIS2015_submission_54.pdf, visited on: 08/20/2016.

[Ho20] Hochschule Augsburg: Stabsstelle Informationssicherheit bayerischer Hochschulen und Universitäten, Last accessed: 01.09.2020, 2020, URL: https://www.hs-augsburg.de/Rechenzentrum/Stabsstelle-Informationssicherheit.html.

[HP17] Hechler, D.; Pasternack, P.: Digitalisierungstrategien und Digitalisierungspolicies an Hochschulen. Die Hochschule: Journal für Wissenschaft und Bildung 26/2, pp. 84–105, 2017.

[JG07] Johannsen, W.; Goeken, M.: Referenzmodelle für IT-Governance. d-punkt, Heidelberg, 2007.

[Ju20a] Julius-Maximilians-University of Wuerzburg: IT-Recht, Last accessed: 01.09.2020, 2020, URL: https://www.rz.uni-wuerzburg.de/dienste/it-recht/.

[Ju20b] Julius-Maximilians-University of Wuerzburg: WebShop-Anleitung, Last accessed: 01.09.2020, 2020, URL: https://www.rz.uni-wuerzburg.de/dienste/shop/webshop-anleitung/.

[KIT17] IV-Governance-FrameworkGovernance der digitalen Informationsverarbeitung und -versorgung des KIT, 2017, URL: https://www.kit.edu/downloads/cio/IV-Gov-Framework.pdf, visited on: 07/13/2020.

[Ko13] Koordinierungsstelle für Datenverarbeitung in der Hochschulverwaltung an den staatlichen Fachhochschulen in Bayern: Die Koordinierungsstellefür die Datenverarbeitung in der Hochschulverwaltung an den staatlichen Fachhochschulen in Bayern (KDV), Last accessed: 01.09.2020, 2013, URL: https://www.kdv-fh-bayern.de/cms/.

[Le20] Leibniz Supercomputing Centre of the Bavarian Academy of Sciences and Humanities: Leibniz Supercomputing Centre, Last accessed: 01.09.2020, 2020, URL: https://www.lrz.de/english/.

[Re20] Regional data center Friedrich-Alexander-University of Erlangen-Nuremberg: FAUbox, Last accessed: 01.09.2020, Sept. 2020, URL: https://www.rrze.fau.de/serverdienste/server/faubox/.

[RWR06] Ross, J. W.; Weill, P.; Robertson, D.: Enterprise Architecture As Strategy: Creating a Foundation for Business Execution. Harvard Business Review Press, Boston, Mass, 2006, ISBN: 978-1-59139-839-4.

[Sc09] Schwabe, G.: IT-Governance an Universitäten in Deutschland, Schweiz und Österreich. VM Verwaltung & Management 15/6, pp. 317–325, 2009, URL: http://www.zki.de/fileadmin/zki/Archiv/tagungen/cottbus-2010/02_Schwabe_IT-Governance_VM_Teil_2.pdf, visited on: 08/20/2016.

[Su16] Suchodoletz, D. v.; Schulz, J. C.; Leendertse, J.; Hotzel, H.; Wimmer, M.: Kooperation von Rechenzentren: Governance und Steuerung - Organisation, Rechtsgrundlagen, Politik. Google-Books-ID: CXCtDgAAQBAJ, Walter de Gruyter GmbH & Co KG, 2016, ISBN: 978-3-11-045975-3.

[Un19] University of Regensburg: Investition in die Infrastruktur, Last accessed: 01.09.2020, 2019, URL: https://www.ur.de/pressearchiv/pressemitteilung/1017752.html.

[WWB12] Wulf, J.; Winkler, T. J.; Brenner, W.: Organisationsgestaltung der Demand-IT. In (Goltz, U.; Magnor, M.; Appelrath, H.-J.; Matthies, H. K.; Balke, W.-T.; Wolf, L., eds.): INFORMATIK 2012. Gesellschaft für Informatik e.V., Bonn, pp. 746–758, 2012.

Die Sprache «SemaLogic» als semantische Repräsentation

Eine anforderungsbasierte Sprache zur Modellierung von Prüfungsordnungen und Abbildung von Studienverläufen

Markus von der Heyde [1], Matthias Goebel[2]

Abstract: Eine wesentliche Aufgabe im Zuge der Digitalisierung der Verwaltungsprozesse einer Universität ist die Abbildung der Prüfungs- und Studienordnungen (PStO) im Campus-Management-System (CMS). Denn nur wenn die in diesen Ordnungen enthaltenen Regeln logisch auswertbar sind, können sie Abläufe im Student-Life-Cycle-Prozess steuern. Dieser Beitrag schlägt eine drastische Vereinfachung als technische Lösung vor. Auf Basis einer semantischen Repräsentation, die direkt aus den Ordnungen erstellt werden kann, wird die menschlich lesbare Form der Ordnung zum „Programmcode" und kann direkt Einfluss auf die Umsetzung im CMS nehmen.

Für die Erzeugung einer solchen sprachlichen Übersetzung, welche die notwendige Flexibilität, aber eindeutige Repräsentation erzeugen kann, wurden allgemeine Anforderungen aus der Literatur sowie 20 Studienordnungen analysiert. Die getroffene Auswahl deckt dabei alle Bundesländer, den Großteil der Hochschulformen (Volluniversitäten, Technische Universität, Hochschulen für angewandte Wissenschaften, Musik- und Kunsthochschulen), ein breites Fächerspektrum und vier verbreitete Campus-Management-Systeme ab. Im Weiteren wird aufgezeigt, warum die auf dieser Grundlage entwickelte Sprache SemaLogic die dargelegten Anforderungen erfüllt und dass der Entwicklungsstand der darauf basierenden prototypischen Umsetzung bereits in der Lage ist, die semantische Repräsentation zu erzeugen und logische Prüfungen durchzuführen. Ein Ausblick auf die potentiellen Anwendungsfälle schließt den Beitrag ab.

Keywords: Logische Sprache, Semantik, Studienordnung, Prüfungsordnung, Modellierung

1 Einleitung

Es gibt eine Reihe von organisatorischen Problemstellungen, wenn Hochschulen und Universitäten neue Campus-Management-Systeme etablieren bzw. diese im Zuge der Digitalisierung ihrer Prozesse intensiver nutzen. Eine systematische Analyse der Erfolgsfaktoren derartiger Entwicklungen wurde von Schreiter et al. [SAA12] vorgenommen. Gleichzeitig ist die Modellierung von geeigneten Datenstrukturen, die in technischen Systemen nachfolgend umgesetzt werden, eine Herausforderung, die seit mehr als 25 Jahren die Hochschulen begleitet [SM95; BCS13]. In einer Vielzahl von Projekten wurde insbesondere die Unsicherheit bei der Abschätzung vom erforderlichen Aufwand für die technische Abbildung

[1] SemaLogic, Weimar, Germany, markus.von.der.heyde@semalogic.de, https://orcid.org/0000-0002-6026-082X
[2] SemaLogic, Weimar, Germany, matthias.goebel@semalogic.de

doi:10.18420/inf2020_48

der Prüfungs- und Studienordnungen (PStO) im Campus-Management-System (CMS) untersucht [Ba11].

Logische Sprachen bzw. alle prozeduralen Sprachen repräsentieren Wissen, Abläufe oder Zusammenhänge in digitaler, durch den Computer ausführbarer Form. Um Regelwerke und Abläufe aus natürlichsprachlichen Texten erzeugen zu können, müssen diese bisher zunächst fachlich interpretiert und in einem zweiten Schritt programmiert werden. Ob eine in aktuellen CMS erzeugte Programmierung mit dem ursprünglichen Text übereinstimmt, kann de facto nicht geprüft werden.

Habtemariam untersuchte bereits 2006 die Anwendbarkeit von logischen Sprachen auf die Modellierung von PStO [Ha06]. Die gewählte Sprache glich aber in der Komplexität einer üblichen Programmierung und war nur für Experten auf dem Gebiet der Studienorganisation sowie der verwendeten Programmiersprache geeignet. Zudem war die Laufzeit der Prüfung der abgebildeten Ordnung auf Konsistenz ungeeignet, um interaktiv Regeln verändern zu können und Optimierungen während der Designphase der Ordnung berücksichtigen zu können. Die semantische Modellierung und Simulation von PStO in der Sprache CLP erfolgte in [Te08], führte aber nicht zur einer Umsetzung in den verwendeten CMS, da auch hier die Hürde der Programmierung eine gleichzeitige Verwendung von Ordnungstext und semantischer Repräsentation der Ordnung verhinderte. Wie würde sich die Welt aber verändern, wenn der Computer einen natürlichsprachlichen Text „verstehen“ und die darin enthaltenen Regeln und Anweisungen klar und verlässlich abbilden würde? Könnte der Computer dann bereits während der Formulierung textliche Unklarheiten erkennen und korrigieren und so den Menschen unterstützen? Um diese Unterstützung möglich zu machen, ist es notwendig, dass der im Text enthaltene Sinn, also die Bedeutung der Worte, in beiden Interpretationen übereinstimmt. Eine neutrale semantische Repräsentation müsste hierzu aus einem Text verlässlich in ein technisch lesbares und semantisch identisches Abbild überführt werden. Wenn dies gelänge, kann die Interpretation und die anschließende Programmierung – wie sie beispielsweise bei der Abbildung von PStO im CMS stattfinden muss – in wesentlichen Teilen entfallen. Es würden damit potentielle Fehler während der Erzeugung des Textes, der Interpretation bis hin zur Kodierung ausgeschlossen, solange die schriftliche Form eines Textes wie einer PStO tatsächlich die intendierten Regelungen enthält.

2 Ableitung von Anforderungen aus der Literatur

Zwei grundsätzliche Lösungen existieren, damit aus einem Text der PStO eine logische, semantische Struktur z.B. der Modulreihenfolge abgeleitet werden kann: a) Die Disziplin des Natural Language Processing leitet mithilfe von Heuristiken aus der natürlichen Sprache eine wahrscheinlich gemeinte Struktur ab [KM93], als Review siehe [Ar14]. b) Eine Sprachdefinition formuliert eineindeutig logische Operatoren bzw. formale Aussagen (z.B. Constraints) und durch einen Parser werden die Sprachkonstrukte in die semantische Repräsentation überführt [Ha06; Te08]. Otto und Anton [OA07] kategorisierten diese

grundsätzlichen Ansätze feiner in ein ganzes Spektrum von hybriden Ansätzen, welche sich primär durch die technische Realisation und die Eigenschaften der vorgeschlagenen spezialisierten Sprachkonstrukte unterscheiden. Für alle Fälle ist es zunächst aber essentiell, die in tatsächlichen Texten vorkommenden logischen Operatoren zu klassifizieren und in der Zielrepräsentation die Vollständigkeit sicherzustellen.

Bereits 1957 formulierte Allen in der Dissertation [Al57] sechs grundlegende logische Operatoren, die in juristischen Texten maßgeblich präsent sind:

- Conjunction – das logische *UND*, welches alle Bestandteile zusammenbindet
- Exclusive disjunction – das ausschließliche *ODER*, dem genau jeweils eine von mehreren Alternativen gewählt werden kann
- Inclusive disjunction – ein *ODER*, bei dem eine oder mehrere Wahloptionen gewählt werden können
- Negation – das *NICHT*, bei der die Aussage negiert wird
- Implication – als Folgerung, wenn aus einer Aussage eine zweite folgt
- Coimplication – die gegenseitige Folgerung, bei der aus dem jeweiligen Wahrheitswert einer Aussage die jeweils andere abgeleitet werden kann.

Allen formulierte eine neue Schreibweise der Anordnung von Aussagen, in der diese Verbindungen auf natürliche Texte angewendet werden sollten (systematically-pulverized form), damit die von Juristen formulierten Aussagen symbolisch eindeutig würden.

Peek geht in [Pe97] darüber hinaus, in dem die Informationsstruktur, die den Aussagen unterliegt, klassifiziert wird. Den Worten wird dabei innerhalb eines Satzes eine grammatikalische Bedeutung zugewiesen, welche die Funktion der Worte im Satzgefüge eindeutig festlegt. Die Analyse der Sätze erfolgt durch denjenigen, der die Transkription des ursprünglichen juristischen Textes vornimmt. Tetzner und Riedewald schlagen ferner eine Anwendung von symbolischer Simulation zur Prüfung zeitlicher Abhängigkeiten in Regelwerken wie PStO vor [TR05]. Der dort formulierte Ansatz erweitert endliche Automaten um die Auswertung der zeitlichen Regelwerte und ermöglicht, dass komplexe Zusammenhänge problemadäquat modelliert werden können.

Habtemariam wendete in seiner Dissertation [Ha06] die Re-Formulierung der PStO Texte als Constraints an, um mit existierenden Werkzeugen den formalen Studienablauf als Constraint-Erfüllbarkeitsproblem (Constraint Satisfaction Problem, CSP) bzw. Constraint Optimierungsproblem darzustellen und zu lösen. Als Notation der Constraints werden dazu die endlichen Wertemengen der verwendeten symbolische Variablen und deren Zusammenhänge formuliert. Diese Zusammenhänge sind ungerichtete, also bi-direktionale, Beschreibungen, welche potentiell auch nur in partiellen Bedingungen (z.B. im mathematischen Sinne) zwischen zwei Variablen bestehen. Die Auflösung komplexer Aussagen

mit mehreren Variablen in einfache, welche maximal zwei Variablen besitzen, wurde durch grundlegende Arbeiten zur Programmierung mit Constraints (Constraint Logic Programming - CLP) gezeigt.

In ihrer Dissertation erweiterte Tetzner [Te08] die Constraint Simulation der CLP durch die nutzerfreundlichen Sprachen MODEL-HS und VYSMO, welche die Formulierung von hybriden endlichen Automaten unter der Verwendung von zeitlichen Constraints erlaubte. Die Mächtigkeit der Aussagen blieb aber durch die grundlegenden Aussagen der CLP konstant. Der nutzerfreundliche Text der MODEL-HS gleicht weiterhin eher einer Programmiersprache als einer natürlichen Sprache, erlaubt aber im Gegensatz zur natürlichen Sprache eine eindeutige Formulierung der intendierten Aussagen. Otto und Anton [OA07] formulierten ebenfalls die Anforderung, dass im Fall einer Prüfung der Bedingungen aus den Regelungstexten im Sinne von Audits eine Rückverfolgung von Aussagen auf deren Quelle unerlässlich sei.

Spitta und Mordau [SM95] schlugen eine konsistente Bezeichnung der Strukturelemente von PStO-Begriffen im Sinne einer Ontologie vor, welche durch Brune et al. [BCS13] weiter entwickelt und umgesetzt wurde. Sicherlich ist für die Austauschbarkeit zwischen Hochschulen ein gemeinsames Verständnis der Begriffe essentiell; für die Prüfung der Konsistenz eines Regelwerkes ist diese allgemeine Ebene nicht erforderlich.

Zusammenfassend kann aus der Literatur geschlossen werden, dass zur Formulierung von juristischen Texten, welche eine Struktur formulieren, folgende logische Operatoren zur Gliederung der Anforderungen notwendig sind:

- *UND* zur Verbindung von Aussagen
- *ODER* als aus- und einschließende Variante
- *NICHT* zur Negation von Aussagen
- *BEDINGT* zur Formulierung von Implikationen bzw. Äquivalenz
- *VOR/NACH* zur zeitlichen Abgrenzung
- *FOLGT AUS* zur Angabe der Quelle
- *BASIERT AUF* als Möglichkeit der Versionsangabe

Eine semantische Repräsentation, die konkret dazu geeignet sein soll, PStO strukturell semantisch abzubilden, muss diese bzw. äquivalente Sprachkonstrukte anbieten.

3 Ableitung von Anforderungen aus PStO

Aus existierenden PStO in Deutschland kann stichprobenartig eine Liste von Anforderungen erstellt und die aus der Literatur abgeleitete Liste der logischen Operatoren dagegen geprüft

werden. Dazu ist es wünschenswert, eine möglichst breite Varianz von PStO mit Stichproben aus verschiedenen Hochschulformen, Fächern und Rechtsräumen zu analysieren. Nicht in die Analyse einbezogen sind Anforderungen, die sich außerhalb eines Regelwerkes zur Abbildung logischer Operatoren befinden. Typische Beispiele dafür sind organisatorische Voraussetzungen oder Zusammenhänge, die bereits durch einfache Datenfelder innerhalb vorhandener CMS abgebildet werden. Dies betrifft insbesondere konkrete Prüfungsregularien (z.B. Anwesenheit von einer bestimmte Anzahl von Prüfern), Fristen zur Anmeldung und Prüfungsdurchführung sowie grundsätzliche Zugangsvoraussetzungen.

3.1 Darstellung der Auswahl von PStO

In einer grundlegenden Analyse wurden insgesamt 20 PStO auf Basis der Kriterien Hochschulform, geographische Verteilung und Abdeckung des Fächerspektrums ausgewählt und in Hinblick auf die darin verwendeten logischen Operatoren untersucht.

Die analysierten PStO gliedern sich in die Hochschulformen Universität(8), Fachhochschule(6), Technische Universität(4) und Kunst & Musikhochschule(2). Die geographische Verteilung berücksichtigt Hochschulen aller 16 Bundesländer, wobei Bayern(2), Hessen(2) und Nordrhein-Westfalen(3) mit mehr als einer PStO in die Untersuchung aufgenommen wurden. Bei Zuordnung zum Fächerspektrum lassen sich Geisteswissenschaften(4), Naturwissenschaften(4), technische Fächer(3), Sozialwissenschaften(4), Kunst oder Musik(2), Theologie(1) und Pädagogik(2) unterscheiden.

3.2 Auswertung der Analyse

Im Rahmen der Analyse wurden insgesamt 1746 unterschiedliche Konstrukte untersucht und in Bezug zu struktur-definierenden, logischen Operatoren gesetzt. Von den untersuchten Konstrukten lassen sich rund 62% (1082) den logischen Konstrukten, die aus der Literatur abgeleitet wurden, zuordnen. Die übrigen 38% (664) Konstrukte formulieren Nebenbedingungen und sprachliche Vereinfachungen von logischen Aussagen (s.g. dynamische Gruppen), die für PStO typisch erscheinen.

Die Mehrheit der Konstrukte - annähernd 58% (1018) - basieren auf wenigen Basis-Operatoren aus der Literatur: 26% (451) auf *UND*-Verknüpfungen, 17% (289) auf *ODER*-Operatoren und 16% (278) zeigen eine zeitliche Bedingung oder Zuordnung auf. Diese Standardverknüpfungen bilden damit die zentralen logischen Konstrukte bei der Formulierung der PStO und müssen bei jeder semantischen Repräsentation beachtet werden. Die Nutzung von *NICHT*-Konstrukten 4% (64) stellt dagegen eine Seltenheit in den PStO dar; davon wurden 60 Fälle von einer Universität als explizite Formulierung verwendet. In allen anderen PStO wird der Ausschluss von Möglichkeiten über Positivformulierungen erreicht, die ein *ENTWEDER/ODER* darstellen und damit in der Analyse dem logischen Konstrukt des *ODER* hinzugerechnet wurden.

Der logische Operator *FOLGT AUS* definiert kein strukturgebendes Konstrukt innerhalb der PStO, sondern stellt wie beschrieben den nachvollziehbaren Zusammenhang zwischen einer definierten Konsequenz und der zugrundeliegenden Voraussetzung her. Neben den im Detail analysierten Konstrukten ist der logische Operator *BASIERT AUF* bei allen untersuchten Institutionen aufgrund unterschiedlicher Versionsstände der jeweiligen PStO grundsätzlich gegeben und muss daher zwingend in einer vollständigen semantischen Repräsentation berücksichtigt werden. Darüber hinaus werden Bezüge auf fächerübergreifende oder institutsübergreifende Ordnungen genutzt, die Vorgaben und übergreifende Regeln beinhalten. Dies motiviert die Berücksichtigung eines *INCLUDE* als Funktion der Sprache.

Neben den bereits aus der Literatur bekannten Konstrukten und Operatoren wurden bei der Analyse 38% (664) Formulierungen identifiziert, die für den Studienverlauf strukturgebend wirken: Nebenbedingungen 9% (160), Zuweisung von Werten 2% (41) durch Formeln 15% (264) und Gruppierung ohne logische Verknüpfung 10% (171) sowie Empfehlungen 2% (28). Die Prüfung von Nebenbedingungen erfolgte z.B. durch *„Zur Abschlussarbeit kann nur zugelassen werden, wer mindestens xxx Credit-Points erbracht hat“*. Einfache mathematische Formeln und die Zuweisungen vom Ergebnis wurde z.B. formuliert mit: *„Das Modul gilt als bestanden, wenn mindestens 50% der Punkte erreicht wurden“*. Zum Teil nahmen die Formulierungen gegenseitig Bezug aufeinander: *„Die Gesamtnote der Bachelorprüfung wird als gewichtete Durchschnittsnote berechnet. Die Gewichtung erfolgt nach den Credit-Points der einzelnen benoteten Module“*. Schließlich wurden mehrere Module/Veranstaltungen, um darauf weiter Bezug nehmen zu können, zu Gruppen mit statischen bzw. dynamischen Benennungen zusammengefasst (bspw. durch *„Die Module WP1 bis WP8. . . “*). Eine Sprache, die den semantischen, strukturgebenden Gehalt der PStO abbildet, muss demnach auch diese vier Kategorien enthalten.

Auffällig in der Analyse war, dass im Umfeld der PStO sprachliche Wenn-Dann-Konstrukte verwendet werden, um instanziierte Regelfolgen oder Wertableitungen abzubilden. Eine Formulierung wie *„Wenn der Studierende den Schwerpunkt [. . .] wählt, so müssen aus dem Wahlpflichtbereich [. . .] zwei Module ausgewählt werden.“*, entspricht dabei aber eigentlich einer Folgerung aus den bestehenden *UND* und *ODER*-Regeln, die für diesen Studiengang definiert wurden. Andere Formulierungen wie *„Ein Modul ist bestanden, wenn die Modulprüfung mindestens mit ‚ausreichend‘(4,0) bewertet wurde“* stellen lediglich eine mathematische Ableitung eines Wertes und ebenfalls kein bedingendes logisches Konstrukt dar.

3.3 Schlussfolgerungen

Die Analyse der PStO bestätigt die Notwendigkeit der grundlegenden, in der Literatur genannten logischen Operatoren. Auch wenn die Varianten der bedingten Formulierung von Implikationen und die Negation von Aussagen nur selten bzw. nicht in PStO vorkommen, so sollten sie dennoch in einer semantischen Repräsentation in ihrem Ergebnis zumindest

über entsprechende Formulierungsmöglichkeiten in Form von *UND* oder speziellen *ODER*-Operatoren verfügbar sein.

Ein weiteres wesentliches Ergebnis der Analyse ist, dass sich Lösungen zur Abbildung von PStO nicht nur auf die Formulierung von logischen Operatoren beschränken dürfen. Es ist essentiell, dass die semantischen Regeln gleichzeitig dafür genutzt werden können, um Nebenbedingungen und mathematische Funktionen auf dieser Basis beschreiben zu können. Um einfache Bezüge ohne eine logische Einordnung von Modulen vornehmen zu können, müssen überdies Gruppen gebildet werden können.

4 Anforderungen aus bisherigen Implementationen

Die überwiegende Menge der Abbildungen von PStO in Deutschland erfolgte in den Systemen der Hersteller HIS (HISinOne, HIS GX), Datenlotsen (CampusNet), SAP (SLCM) und durch das System der TU Graz (CAMPUSonline). Um für den deutschen „Markt“ eine relevante Lösung zu präsentieren, sollten die Fähigkeiten dieser Systeme inkludiert sein. Die öffentlich zugänglichen Dokumente sowie Dokumente aus der Zusammenarbeit im ZKI (AK-Campus-Management) wurden dementsprechend analysiert.

Die Beschreibungen der Fähigkeiten der Systeme auf den Webseiten sind sehr allgemein gehalten. Zum Teil wird die Anwendbarkeit für ein typisches Fächerspektrum genannt. Allgemeine Aussagen zur Versionierung und Wiederverwendbarkeit von Modul-Definitionen lassen auf die strukturierte Speicherung der Information in Datenbanken schließen. Eine Verbindung zwischen dem PStO Texten und der im CM-System hinterlegten Abbildung wurde nicht gefunden.

5 Entwicklung der Sprache

Unter Beachtung der Anforderungen aus der Literatur und den Ergebnissen der Analyse von bestehenden PStO wurde eine neue semantische Repräsentation als Kodierung von Regeln entwickelt. Dazu passend wurde eine neue technologische Lösung erstellt, welche die Wandlung von natürlichem Text in diese eindeutige strukturelle Repräsentation stark vereinfacht. Texte können damit natürlichsprachlich aussehen und sowohl vom Menschen als auch vom Computer gleichermaßen gedeutet werden. Somit könnten PStO von den Gremien formuliert und direkt aus der vorhandenen Formulierung vom Computer verlässlich interpretiert werden.

Durch die Vereinbarung einer klaren Satzstruktur entfallen Mehrdeutigkeiten, die in natürlicher Sprache vorkommen. Auf eine Klassifikation der Satzbausteine (im Sinne [Pe97]) in der natürlicher Sprache wird verzichtet und stattdessen eine der menschlichen Sprache nahe und vereinfachte Grammatik vereinbart, die ein größere Flexibilität als klassische Programmiersprachen ermöglicht.

5.1 Beschreibung der Sprache

Die semantische Repräsentation enthält u.a. die folgenden Konstrukte (und passende Beispiele bzw. Anwendungsszenarien jeweils in Klammern):

- Unterscheidung von Definitionen, Nebenbedingungen und Empfehlungen
- Intervalle von Werten aus ganzen oder reellen Zahlen bzw. Symbolen
- Exklusiver und inklusiver Oder-Operator mit Anzahl der gültigen Optionen (mindestens 3, höchstens 5 von insgesamt 8 Alternativen)
- Verbindung von notwendigen Aussagen durch *UND*-Operator
- bedingte mathematische und boolesche Funktionen zur Ableitung und Festlegung von Werten (Summe „Workload-Punkte“ für bestandene Module) auf Basis der über Operatoren verbundenen Symbole
- Statische und dynamische Gruppen (alle Kurse Biologie Grundlagen 1 bis Biologie Grundlagen 9)
- Abbildung der zeitlichen Abfolge (Modul A liegt vor Modul B)
- Versionierung (für alle Definitionen, Nebenbedingungen und Empfehlungen)
- angekündigte Werte (für erwartete Benotungen)
- Inkludierung von Regelwerken (Nutzung und Definition von Rahmenregelwerken)
- Kaskadierung von begrifflichen Ebenen (Biologie.Grundstudium)
- Festlegung von Wertebereichen (z.B. für Notenwerte oder Semesterlaufzeiten)
- Instanziierung (Belegung von Werten mit individuellen Studienergebnissen)

Folgende Konstrukte werden nicht als Sprachelemente explizit angeboten, da diese durch andere Konstrukte substituierbar sind: Die Negation als logische Basis wird über den Ausschluss innerhalb der *ODER*-Mengen realisiert, da dort auch null von null Optionen einer Menge gewählt werden können. Die Implikation (*Wenn-Dann*) wird unterschieden nach logischen und zeitlichen Bedingungen, entweder durch eine Gleichheit (*Äquivalenz*) oder durch eine Zusammenfassung von Aussagen.

5.2 Konsistenzprüfung

Im Rahmen der Analyse wurden semantische Problemstellungen, die in aktuell beschlossenen PStO vorliegen, erfasst und beispielhaft fragwürdige Stellen gesammelt. Eine bereits in [Al57] ausgiebig analysierte sprachliche Unschärfe ergibt sich, da im Fließtext Aussagen

nicht klar gruppiert werden können. So wird z.B. das Wort „und“ sowohl als Verbindung einer Aufzählung (auch mit *ODER*-Charakter) als auch als Verbindung von gleichzeitig notwendigen Voraussetzungen (boolesches *UND*) verwendet. Aussagen, die mit *UND* bzw. *ODER* verbunden werden, stellen somit keinen eindeutigen sprachlich logischen Bezug her.

Zur Illustration einige Beispiele:

- *„Wer das Wahlpflichtmodul WP 76 wählt, darf nicht die Wahlpflichtmodule WP 30 und WP 40 wählen.“* Ist hier gemeint, dass man, wenn WP 76 gewählt wurde, a) weder WP 30 noch WP 40 wählen darf, oder dass man b) WP 30 und WP 40 nicht gemeinsam wählen darf? Vermutlich ist Interpretation a) gemeint, aber sprachlich sauberer wäre dann: Wer X wählt, darf weder Y noch Z wählen.
- *„[Dazu sind] für den Wahlpflichtbereich xyz die Wahlpflichtmodule (WP 14 oder WP 15), WP 40 und WP 55 zu wählen.“* Hier liegt die Interpretation, dass WP 14 und WP 15 sich gegenseitig exklusiv ausschließen, nahe. Sprachlich eindeutig sind so formulierte *ODER*-Bedingungen nicht, so dass eine Wahl von WP 14 und WP 15 nicht eindeutig vermieden wird.

Derartige sprachliche Probleme betreffen zunächst nur den Übersetzungsschritt aus einem natürlichsprachlichen Text in eine eindeutige Semantik. Die neu entwickelte Sprache bietet dazu eine flexible, jedoch die Eindeutigkeit erhaltende Basis.

Wenn die aus dem Text entstehende mathematische Bedeutung der Repräsentation eindeutig ist, kann die formulierte Ordnung durch die neue Technologie zusätzlich auf Konsistenz geprüft werden: Logische Bezüge zwischen Begriffen und zeitliche Abfolgen werden auf Zirkelbezüge geprüft; Wahloptionen, die nur scheinbar existieren, aber durch Nebenbedingungen letztlich ausgeschlossen werden, werden aufgedeckt; Lücken im Übergang zwischen Versionen werden erkannt. Eine erste Liste von mehr als 30 Konstellationen wurde im Rahmen der Implementierung des Prototyps erstellt. Diese können zwar durch die Flexibilität der Sprache in einer semantischen Repräsentation abgebildet werden, sind in sich aber nicht notwendigerweise konsistent. Ein umfangreicher Nachweis, dass derartige Fehler tatsächlich in PStO vorliegen, kann allerdings erst erbracht werden, wenn die Technologie zur Abbildung von vielen Ordnungen verwendet wird und dabei diese Fehler identifiziert werden. Eine Anwendung der Technologie in der Breite steht aber noch aus, so dass zu diesem Zeitpunkt keine Übersicht erstellt werden kann.

5.3 Allgemeingültigkeit

Die durch die Anwendung der strukturgebenden semantischen Sprache erzeugte Repräsentation ist allgemeingültig und unabhängig von spezifischen Eigenschaften der Umsetzung eines konkreten CMS.

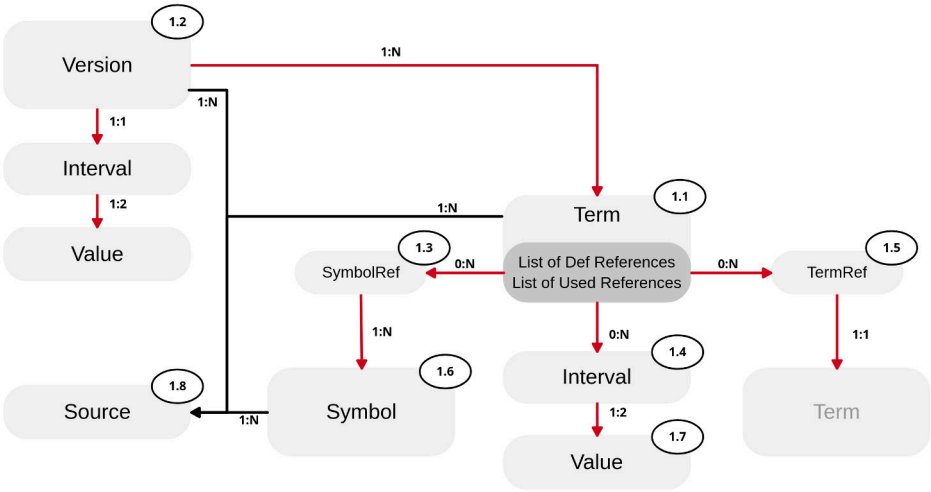


Abb. 1: Allgemeine Objektstruktur der semantischen Repräsentation

Abbildung 1 zeigt die gewählte konkrete Objektstruktur der semantischen Repräsentation. Grundsätzlich werden alle Sprachkonstrukte (Definitionen, Nebenbedingungen und Empfehlungen) als Term **1.1** aufgefasst, der einer Versionierung **1.2** unterliegt. Terme referenzieren potentiell Symbole **1.3**, Intervalle **1.4** und weitere Terme **1.5**. Sie dienen damit sowohl der Abbildung logischer Operatoren als auch der Abbildung von Funktionen oder mathematischen Rechenoperationen. Symbole **1.6** dienen der Kaskadierung und Abstraktion von Namensräumen. Die Intervalle werden verwendet, um Versionen oder die Menge der *ODER*-Auswahl zu repräsentieren, deren Unter- und Obergrenze durch Werte **1.7** dargestellt werden. Ein Intervall mit einem Wert entspricht der normalen Speicherung von Ganzzahlen, Fließkommazahlen, Strings oder symbolischen Werten wie ∞. Schließlich werden alle Sprachkonstrukte mit Bezug zum Quelltext (Source) **1.8** verwaltet. Die rekursive Schachtelung aller Terme erlaubt beliebige Kaskadierung der abzubildenden semantischen Struktur. Die Unterscheidung verschiedener Statements ist durch die Typisierung der Terme jederzeit erweiterbar.

Diese Allgemeingültigkeit der logisch semantischen Regeln ist entscheidend, damit grundsätzlich die Möglichkeit gegeben ist, diese Repräsentation sowohl in einen Text zurückzuführen als auch für technische Systeme wie die CMS bereitzustellen.

Somit wäre eine technische Lösung potentiell sogar in der Lage, automatisch auf Eigenschaften oder Regelformulierungen der PStO aufmerksam zu machen, deren Abbildung in einem konkreten System problematisch sein könnten. Gleiches gilt für PStO, die von einem CMS in ein anderes übertragen werden sollen. Hier kann die Zwischenebene der neutralen Repräsentation genutzt werden, um Vereinfachungen der Regelformulierungen so vorzunehmen, dass ein konkretes Zielsystem optimal unterstützt wird. Voraussetzung dafür

ist die generelle Unterstützung einer einheitlichen semantischen Repräsentation innerhalb der Systeme nach einem gemeinsamen Standard oder zumindest die Erstellung geeigneter Import- & Export-Konfigurationen.

6 Aktueller Stand der Umsetzung

Das in diesem Beitrag vorgestellte Projekt befindet sich aktuell in der Erstellung eines Prototypen zur Verifikation der getätigten Aussagen.

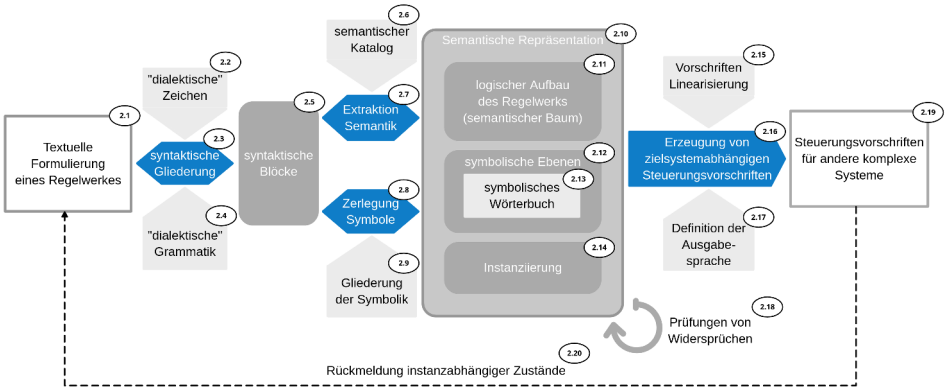


Abb. 2: Funktionaler Aufbau des Prototyps

Abbildung 2 stellt den funktionalen Aufbau des Prototyps dar, der auf der oben dargestellten Form der semantischen Repräsentation basiert. Die von uns implementierte Standard-Sprache erlaubt die Erweiterung um beliebige Symbole: Alle Standard-Symbole sind flexibel für s.g. Dialekte neu oder zusätzlich zu definieren **2.2**; dies gilt selbst für Kommentar-Start- und -Ende-Zeichen, Leerzeichen und Zeilenenden. Die zum Einlesen von Texten **2.1** verwendete Syntax ist auf Basis der Übergänge einer State Machine konfigurierbar **2.4** und ohne Programmänderung anpassbar. Die Möglichkeit, die Ausgabe auf Basis einer Syntax-Konfiguration gestalten zu können, bereitet die schnelle Entwicklung von Input/Export-Schnittstellen für CMS-Systeme vor (**2.15**, **2.17**).

Aktuell ist der Prototyp bereits in der Lage, sämtliche Sprachkonstrukte korrekt zu identifizieren **2.3**, einzulesen und die semantische Repräsentation **2.10** zu erzeugen (**2.7**, **2.8**). Diese kann vollständig in einem standardisierten Format ausgegeben werden und dem Menschen damit den Einblick in das repräsentierte Regelwerk bzw. eine repräsentierte Instanz (z.B. einen Studienverlauf) geben. Für die semantische Prüfung ist die Erkennung von logischen Fehlern **2.18** weitgehend umgesetzt und beinhaltet: Schleifen in jeglichen semantischen Konstrukten (mathematische Formeln, Zeit, logische Ausdrücke); Widersprüche in komplexen *ODER- / UND*-Kombinationen; Widersprüche zwischen Versionen des Regelwerks.

6.1 Beispiel

Der Abdruck einer kompletten PStO wäre im Rahmen des Artikels nicht sinnvoll, da die Sprache primär die strukturgebenden Aspekte abbildet. Es wurde daher eine Passage aus einer analysierten Studienordnung ausgewählt, die auf kleinem Raum viele verschiedene Sprachkonstrukte benutzt. Zur Übersichtlichkeit werden alle enthaltenen Aussagen nummeriert. Zunächst wird der Abschnitt im Original zitiert. Anschließend wird der semantische Gehalt in SemaLogic dargestellt. Abschließend wurde eine erweiterte Sprach-Symbolik gewählt, die den Text möglichst nahe am originalen Text bzw. der deutschen Sprache belässt, aber vollständig und eindeutig in die semantische Repräsentation überführbar ist.

```
(1) Aus den Wahlpflichtmodulen WP 18 bis WP 26 und WP 68 bis WP 77 sind
    insgesamt zwei Wahlpflichtmodule zu wählen.
(2) Dabei soll im 3. und 6. Fachsemester jeweils ein Wahlpflichtmodul
    gewählt werden.
(3) Wer das Wahlpflichtmodul WP 19 wählt, darf nicht die
    Wahlpflichtmodule WP 2 und WP 3 wählen.
(4) Wer das Wahlpflichtmodul WP 76 wählt, darf nicht die
    Wahlpflichtmodule WP 30 und WP 40 wählen.
```

List. 1: Text der PStO

Als Standard-Sprache formuliertes Beispiel:

```
(1) Wahlpflichtmodule 2|2 {Wahlpflicht ~ WP 18 | WP 26, WP 68 | WP 77 ~}
(2) ! 3te Fachsemester 1|1 {Wahlpflicht}
    ! 6te Fachsemester 1|1 {Wahlpflicht}
(3) [WP 19, 0|0 {WP 2, WP 3}]
(4) [WP 76, 0|0 {WP 30, WP 40}]
```

List. 2: SemaLogic als Standard-Sprache

Als Sprachanpassung formuliertes und semantisch identisch überführbares Beispiel:

```
(1) Wahlpflicht umfasst WP 18 bis WP 26, WP 68 bis WP 77 als Gruppe.
    Wahlpflichtmodule enthält 2 bis 2 Alternativen aus Wahlpflicht.
(2) Empfohlen wird: 3te Fachsemester enthält 1 bis 1 Alternativen aus
    Wahlpflicht.
    Empfohlen wird: 6te Fachsemester enthält 1 bis 1 Alternativen aus
    Wahlpflicht.
(3) Nebenbedingung 1 besteht aus WP 19 sowie 0 bis 0 Alternativen
    aus WP 2, WP 3..
(4) Nebenbedingung 2 besteht aus WP 76 sowie 0 bis 0 Alternativen
    aus WP 30, WP 40..
```

List. 3: GI-Paper Dialekt

Für den letzten Abschnitt wurden folgende Definitionen vorgenommen, so dass der Prototyp die Sprachelemente zusätzlich auf die Standard-Sprache mappen kann:

- Start einer Gruppe: „ umfasst “ ⇒ „~“
- Ende der Gruppe: „ als Gruppe“ ⇒ „~“
- Trennzeichen für Intervalle: „ bis “ ⇒ „|“
- Beginn einer Empfehlung: „Empfohlen wird:“ ⇒ „!“
- Überlesen als Leerzeichen: „ enthält “ ⇒ „ “
- Start einer *ODER*-Regel: „ Alternativen aus “ ⇒ „{“
- Start einer *UND*-Regel: „ besteht aus “ ⇒ „[“
- Trenner zwischen Symbolen: „ sowie “ ⇒ „ ,“
- Ende von *UND*- und *ODER*-Mengen: „.“ ⇒ „}“ und „]“

Zugegebenermaßen leidet der flüssige deutsche Satzbau etwas unter den Restriktionen der statischen Grammatik der bisher definierten Sprache. Dennoch sind die im letzten Abschnitt formulierten Sätze bereits näher an der natürlichen Sprache als an der hier definierten Standard-Sprache oder einer anderen formalen Programmiersprache wie Prolog. Dem Nutzer steht es frei, durch weitere Definitionen die Formulierung noch weiter an den üblichen Sprachgebrauch anzupassen.

7 Ausblick

Die aktuelle standardisierte Sprache und die damit verbundene semantische Repräsentation ist vollständig auf wenigen Seiten beschreibbar und wird in Kürze veröffentlicht. Die technische Umsetzung zur Steuerung oder Konfiguration komplexer Plattformen ist zusammen mit der Sprachdefinition zum Patent angemeldet. Es wird ferner angestrebt, die aktuellen Arbeiten in die Bemühungen um den XHochschul-Standard einfließen zu lassen.

7.1 Semantik als Basis für weitere Anwendungsfälle

Die von uns definierte semantische Repräsentation eignet sich sowohl zur Darstellung eines Regelwerks einer PStO als auch zur Speicherung von Studienverläufen. Weiterhin kann auf Basis der semantischen Repräsentation die Zahl der tatsächlich unterschiedlichen Varianten eines Studiums berechnet werden und vieles mehr. Da die Repräsentation auch mit Werten einer individuellen Instanz - also einem konkreten Studienverlauf - befüllt werden kann, ist eine Prüfung von hypothetischen Studienverläufen möglich. Wenn

zukünftig mehrere dieser Hypothesen gleichzeitig getestet werden, entstehen dadurch für die modellierte PStO statistische Vorhersagen des Kapazitätsbedarfs bei gegebenen Annahmen von Studierendenzahlen.

Gemeinsam mit der zugrunde liegenden semantischen Beschreibung der PStO bietet die semantische Repräsentation eine ideale Plattform für den Austausch bzw. die Anerkennung von Studienleistungen auf nationaler und europäischer Ebene. Die Besonderheit besteht dabei in der Neutralität des gespeicherten Wissens und der Möglichkeit, bei der Ausgabe eine andere syntaktische Grammatik und Sprache zu wählen als bei der Kodierung. Natürlich müssten verwendete Begriffe wie „Biologie“ extern als Matching bereitgestellt werden, da unsere Technologie die Übersetzung einzelner, einfacher Begriffe voraussetzt und sich auf die semantische Kodierung der Zusammenhänge fokussiert.

7.2 Einschränkungen

Die vorgeschlagene semantische Repräsentation, welche strukturierte, logische Zusammenhänge, die durch Regelwerke formuliert werden, abbilden kann, ist nicht für alle denkbaren Szenarien von juristischen Texten geeignet. So werden z.B. keine dehnbaren Begriffe („in geeigneter Weise“) oder solche mit klaren juristischen Bedeutungen („unverzüglich“ entspricht „ohne schuldhaftes Zögern“) abgebildet. Die vorgeschlagene Sprache fokussiert auf strukturgebenden, logischen Regeln, welche die Voraussetzungen zum Erreichen eines angestrebten Zustands (z.B. dem Abschluss eines Studiums) definieren.

Die o.g. Liste der Konstellationen, bei denen potentielle Widersprüche und andere Mängel der Konsistenz auftreten, ist noch nicht abgeschlossen. Ferner gehen wir aktuell davon aus, dass eine solche Liste endlich sein kann, damit letztlich alle Mängel durch ein automatisches Verfahren gefunden werden können. Ein Beweis dieser Annahme steht aus.

Schließlich wird die Anerkennung von Leistungen potentiell von der semantischen Struktur profitieren können. Eine Zuordnung der Inhalte ist Bestandteil der Modulbeschreibungen und nicht Teil der Aufbaustruktur der in den PStO definierten Studiengänge. In Konsequenz wird diese Ebene nicht in der hier vorgestellten Sprache abgebildet.

7.3 Schlussfolgerungen

Um die Wirksamkeit der semantischen Prüfung zu erreichen, wurde die in diesem Beitrag vorgeschlagene Sprache in Relation zur natürlichen Sprache stark vereinfacht. Die große Flexibilität in der Nutzung von vereinbarten Sprachsymbolen und einer ggf. durch Konfiguration adaptionsfähigen Grammatik erlaubt Satzstrukturen, die dem korrekten (z.B. deutschen) Satz sehr ähnlich sind und intuitiv verstanden werden. Eine Neuformulierung von Ordnungen in den Passagen, die den studienorganisatorischen Aufbau betreffen, muss dennoch in angemessener Weise erfolgen, um das Konzept anwenden zu können. Für alle

Ordnungen, die in solcher Weise umformuliert werden, kann ein Kompromiss gefunden werden, bei dem der Text sowohl vom menschlichen Leser als auch vom Computer identisch verstanden wird. Die zusätzliche Umsetzung in Campus-Management-Systemen zur Strukturierung des Studiums in Modulen und deren gegenseitige Abhängigkeiten kann dann entfallen.

8 Danksagung

Die Autoren möchten sich beim Team der LMU für den offenen Austausch zu den Schwierigkeiten bei der Abbildung von PStO bedanken. Weiterhin bedanken sich die Autoren bei den Gutachter:innen für die nützlichen Hinweise zum Manuskript.

Literatur

[Al57] Allen, L. E.: Symbolic Logic: A Razor-Edged Tool for Drafting and Interpreting Legal Documents, Diss., Yale Law School, 1957.

[KM93] Kim, J.-T.; Moldovan, D.: Acquisition of semantic patterns for information extraction from corpora. In: Proceedings of 9th IEEE Conference on Artificial Intelligence for Applications. Proceedings of 9th IEEE Conference on Artificial Intelligence for Applications. S. 171–176, März 1993.

[SM95] Spitta, T.; Mordau, J.: Entwicklung und Ergebnisse eines allgemeingültigen Fachkonzeptes für die Prüfungsverwaltung an Hochschulen, Universität Bielefeld, Okt. 1995, S. 127–147, URL: https://pub.uni-bielefeld.de/download/2669014/2679395/PrufungsVerwaltung-Siegen95.pdf.

[Pe97] Peek, N.: Representing Law in Partial Information Structures. Artificial Intelligence and Law 5/4, S. 263–290, 1. Dez. 1997, ISSN: 1572-8382, URL: https://doi.org/10.1023/A:1008238332032, Stand: 18. 07. 2020.

[TR05] Tetzner, E.; Riedewald, G.: Spezifikation und Verifikation in regelbasierten Beratungssystemen auf der Grundlage hybrider Automaten. In: Programmiersprachen und Rechenkonzepte. 22. Workshops der GI-Fachgruppe 2.1.4. Bd. Bericht Nr. 0513, Bad Honnef, S. 77, 2005, URL: http://www-ps.informatik.uni-kiel.de/fg214/Honnef2005/TechnischerBericht0513.pdf#page=85.

[Ha06] Habtemariam, D. T.: Simulation von Prüfungsordnungen und Studiengängen mit Hilfe von Constraint-logischer Programmierung, Diss., Frankfurt am Main, 2006, URL: https://d-nb.info/1046848909/34.

[OA07] Otto, P. N.; Anton, A. I.: Addressing Legal Requirements in Requirements Engineering. In: 15th IEEE International Requirements Engineering Conference (RE 2007). 15th IEEE International Requirements Engineering Conference (RE 2007). ISSN: 2332-6441, S. 5–14, Okt. 2007.

[Te08] Tetzner, E.: Nutzerfreundliche Modellierung mit hybriden Systemen zur symbolischen Simulation in CLP, Diss., Rostock: Universität Rostock, 5. Sep. 2008, 282 S., URL: https://d-nb.info/993978975/34.

[Ba11] Bauer, N.-J.: Sieben Jahre integriertes Campus Management an deutschen Hochschulen. In: Allgemeine Aspekte zur prozessorientierten Hochschule. Bock und Herchen Verlag, S. 39–50, Aug. 2011, URL: http://www.dini.de/fileadmin/docs/Prozessorientierte_Hochschule_2011.pdf#page=39, Stand: 13. 07. 2020.

[SAA12] Schreiter, J.; Alt, R.; Auth, G.: Business Engineering bei der Einführung von Campus-Management-Systemen–Herausforderungen und Potenziale. INFORMATIK 2012/, Publisher: Gesellschaft für Informatik eV, S. 642–656, 2012, URL: https://dl.gi.de/handle/20.500.12116/17866.

[BCS13] Brune, H.; Carolla, M.; Spitta, T.: Studiengangsmodellierung - Ein implementierter Diskussionsansatz -./, Accepted: 2017-12-06T09:52:38Z Publisher: Gesellschaft für Informatik e.V., 2013, ISSN: 1610-5753, URL: http://dl.gi.de/handle/20.500.12116/8849, Stand: 13. 07. 2020.

[Ar14] Araujo, D. A. d.; Müller, C.; Chishman, R.; Rigo, S. J.: Information extraction for legal knowledge representation – a review of approaches and trends. Revista Brasileira de Computação Aplicada 6/2, S. 2–19, 29. Mai 2014, ISSN: 2176-6649, URL: http://seer.upf.br/index.php/rbca/article/view/3542, Stand: 18. 07. 2020.

Community-basierte Methode zur transdisziplinären Gestaltung von Lernräumen an Hochschulen

Lars Schlenker[1], Carmen Neuburg[2]

Abstract: Die Hochschule als digitalisierten Lernort gestalten zu wollen, erzeugt neue Herausforderungen für die Prozesse der Planung und die daran beteiligten Akteure. Um den komplexen Anforderungen an die Zusammenarbeit unterschiedlicher Fachdisziplinen (Architektur, Medientechnik/IT und Pädagogik) gerecht zu werden, ist es notwendig, ihre Vertreterinnen und Vertreter über einen transdisziplinären und partizipativen Diskurs an der Lehr- und Lernraumentwicklung zu beteiligen. Dazu wurden an der TU Dresden im Rahmen eines BMBF-Projekts Design Patterns entwickelt und eine Methode ihrer community-basierten Weiterentwicklung erprobt, um das unterschiedliche Wissen über die Gestaltung von Lernumgebungen transparent zu machen und die Kommunikation und die Zusammenarbeit in transdisziplinären Planungsprozessen zu unterstützen.

Keywords: Design Pattern; Lernräume; Planungsinstrumente

1 Herausforderung Hochschulplanung

Der Veränderungsdruck an Hochschulen nimmt weiter zu. Allein in den letzten 10 Jahren stiegen die Studierendenzahlen deutschlandweit um fast 700.000 [St19]. Hochschulen müssen sich auf weiter steigende Studierendenzahlen, die Digitalisierung der Bildung, aber auch auf neue Lehr- Lernkonzepte nicht nur vor dem Hintergrund der Covid19-Pandemie einstellen. Räume für neue Formen des Lehrens und Lernens, wie das selbstgesteuerte Lernen oder die örtlich flexible Arbeit mit digitalen Formaten und Inhalten, werden dringend benötigt. Über diese Veränderung von Lernräumen an Hochschulen hin zu sogenannten „CrossActionSpaces“ [Ja15] herrscht im theoretischen Diskurs weitestgehend Einigkeit. Räume aber sind Teil von Umgebungen. Sie stehen im Zusammenhang mit anderen nicht nur räumlichen Infrastrukturen. Neue Räume erfordern daher auch neue Umgebungskonzepte, die die Integration zeitgemäßer Lehr- und Lernszenarien zulassen. Entsprechend innovative Konzepte verbinden die physischen Räume der Hochschule mit dem digitalen Raum und adressieren soziale Kommunikationsaspekte ebenso wie vielfältige didaktische Szenarien. Dafür müssen in den Planungsprozess auch von akademischen Neu- oder Umbauten neben Architekten und Medienplanern zunehmend weitere Stakeholder, allen voran Pädagogen einbezogen werden [NJ18].

Transdisziplinäre Planungsprozesse zu etablieren, trägt dem wachsenden Bedürfnis der Planungsforschung nach flexibleren Vorgehensweisen ebenso Rechnung wie der Integration

[1] TU Dresden, Bildungstechnologie, Weberplatz 5, 01217 Dresden, Germany, lars.schlenker@tu-dresden.de
[2] TU Dresden, Bildungstechnologie, Weberplatz 5, 01217 Dresden, Germany, carmen.neuburg@tu-dresden.de

doi:10.18420/inf2020_49

unterschiedlicher Perspektiven in der Planungspraxis. Nach Nissler und Prey [NP18] gehören dazu im Hochschulkontext auch die Perspektiven von Studiendekanen, Gebäudemanagern sowie Akteuren von Bibliotheken und Medienzentren. Die praktische Umsetzung einer solchen Zusammenarbeit unter Berücksichtigung verschiedener disziplinärer Hintergründe stellt allerdings hohe Anforderungen an den Planungsprozess und findet trotz wiederholter Forderungen danach selten statt. Oft scheitert eine Zusammenarbeit, da der zeitliche und personelle Aufwand einer Bedarfsplanung, bei der die Anforderungen der verschiedenen Akteure berücksichtigt werden, als zu hoch und aufwendig erscheint. Die auch für die Abrechnung von Leistungen bei der Planung und dem Bau von Bildungsbauten maßgebliche Honorarordnung für Architekten und Ingenieure (HOAI) weist eine solche projektvorbereitende Phase als Phase 0 bereits aus. Sie wird bis dato nur als Vorleistung geführt. Planungsexperten und Projektmanager sehen im Gegensatz dazu eine intensive Bedarfsplanung angesichts einer wachsenden Zahl von Planungsakteuren und damit verbundenen zunehmenden Komplexität der Planungsaufgaben als unerlässlich an. [HS19]. Instrumente, die entsprechende Prozesse unterstützen und darüber hinaus das vielfältige Wissen und die Erfahrungen der Beteiligten darstellen und abgleichen sowie Verständnis für die unterschiedlichen Sichtweisen und das daraus resultierende Rollenhandeln sowie Anknüpfungspunkte für die transdisziplinäre Kommunikation erzeugen, fehlen zudem aktuell weitestgehend.

2 Patternsammlung LR_D

Am Diskurs über Lernräume nehmen neben Fachplanern eine Vielzahl an weiteren Akteuren sowie Stakeholdern teil. Sie betrachten Lernräume aus verschiedenen Perspektiven und diskutieren auf der Basis von unbewussten Selbst- und Fremdbildern. Deshalb sind Methoden und Prozesse von zentraler Bedeutung, bei denen frühzeitig ein sogenannter Common Ground [CB91], eine gemeinsame Wissensbasis der am Planungsprozess Beteiligten, geschaffen wird. Vor dem Hintergrund dieser Herausforderungen setzte sich eine transdisziplinäre Arbeitsgruppe aus den Bereichen Bildungswissenschaft, Architektur und Medientechnologie der TU Dresden gemeinsam mit Nutzerinnen und Nutzern sowie Akteuren kommunaler, wie privater Bildungsträger und Unternehmen im Rahmen des vom BMBF geförderten Projekts LR_D[3] mit der Entwicklung und Erprobung geeigneter partizipativer und kollaborativer Methoden und Instrumente für die gemeinsame Planung und Gestaltung von Lernräumen auseinander.

Das Projekt LR_D erzeugte Ergebnisse auf unterschiedlichen Ebenen. Ein Ergebnis auf der konzeptionellen Ebene ist die Forderung, digitale und physische Bestandteile von Lehr- und Lernumgebungen nicht getrennt zu betrachten. Stattdessen sollten sie anhand gemeinsamer raumbezogener Anforderungen eines „Umgebungskonzepts Lehren und Lernen“ als eine Einheit adressiert und gestaltet [SNK18] werden. Die daraus abgeleiteten

[3] Das Projekt Lehrraum_digital (LR_D) wird vom deutschen Bundesministerium für Bildung und Forschung (BMBF) im Rahmen der Förderbekanntmachung Digitale Medien in der beruflichen Bildung im Zeitraum vom November 2016 bis April 2019 gefördert. Projektwebsite [https://blog.tu-dresden.de/lehrraum-digital/].

methodischen Entscheidungen überprüfte die Projektgruppe LR_D ab Frühjahr 2018 in konkreten Planungsstudien an zwei Schulen der beruflichen Bildung in Dresden und Bamberg. Die zentralen Ergebnisse des Projekts bewegen sich dementsprechend ebenfalls auf der methodischen Ebene. Dabei handelt es sich zum einen um den partizipativ gestalteten Planungsworkshop LR_D [SNB18] und zum anderen um eine Sammlung von transdisziplinären Gestaltungsmustern (Design Patterns), die der Projektgruppe als Wissensbasis zum Thema Lehr-Lernraumplanung diente. Beide Elemente wurden als methodische Formate im Rahmen der Planungsstudien und von projektbegleitenden Gesprächen und Workshops mit Experten und Stakeholdern entwickelt und erprobt.

2.1 Design Pattern als transdisziplinäres Instrument

Bei Entwurfsmustern (Design Pattern) handelt es sich um eine erstmals vom Architekt und Architekturtheoretiker Christopher Alexander für die Architektur [AIS78] beschriebene inzwischen aber auch in der Designtheorie, der Informatik und Pädagogik verankerte Vorgehensweise. Design Patterns stellen widerverwendbare Konzepte dar, die eine Problemstellung mit ihrem entsprechenden Kontext abbilden und verschiedene Lösungsvarianten aufzeigen. Sie dokumentieren typische Arbeitsaufgaben sowie die damit verbundenen Problemstellungen und Herausforderungen, erheben aber nicht den Anspruch, standardisierten Lösungen dafür anzubieten. Da Patterns flexibel und unabhängig voneinander nur in Abhängigkeit von der jeweiligen Problemstellung (Kontext) eingesetzt werden können, ermöglicht das Arbeiten mit ihnen einen individuellen Lösungsprozess [Le07]. Dabei beruht der Pattern-Ansatz auf der induktiven Ableitung von Erkenntnissen, meist durch Beobachtungen und Analyse von validen oder ungeeigneten Beispielen aus der Praxis und der Ableitung von Lösungsansätzen. Zur weiteren Analyse dieser Beispiele können Experteninterviews und Gruppendiskussionen eingesetzt werden. Die Grundmotivation der Pattern Language von Alexander sieht vor, Vorkommnisse zu systematisieren und den Nutzer inklusive seiner Bedürfnisse besser einzubeziehen [AIS78]. Das kann zu einem besseren gemeinsamen Verständnis (Common Ground) führen, sofern die Entwurfsmuster auch ohne spezifisches Expertenwissen verstanden werden können. Dabei ist jedes Muster eigenständig und auch individuell lesbar. Um es für die verschiedenen Akteure trotzdem verständlich und nachvollziehbar zu machen, folgt der inhaltliche Aufbau eines jeden Musters einer festen, wiederkehrenden Struktur [Le07].

2.2 Entstehung und Darstellung

Die Patterns des Projekt LR_D dokumentieren Lösungsansätze für die Gestaltung von Lehr- und Lernräumen. Sie berücksichtigen unterschiedliche fachliche Perspektiven, um den unterstützen transdisziplinären Austausch zu unterstützen. Ausgangspunkt der Patternsammlung waren die von der Forschungsgruppe LR_D, bestehend aus Akteuren der Architektur, der Medienplanung und der Pädagogik, identifizierten Themen. Regelmäßige Feedbackschleifen

mit Stakeholdern und Akteuren aus unterschiedlichen Bildungskontexten, nutzerzentrierte Workshops mit Lernenden und Lehrenden sowie die Auswertung vom im Projekt durchgeführten Planungsstudien schufen eine Grundlage, auf der eine Auswahl von Patterns mit thematisch und inhaltlich hoher Relevanz für den Planungs- und Gestaltungsprozess von Lernräumen entstehen konnte.

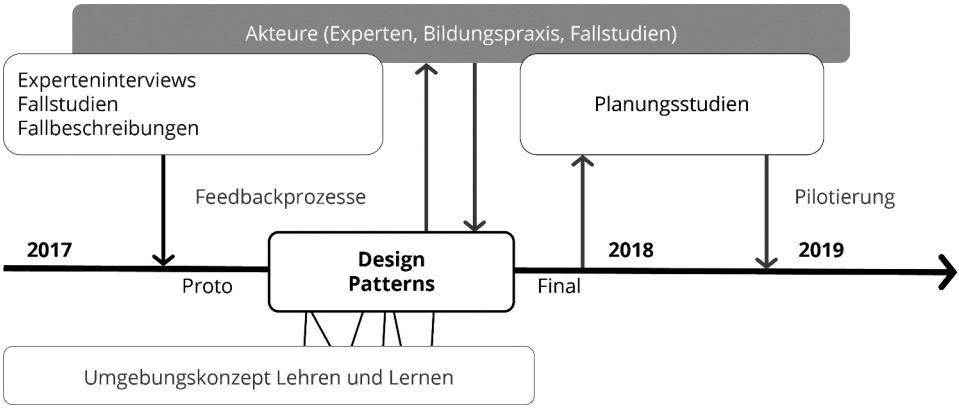


Abb. 1: Entstehungsprozess Pattern LR_D

Neben der inhaltlichen Ausgestaltung der Muster spielte die Darstellungsform der Patterns in Abhängigkeit vom Einsatzkontext eine wichtige Rolle. Patterns sind inzwischen in unterschiedlichen fachlichen Disziplinen verbreitet. Neben der Architektur und die Designtheorie gehören dazu auch die Informatik und die Pädagogik. In Abhängigkeit vom fachlichen Kontext weichen die Darstellungsform und –struktur der Design Patterns deutlich voneinander ab. Ein Charakteristikum der von Christopher Alexander entwickelten Design Patterns ist das Arbeiten mit visuellen Ankerbildern [AIS78]. Horst Rittel [Ri13] als Vertreter der Designtheorie überführte ihre Darstellung in Concept-Maps. Sie dienten ihn der Bearbeitung so genannten „bösartiger“ Probleme, bei deren Lösung unterschiedlichste Kontexte und konkurrierende Variablen berücksichtigt werden müssen und der Planer bzw. Entwerfer sich in einem ständigen Lernprozess befindet [Ri12]. In den im Projekt LR_D entwickelten Concept-Maps in Anlehnung an die Rittelsche Planungsmethodik wurden dem Muster (Objekt-Modell) verschiedene mögliche Rahmenbedingungen (Kontext-Modell) vorangestellt, um sie daran anschließend in unterschiedliche Handlungsvarianten (Performance-Modell) zu überführen (siehe Abb. 1). In der im Projekt LR_D Methodik wurden die Designvariablen vom im Projekt entwickelten Umgebungskonzept [SNK18] erzeugt. Das Rittelsche Performance-Modell wurde beibehalten, aberh durch eine abschließende Bewertung ergänzt, deren Ergebnis wiederum das Kontext-Modell beeinflusst. Auf der Basis dieses Strukturmodells wurden die Pattern im Projekt LR_D in Concept-Maps überführt und innerhalb von Fallstudien mit unterschiedlichen Nutzergruppen und Experten aus unterschiedlichen Bildungskontexten getestet und weiterentwickelt [BS20]. Die strukturierte und übersichtliche Darstellung in Concept-Maps unterstützte dabei vor allem die Zusammenarbeit mit Akteuren ohne spezifisches Planungs- und Entwurfswissen.

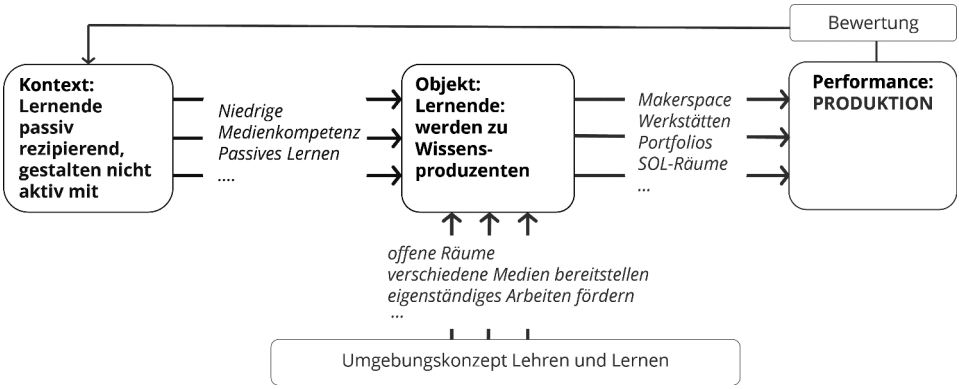


Abb. 2: Beispiel Pattern *Produktion* (Concept-Map in Anlehnung an Horst Rittel)

Um die Vorteile verschiedener Darstellungsformen zu erschießen, wurden die Concept-Maps abschließend in frei zugänglichen Wiki-Artikeln verschriftlicht und ausformuliert. Neben beispielgebenden Bildern wurde zusätzlich die den Pattern zugrundeliegenden Concept-Maps verlinkt.

Die finale Dokumentation der Pattern LR_D bietet ihren Nutzern verschiedene Zugänge und Auswahlmöglichkeiten. Der Einstieg in ein Pattern kann über ein konkretes Pattern-Thema, wie z.B. „Rückzugsraum“ oder „Lernortvernetzung“ erfolgen. Dazu werden die Textlinks, die zu den einzelnen Pattern-Seiten führen, in der Sammlungsübersicht durch themenbezogene und beispielgebende Bilder (visueller Anker) ergänzt (siehe Abb. 3). Der Zugang zu den einzelnen Patterns ist alternativ auch über eine bestimmte Akteursgruppe bzw. Rolle, wie beispielsweise die des Architekten möglich. Dabei wird für jeder Rolle eine individuelle Patternansicht generiert, auf der die einzelnen Patternbilder in Abhängigkeit von ihrer Bedeutung für die jeweilige Rolle in unterschiedlichen Größen dargestellt werden. Die Darstellungsstruktur jedes Pattern ist, um die Orientierung zu erhöhen, identisch aufgebaut. Neben dem themenbezogenen Bild beinhaltet der Eintrag zunächst eine kurze Beschreibung des Patterns. Es folgen typische damit verbundene Herausforderungen und Problemstellungen, einflussgebende Parameter und konkrete Umsetzungsbeispiele. Abschließend werden geeignete Bewertungskriterien, die zur Messung einer erfolgreichen Intervention geeignet sind, aufgelistet. Ergänzend wird auf verwandte Design Patterns verwiesen.

2.3 Veröffentlichung und Partizipation

Bei der Veröffentlichung der Patterns als zentrale Projektergebnisse waren sich die Verbundpartner einig, sie Nachnutzern nicht nur zugänglich zu machen, sondern ihnen auch die Möglichkeit einzuräumen, sie im Dialog mit anderen teilen und im Sinne einer nicht

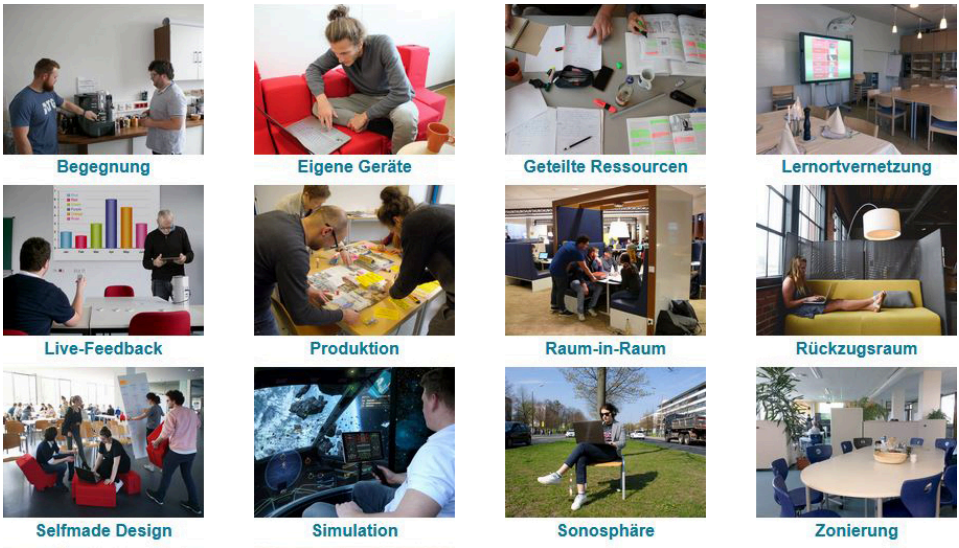


Abb. 3: Übersicht Patternsammlung LR_D

abgeschlossen Sammlung weiterentwickeln und ergänzen zu können. Verschiedene am Prozess der Planung von Hochschulen beteiligten Akteure, Fachexperten und Stakeholder sollten sich informieren, aber auch gleichberechtigt ihre Erfahrungen und ihr Wissen über eine frei zugängliche Plattform weitergeben, austauschen und dokumentieren können. Den inhaltlichen Nukleus der Sammlung bildeten die im Projekt LR_D entwickelten Patterns zur Gestaltung digitaler Lehr- und Lernräume. Bei der technischen Basis fiel die Wahl aufgrund seiner niedrigschwelligen Handhabung und hohen Verfügbarkeit auf das Wiki einer Lernplattform. Digitale Lernplattformen sind an Hochschulen inzwischen stark verbreitet und bieten vergleichbare Funktionen und Routinen an. Mit Dritten können über Lernplattformen Inhalte und Ressourcen im Sinne von Open Educational Ressources (OER) geteilt und bearbeitet werden.

3 Pilotierung mit Fachexperten

3.1 Feedback zur Gestaltung

Um die Veröffentlichung der Patterns als Open Educational Ressource und die Möglichkeit ihrer community-basierten Weiterentwicklung innerhalb eines Wikis zu bewerten, wurde eine Pilotierung mit zwölf Experten aus Architektur, Medienplanung und Pädagogik sowie Leitungskräften von Bildungseinrichtungen (Entscheider) durchgeführt. Bis auf die beteiligten Pädagogen, besaßen alle Teilnehmenden mehrjährige Erfahrungen in Planungsprozessen. Alle Beteiligten wurden gebeten, sich durch die Ansicht der Patterns zu navigieren und

im Anschluss schriftlich ihre Bedienbarkeit, Übersichtlichkeit und Attraktivität in einem Fragebogen zu bewerten. Die Ergebnisse der Bewertung zeigen deutlich, dass das Wiki grundsätzlich, unabhängig von der Gruppenzugehörigkeit, vor allem als leicht bedienbar wahrgenommen wird. Kritik an der Verwendung einer akademischen Lernplattform zur Veröffentlichung der Patterns wird vor allem von beteiligten Architekten geäußert, die nicht von der Hochschulnähe des Systems profitieren. Das Fehlen dynamischer Elemente wird von mehreren der Teilnehmenden als veraltet angesehen. Gleichzeitig wird die Reduktion auf das Wesentliche vorn allen Beteiligten befürwortet.

Es zeigt sich, dass der sich wiederholende Aufbau der Patterns Orientierung bietet und vor allem die Auseinandersetzung mit weniger bekannten Themen unterstützt. Die den Text ergänzenden beispielgebenden Bilder stärken die Anschaulichkeit der Patterns zusätzlich. Sie wecken das Interesse und die Neugier der Nutzer und motivieren, sich mit neuen oder bisher unbekannten Themen bzw. Mustern zu beschäftigen. Die verschiedenen Zugangsmöglichkeiten zu den Patterns wurden von den unterschiedlichen Akteuren ebenfalls positiv aufgenommen. Die Beobachtungen in der Bearbeitungsphase zeigen, dass vor allem Architekten die Darstellung der Patterns als Concept-Maps bevorzugen. Die Akteure aus Leitungsebene und der Medienplanung nutzten dagegen vorrangig die beschreibenden Texte.

3.2 Feedback zu den Inhalten der Pattern

Inhaltlich wurden die behandelten Themenfelder der Patterns der Sammlung als relevant wahrgenommen. Vor allem die Darstellung der verschiedenen Sichtweisen der Akteursgruppen auf ein Muster wurde als sehr hilfreich angesehen. Die unterschiedlichen Perspektiven machen es möglich, die für die anderen Gruppen ausschlaggebenden Anforderungen in die eigene Arbeit einbeziehen zu können. In diesem Zusammenhang ist das Aufzeigen von Beispielen für die Teilnehmenden von zentraler Bedeutung. In der Pilotierung wurden pro Pattern drei Beispiele präsentiert. Die Rückmeldung durch die beteiligten Akteure ergab, dass sie sich noch mehr konkrete Fälle gewünscht hätten. Die Möglichkeit der finalen OER-Version, dass Akteure eigene Design Patterns erzeugen oder bestehende weiterentwickeln können, nimmt darauf direkten Bezug. Intensiv wurde über den Zugang über die Rollenansichten (Architektur, Didaktik, Medientechnik diskutiert. Diese werden als Stärke angesehen, wenn es darum geht, einen gezielten Perspektivwechsel vorzunehmen. Gleichzeitig äußern die beteiligten Akteure die Befürchtung, dass eine entsprechende Darstellung das Denken in bestehenden Rollenbildern stabilisiert.

3.3 Gegenseitiges Disziplinverständnis (Common Ground)

Um den Einfluss der Patterns auf die gegenseitige Wahrnehmung zu überprüfen, wurde vor und nach der Bewertung die Einstellung der Beteiligten zu den jeweils anderen Akteursgruppen erfasst. Die offenen Antworten wurden qualitativ mit Hilfe einer Inhaltsanalyse

ausgewertet und mit den Ergebnissen aus dem Posttest verglichen. Dabei zeigt sich, dass fast alle Beteiligten nach der Arbeit mit dem Wiki deutlich mehr Verständnis gegenüber den anderen Akteuren aufbringen als davor. Besonders groß fällt dieser Effekt insgesamt bei der Gruppe der Entscheider und der Pädagogen aus. Beide Gruppen gehören keinen traditionell planenden Disziplinen an, noch sind sie in die konkreten Prozesse der Gebäude- und Medienplanung eingebunden. Entsprechend ist das Wissen über die Arbeitsweisen der anderen (planenden) Akteure gering. Es überrascht daher nicht, dass beide Akteursgruppen besonders stark von den Patterns profitieren. Von Bedeutung für Medienplaner und Architekten sind vor allem Kenntnisse über sich verändernde Lehr- und Lernmethoden und den dazugehörigen Alltag jenseits ihres professionellen und verstetigten Planungswissens. Sie profitieren besonders von den Erfahrungen und Einschätzungen von Pädagogen als Nutzer. Die vorliegenden Ergebnisse zeigen deutlich, wie wichtig es, frühzeitig Informationen über die grundlegenden Aufgaben und fachlichen Anforderungen der beteiligten Akteure zu teilen. Dieses vom jeweiligen Prozess abhängige Wissen ergänzt das spezifische Fachwissen der Beteiligten und erleichtert die prozessbegleitende Kommunikation.

3.4 Rolle des Wikis im Planungsprozess

Neben dem beschriebenen Erkenntnisgewinn in Bezug auf die Aufgaben der beteiligten Akteure, lässt sich bei den beteiligten Pädagogen zusätzlich ein subjektiver Zugewinn beim Prozessverständnis verzeichnen. Dieser konnte auf der Basis von quantitativen Selbsteinschätzungsskalen nachgewiesen werden. Besonders positiv werden die Patterns von allen Beteiligten gesehen, wenn es darum geht, erste Informationen für den Einstieg in den Prozess der gemeinsamen Planung zu teilen. Eine hohe Beteiligungsbereitschaft vorausgesetzt, können vor allem frühe Planungsphasen, in denen Grundlagen und Bedarfe ermittelt werden, von den Patterns profitieren. Ihre darüberhinausgehende Verankerung im laufenden Planungsprozess können sich die Teilnehmenden im Sinne eines „Nachschlagewerks“ vorstellen. Eine entsprechende gemeinsame Wissensbasis hätte das Potential, prozessbezogene Informationen und Hinweise disziplinübergreifend transparent zu machen und Entscheidungs- und Genehmigungsprozesse zu unterstützen.

3.5 Wiki als offenes System

Alle Teilnehmende begrüßten grundsätzlich die Umsetzung planungsunterstützender Instrumente in einem öffentlich zugänglichen Format (OER). Entsprechend wurde die zeitlich und örtlich flexible und freie Verfügbarkeit der Patterns besonders positiv wahrgenommen. Keiner der Beteiligten hatte beim Aufrufen oder Bearbeiten Probleme oder Einschränkungen. Medienplaner können sich die Realisierung der Patterns alternativ auch in einem Datenbank-Managementsystem vorstellen. Vorbehalte meldeten dagegen die beteiligten Akteure gegen die community-basierte Arbeitsweise an. Die Akteure stellten in Frage, ob die aktuell starke inhaltliche Vernetzung zwischen den Pattern auch dann noch aufrecht

erhalten werden kann, wenn der Grad der Verlinkung von den unterschiedlichen Nutzern und deren Initiative, komplexe Zusammenhänge darzustellen, abhängig sein wird. Die Mehrheit der Beteiligten an der Pilotierung hat zudem Bedenken, die Pflege der Inhalte und ihre Validität einem kollaborativen Prozess zu überlassen. Sie wünschen sich eine Überprüfung und fachliche Kontrolle der Inhalte. Ob diese Bedenken berechtigt sind, konnte innerhalb der Pilotierung nicht geklärt werden. Vor allem Entscheider und Medienplaner können sich zudem nur schwer vorstellen, dass die fachlichen Inhalte der Patterns von weiteren Stakeholdern wie u.a. von Nutzern, im Sinne von Nicht-Fachexperten, bearbeitet werden. Die Aussagen zeigen deutlich, dass gegenüber community-basierten Methoden vor allem Vorbehalte bei der Entwicklung von Inhalten bestehen, die sich durch eine hohe Fachlichkeit auszeichnen.

4 Fazit und Ausblick

Community-basierte Planungswerkzeuge als Open Educational Ressources (OER) bieten die Möglichkeit einer breiten Beteiligung. Unterschiedliche Akteure und Stakeholder können disziplin- und einrichtungsübergreifend ihre Erfahrungen und ihre Lösungen in Bezug auf die Planung und Gestaltung von akademischen Lernräumen abbilden und teilen. Entsprechende Sammlungen von Design Patterns können Impulse in Richtung Wissensmanagement und Partizipation geben. Die Experten aus der Pilotierung betonen, dass die Patterns nicht den persönlichen Austausch zwischen den Disziplinen ersetzen, sondern ergänzend eingesetzt werden sollten. Dieser Austausch ist wünschenswert und notwendig vor allem als Teil der Phase 0, zu der eine Bedarfsermittlung ebenso gehört, wie die Auswahl anderer an der Planung fachlich Beteiligter. Zu den Leistungen von Architekten und Ingenieuren, die lt. Honorarordnung, vergütet werden, gehören sie nach wie vor nicht. Ein Jahr nach Ende des Projekts LR_D muss entsprechend festgestellt werden, dass eine Weiterentwicklung der Patterns seitens der unterschiedlichen am Projekt beteiligten Planungsexperten bis dato nur vereinzelt stattfindet.

Die Arbeiten von Farías und Criado (2019) legen daher nahe, die kritische Auseinandersetzung mit der eigenen Rolle und Entwurfs- bzw. Planungspraxis von Experten wie Architekten frühzeitig zu etablieren und bereits in ihrer Ausbildung für eine Demokratisierung von Technikgestaltung zu sensibilisieren [FC19]. Dies gilt nicht nur für die Ausbildung von Architekten. An der TU Dresden setzen sich aktuell vor dem Hintergrund ihrer (Doppel-)Rolle als Nutzer und pädagogische Planungsexperten vor allem Lehrende und Studierende an der Fakultät Erziehungswissenschaft mit der Patternsammlung des Projekts LR_D auseinander.

Im Hochschulbau aber liegen die größten Herausforderungen für die Lehr-Lernraumplanung. Die Höhe des vom HIS-Institut für Hochschulentwicklung berechneten Investitionsstaus im Bereich der Hochschulinstandhaltung (Neubau inbegriffen) belief sich 2016 für den Planungshorizont bis 2025 auf 35 Milliarden Euro [SS16]. Es ist davon auszugehen, dass der auf den Hochschulen lastende Veränderungsdruck in Folge der Covid19-Pandemie

weiter zugenommen hat. Darin liegt eine Chance aber mehr noch eine Notwendigkeit, die Hochschule als akademischen Ort des gemeinsamen Arbeitens und Lernens grundlegend neu zu erfinden [Sc20] sowie geeignete disziplinübergreifende Konzepte und Umgebungen nachhaltig zu etablieren.

Literaturverzeichnis

[AIS78] Alexander, C.; Ishikawa, S.; Silverstein M.: A Pattern Language. Oxford University Press, New York, 1978.

[BS20] Bei der Kellen, D.; Schlenker, L.: Vererbung – Konzept des Co-Designs. In (Plankert, S. ed.): Entwerfen, Lernen, Gestalten. Transcript, Bielefeld, pp. 211-232, 2020.

[CB91] Clark, H. H.; Brennan, S. E.: Grounding in Communication. In (Resnick, B.; Levine, J.M.; Teasley, S. D.; ed.): Perspectives on Socially Shared Cognistion. American Psychological Association, Washington, DC, pp. 127-149, 1991.

[FC19] Farías, I.; Criado, T. S.: Erfahren. Experimente mit technischer Demokratie in Entwurfskursen. In (Marguin, S.; Rabe H.; Schäffner W.; Schmidgall, F.; ed.), Experimentieren Transcript, Bielefeld pp. 67-80, 2019.

[HS19] Hodulak, M.; Schram, U.: Nutzerorientierte Bedarfsplanung. 2. Auflage. Springer Vieweg, S. 11, 2019.

[Ja15] Jahnke, I.: Digital Didactical Designs. Teaching and Learning in CrossActionSpaces. New York, Routledge, 2015.

[Le07] Leitner, H.: Mustertheorie – Einführung und Perspektiven auf den Spuren von Christopher Alexander. Graz, Nausner & Nausner Verlag, 2007.

[NJ18] Ninnemann, K.; Jahnke, I.: Den drittem Pädagogen neu denken. Wie CrossActionSpaces Perspektiven der Lernraumgestaltung verändern. In (Getto, B.; Hinze, P.; Kerres, M.; ed.): Digitalisierung und Hochschulentwicklung. Proceedings zur 26. Tagung der Gesellschaft für Medien in der Wissenschaft e.V. Waxmann, Münster; New York, pp. 135-147, 2018.

[NP18] Nissler, A.; Prey, G.: Neue Lehre – neue Räume? In (Weich, A.; Othmer, J.; Zickwolf, K. ed.): Medien, Bildung und Wissen in der Hochschule. Wiesbaden: Springer, pp. 225–239, 2018.

[Ri13] Rittel, H.: Thinking Design. Transdisziplinäre Konzepte für Planer und Entwerfer. Birkhäuser, Basel, 2013.

[Ri12] Rittel, H.: Die Denkweise von Designern - Studienhefte Problemorientiertes Design, Heft 1. Adocs, Hamburg, 2012.

[Sc20] Schlenker, L.: Die Neuerfindung des Campus – Digitalisierung als Chance für die Hochschule als Lernraum. In (Bauer, R.; Hafer, J.; Hofhues, S.; Schiefner-Rohs, M.; Thillosen, A.; Volk, B.; Wannemacher, K. ed.): Vom E-Learning zur Digitalisierung - Mythen, Realitäten, Perspektiven, Medien in der Wissenschaft, Bd. 76, Münster, Waxmann, pp. 354-362.

[SNB18] Schlenker, L.; Neuburg, C.; Bei der Kellen, D.; Jannack, A.: Partizipativ planen für die berufliche Bildung – Hybride Lernräume gemeinsam gestalten. Konferenzbeiträge der 21. GeNeMe – Konferenz Gemeinschaften in Neuen Medien. Dresden, TUDpress, pp. 150-154, 2018.

[SNK18] Schlenker, L.; Neuburg, C.; Köhler, T.: Thinking in hybrid environments – new classroom concepts for the digital age. In Proceedings of EDULEARN 2018: 10th International Conference on Education and New Learning Technologies. Barcelona, Spain. 2-4 July 2018, pp. 1328-1332.

[St19] Statistisches Bundesamt: Bildung und Kultur. Schnellmeldungsergebnisse der Hochschulstatistik: Online Destatis, 2019.

[SS16] Stribbe, J.; Stratmann, F.: Finanzierungsbedarf für den Bestandserhalt der Hochschulgebäude bis 2025. Forum Hochschulentwicklung, Hannover: HIS-Institut, 2016.

Start des neuen weiterbildenden Masters „Digitales Datenmanagement – DDM“ während des Corona Lockdowns

Die Ad-hoc-Virtualisierung eines Präsenzstudiengangs

Heike Neuroth [1], Stefan Schmunk [2], Ulrike Wuttke [3], Vivien Petras [4]

Abstract: Der Artikel beschreibt die Einführung eines neuen weiterbildenden Masterstudiengangs Digitales Datenmanagement (DDM) während des Corona-Lockdowns, welche zum Studienstart einen sofortigen Wechsel in die online synchrone Lehre umfasste. Der Beitrag fasst wichtige Anforderungen an einen virtuellen Lehrraum zusammen und diskutiert Lessons Learned aus dem ersten erfolgreichen Studiensemester. Eine Besonderheit des Studiengangs ist die Zusammensetzung der Studierendenkohorte, die sich einerseits durch eine ausgeprägte Methoden- und Wissenskompetenz sowie sehr gute technische Ausstattung auszeichnet, andererseits aber hohe Anforderungen an die Inhaltsvermittlung und didaktischen Fähigkeiten der Dozierenden stellt.

Keywords: Digitales Datenmanagement; digitale synchrone Lehre; Best Practices

1 Einleitung

Das Institut für Bibliotheks- und Informationswissenschaft (IBI)[5] der Humboldt-Universität zu Berlin und der Fachbereich Informationswissenschaften (FB 5)[6] der Fachhochschule Potsdam konzipierten in den letzten Jahren den weiterbildenden Masterstudiengang „Digitales Datenmanagement – DDM“[7], der zum 01. April 2020 startete [NRPK19] [PKNR19]. Der Studiengang vermittelt Kompetenzen der Forschungs- und Handlungsfelder im digitalen Datenmanagement und berücksichtigt dabei die nationalen und internationalen wissenschaftspolitischen, organisatorischen, technischen, methodischen und rechtlichen Rahmenbedingungen. Im Mittelpunkt steht der Erwerb von Kompetenzen im analytischen und

[1] Fachbereich Informationswissenschaften, Fachhochschule Potsdam, Kiepenheuerallee 5, 14469 Potsdam, heike.neuroth@fh-potsdam.de, https://orcid.org/0000-0002-3637-3154
[2] Fachbereich Media, Hochschule Darmstadt, Max-Planck-Str. 2, 64807 Dieburg, stefan.schmunk@h-da.de, https://orcid.org/0000-0001-9706-9757
[3] Fachbereich Informationswissenschaften, Fachhochschule Potsdam, Kiepenheuerallee 5, 14469 Potsdam, ulrike.wuttke@fh-potsdam.de, https://orcid.org/0000-0002-8217-4025
[4] Institut für Bibliotheks- und Informationswissenschaft, Humboldt-Universität zu Berlin, Unter den Linden 6, 10099 Berlin, vivien.petras@ibi.hu-berlin.de, https://orcid.org/0000-0002-8113-1509
[5] https://www.ibi.hu-berlin.de/.
[6] https://www.fh-potsdam.de/studium-informationswissenschaften/.
[7] https://www.ddm-master.de/.

doi:10.18420/inf2020_50

praktischen Umgang mit (Forschungs-)Daten aus den Bereichen Wissenschaft, Wirtschaft, Verwaltung und Kultur. Zum Sommersemester 2020 war der Start des Studiengangs mit einer ersten Präsenzphase im ersten Modulkurs „Theoretische Grundlagen Datenmanagement und Data Literacy“ ganztägig am 24. und 25. April 2020 an der HU Berlin geplant.

Mit dem am 27. März 2020 in Kraft getretenen „Gesetz zum Schutz der Bevölkerung bei einer epidemischen Lage von nationaler Tragweite“ [DB20] wurde das Bundesgesundheitsministerium ermächtigt, bundesweit Anordnungen zu treffen. Das Gesetz regelte die rechtlichen Rahmenbedingungen für die Phase des „harten Lockdowns“ (23. März bis 19. April 2020) in Deutschland. Die zuständigen Landesministerien, insbesondere in Berlin und Brandenburg, untersagten den Hochschulen zunächst den Präsenzbetrieb, der aber nach ersten Einschätzungen ab 20. April wieder aufgenommen werden sollte – dies trat jedoch nicht ein. Faktisch bedeutete dies, dass der neue Studiengang nicht wie geplant in Präsenz durchgeführt werden konnte, sondern unmittelbar bei Start alle bereits vorhandenen inhaltlichen, methodischen und didaktischen Konzepte auf digitale Lehrveranstaltungen umgestellt werden mussten. Bis zu diesem Zeitpunkt kannten sich weder die Studierenden noch die Lehrenden untereinander, noch kannten sie sich mit den Studiengepflogenheiten des neuen Studiengangs oder den technischen Infrastrukturen aus.

Im Folgenden werden erste Anforderungsanalysen skizziert und die genutzten didaktischen sowie technologischen Umsetzungsmöglichkeiten vorgestellt. Die bisherigen Erfahrungen werden genutzt, um Best-Practice-Leitfäden für die DDM-Dozierenden sowie für andere Lehrzusammenhänge zu erstellen. Es gilt bei den folgenden Erläuterungen zu berücksichtigen, dass der Studiengang DDM so konzipiert ist, dass ein Modulkurs überwiegend im Rahmen einer zweitägigen Präsenz-Blockveranstaltung mit 20 Stunden Lehre stattfindet.

2 Technische und kulturelle Rahmenbedingungen

Ohne eine stabile, skalierbare und performante technische Infrastruktur ist die Durchführung von digitalen Lehrangeboten bzw. wie im Falle von DDM eine Komplettumstellung auf digitale Lehre undenkbar. Es muss beispielsweise sichergestellt sein, dass die Studierenden während der virtuellen Präsenzlehre ununterbrochen sowohl auf Kursinhalte zugreifen als auch via Video oder Audio an den Lehrveranstaltungen teilnehmen können. Die Entscheidung für eine Umstellung auf eine virtuelle synchrone Präsenzlehre wurde vor allem dadurch motiviert, dass sie die beste Gelegenheit dafür bot, dass sich die Studierenden und Lehrenden, die sich zu diesem Zeitpunkt untereinander noch nicht persönlich kannten, gegenseitig kennenlernen und in den gerade für weiterbildende bzw. berufsbegleitende Studiengänge so wichtigen intensiven Austausch mit- und untereinander treten. Als Grundvoraussetzung für einen interaktiven und diskursiven Austausch während der virtuellen Präsenz stellten sich eine gute didaktische Vorbereitung und die Fokussierung auf die Methoden- und Wissenskompetenzen der Teilnehmenden heraus.

2.1 Die digitale Lernumgebung für jeden Kurs

Eine wesentliche Voraussetzung, um einen Komplettumzug von analoger zu digitaler Lehre zu ermöglichen, ist eine stabile, funktionale und mit einem Rollen- und Rechtemodell versehene Lernmanagementplattform, die darüber hinaus genügend Speicherplatz bietet, damit umfangreiche Lehrmaterialien, wie z. B. Videos, hinterlegt werden können. Mit dem an der FH Potsdam eingesetzten Moodle-System[8] und den darin integrierten Tools für Wikis, Foren, Etherpads, Umfragen etc. war eine zuverlässige, technisch stabile Plattform vorhanden, die aus Datenschutzsicht unbedenklich ist.

2.2 Der virtuelle Klassenraum

Es war völlig klar, dass ohne ein stabiles und gemeinhin akzeptiertes Videokonferenzsystem die digitale Lehre im Sommersemester 2020 mit virtuell-synchronen Lehreinheiten von 20 Stunden verteilt über zwei Tage nicht zu leisten ist. Dabei standen bei der Auswahl der Videokonferenzsystems die folgenden Vorüberlegungen auf der Feature-Wunschliste:

- Mindestens 50 Teilnehmende pro Lehrsitzung mussten sowohl aktiv als auch passiv teilnehmen und audio-visuell miteinander interagieren können.
- Die (technische) Infrastruktur musste nachweisbar über 10 Stunden stabil laufen.
- Sollte eine Video-Teilnahme auf Seiten eines Studierenden nicht möglich sein, musste eine Teilnahme über Telefon gewährleistet sein.
- Dozierende und Studierende sollten die Möglichkeit zum Teilen des Bildschirms haben.
- Die Möglichkeit, Studierende in Kleingruppen aufteilen zu können (z.B. via Aufteilung in virtuelle verteilte Videokonferenzräume), ohne die eigentliche Sitzung zu unterbrechen, war eine didaktische Anforderung. Dadurch sollte (spontane) Gruppenarbeit ermöglicht werden, z. B. um Fragestellungen in kleineren Teams zu bearbeiten.
- Eine Teilnahme sollte auch ohne Herunterladen einer bestimmten App und ohne Registrierung ermöglicht werden können.
- Die Möglichkeit, Co-Verantwortliche für die jeweilige Sitzungsleitung zu benennen – also ein konfigurierbares Rollen- und Rechtemodell –, sollte gegeben sein. Auf diese Weise sollte verhindert werden, dass Sitzungen ungeplant geschlossen werden, wenn beispielsweise die Internetverbindung der Lehrperson abbricht. Andererseits sollte die Sitzungsleitung zentrale administrative Rechte so vergeben können, dass sich keine Unbefugten Zutritt zum virtuellen Klassenraum verschaffen können bzw. unerwünschte Inhalte teilen (z. B. „Zoombombing“).

[8] https://moodle.de/.

- Und natürlich sollten keine datenschutzrechtlichen Bestimmungen gegen den Betrieb des Videokonferenzsystems sprechen, wobei dies tatsächlich zu Beginn des Studiengangs im März 2020 die größte Hürde darstellte und teilweise immer noch darstellt [BADI20].

Zum jetzigen Zeitpunkt (September 2020) kann festgehalten werden, dass es weltweit kein System gibt, welches alle der oben genannten Anforderungen erfüllt.

Die letztendliche Entscheidung für Zoom[9] als Videokonferenzsystem für die digitale Lehre und digitalen Konsultationstermine im DDM-Studiengang wurde basierend auf der oben genannten Feature-Wunschliste im Vergleich zu anderen Systemen getroffen. Die grundsätzliche Entscheidung, ein Videokonferenzsystem einzusetzen und die Lehre in digital synchronen Sitzungen durchzuführen, hat sich positiv ausgewirkt. Kein Studierender hat das Studium abgebrochen und alle Studierenden konnten seit Anfang April 2020 mit dem neuen Studiengang und den technischen Implikationen der digitalen Lehre vertraut gemacht werden. Ein wichtiges Erfolgskriterium ist aber vor allem, dass bisher alle Studierenden an den Veranstaltungen mit Video und Ton teilnehmen konnten und die technische Infrastruktur von Zoom äußerst stabil und ausfallsfrei lief. Auf diese Weise waren wesentlich mehr kommunikative Formen der Interaktion möglich, als wenn z. B. ausschließlich über Chatfunktionen kommuniziert worden wäre.

Unter Datenschutz-Gesichtspunkten wurde Zoom zum Teil aus unterschiedlichen Gründen kritisiert [Ri20] [Sch20]. Allerdings sind zahlreiche Kritikpunkte bereits während des Sommersemesters 2020 behoben worden. Zudem wurden während des Sommersemesters von einer Vielzahl an Videokonferenzsystemen Nachbesserungen, sowohl was die Erweiterung der Funktionalitäten in Bezug auf klar artikulierte didaktische als auch datenschutzrechtliche Anforderungen betraf, getroffen. Mittlerweile wurden von einzelnen Hochschulen bzw. sogar ganzen Bundesländern eine Reihe von Lizenzierungen durchgeführt – wie beispielsweise das Master Subscription Agreement in Bayern[10] –, die eine Datenschutzkonformität herstellten und somit die kommenden Semester aus datenschutzrechtlicher Perspektive auf eine solide rechtliche Basis stellen.

2.3 Organisatorische und sozio-kulturelle Rahmenbedingungen

Die technischen Rahmenbedingungen für den DDM-Master sind zu gleichen Teilen von den zu vermittelnden Themen und den bereits vorhandenen Methoden- und Wissenskompetenzen der Studierenden abhängig. Eine Besonderheit des DDM-Studiengangs ist, dass es sich um einen qualifizierenden, weiterbildenden Masterstudiengang handelt. Aus diesem Grund weisen alle Studierenden nicht nur umfangreiche Praxiskompetenzen in unterschiedlicher thematischer Ausprägung und Tiefe auf, sondern sind zugleich aufgrund ihrer beruflichen Tätigkeiten – gerade unter Corona-Bedingungen – in einem hohen Maße an die Nutzung

[9] https://zoom.us/.

[10] https://www.rz.uni-wuerzburg.de/dienste/it-recht/it-vertraege/zoom/.

digitaler Werkzeuge gewöhnt. Die Studierenden waren und sind in unterschiedlichen Berufsfeldern tätig, was in der Kombination zu unterschiedlichen Kenntnisständen und Methodenkompetenzen bei einzelnen Themenfeldern führte, die sich jedoch im Rahmen des Plenums, der Gruppen und der Tandems gut ausglichen.

Diese soeben beschriebenen Besonderheiten haben u. a. dazu geführt, dass bei den im Sommersemester 2020 durchgeführten Modulen alle Studierenden dauerhaft per Video an den Veranstaltungen teilnahmen. Allein aufgrund dieser Konstellation und einer Gruppenstärke, die 30 Studierende nicht überstieg, konnten bei den virtuellen synchronen Präsenzveranstaltungen Gestik und Mimik – und somit auch non-verbale Kommunikationsformen – sowohl von den Studierenden als auch von den Lehrenden gegenseitig zumindest zum Teil wahrgenommen und dadurch die Interaktion gesteigert werden. Dies war eine Grundvoraussetzung, um den virtuell-synchronen Lehransatz in diesem Umfang erfolgreich umzusetzen.

Um den Ablauf der Einzelmodule nicht mit organisatorischen Aspekten zu überfrachten, wurden im Vorfeld fünf bis sechs Personen umfassende Arbeitsgruppen verbindlich festgelegt, ebenso Tandems von jeweils zwei Studierenden. Auf diese Weise konnte eine kooperative Arbeitsstruktur und -kultur etabliert werden, die im Vorfeld bzw. im Nachgang zu den virtuellen Präsenzveranstaltungen einen hohen Grad an Verbindlichkeit generierte und den Austausch unter den Teilnehmenden um die Präsenzveranstaltungen herum stark förderte.

Hervorzuheben ist die außerordentliche Bereitschaft der Studierenden, sich aktiv in die synchrone Lehre einzubringen und diese für die gesamte Dauer der einzelnen Lehrveranstaltungen interaktiv mitzugestalten. Dies ist sicherlich keine Selbstverständlichkeit, da die Lehrveranstaltungen eine tägliche Dauer von geplanten 10 Stunden (einschließlich Pausen) hatten, die einzelnen Module an zwei aufeinanderfolgenden Tagen durchgeführt wurden und es sich zugleich ausnahmslos um Studierende handelte, die das Studium neben einer eigenen Berufstätigkeit durchführen.

Zugleich ist festzustellen, dass die Erwartungshaltungen an das Studium aufgrund des unterschiedlichen disziplinären Backgrounds der Studierenden – von geistes- und kulturwissenschaftlichen über naturwissenschaftliche bis hin zu rechts- und verwaltungswissenschaftlichen Abschlüssen – sehr unterschiedlich waren. Insgesamt zeichneten sich die DDM-Studierenden des ersten Jahrgangs durch hohe Einsatzbereitschaft und überdurchschnittliches Engagement aus. Diese Eigenschaften sind kennzeichnend für fast alle weiterbildenden und berufsbegleitenden Studiengänge, bei denen Studierende, die bereits über einen Studienabschluss verfügen (BA, MA) bzw. teilweise schon promoviert sind, sich in der Regel zielgerichtet zur Erlangung eines weiterführend qualifizierenden Studienabschlusses immatrikulieren und dazu bereit sind, hierfür Studiengebühren zu bezahlen. Durch die hohe intrinsische Motivation und Methoden- und Wissenskompetenz konnten sich die DDM-Studierenden schnell auf die im Frühjahr 2020 verändernden Rahmenbedingungen einstellen.

3 Werkzeuge für die digitale Lehre

Generell gilt auch beim E-Learning, trotz der nahezu unbegrenzten Möglichkeiten und Vielfalt, aus Sicht der Autor*innen „Weniger ist mehr“. Aus Gründen der Stabilität, des Datenschutzes und der Usability sollte möglichst ein System genutzt werden, in dem alle zentralen Informationen zusammenfließen. Dazu können dann einige ausgewählte Werkzeuge zur Aktivierung kommen.

Die Kommunikation in den Modulen selbst erfolgte möglichst über ein Moodle-Forum, das alle Kursmitglieder mitlesen können, da auf diese Weise prinzipiell alle eingeschriebenen Kursmitglieder (Studierende, Co-Dozierende) gleichzeitig erreicht werden können. Zudem werden über das Forum alle Fragen und Antworten und somit der gesamte Kommunikationsverlauf dokumentiert, so dass Teilnehmende, die an einzelnen Sitzungen aufgrund privater bzw. beruflicher Umstände nicht teilnehmen können, diese Inhalte jederzeit zeitversetzt rezipieren können.

Alle für die Veranstaltung relevanten Unterlagen wurden den Teilnehmenden mindestens vier Wochen vor dem Start der jeweiligen Lehrveranstaltung zur Vorbereitung über Moodle als zentrale Lernmanagementplattform zur Verfügung gestellt. Ergänzend wurden kleinere Clips, Screencasts, Aufgaben und weitere Lehr- und Lernmaterial über Moodle zugänglich gemacht sowie der Link zu einer gemeinsamen Online-Bibliografie (Zotero[11]), in der für jedes Modul durch die Dozierenden ausgewählte Grundlagen- und Vertiefungsliteratur eingepflegt wurde, zur Verfügung gestellt.

Bereits vor und während der einzelnen Veranstaltungen konnten die Studierenden zudem in speziell dafür in Moodle angelegten Etherpads Fragen, Anmerkungen und Kommentare hinterlegen und diskutieren. Auch für Arbeitsgruppensitzungen wurden in Moodle gruppenspezifische Etherpads angelegt, so dass nach Ende der Veranstaltung Ergebnisse und Diskussionsverläufe dokumentiert waren. Dies ermöglicht die Nachbereitung durch die Studierenden und ist zugleich als Wissensrepräsentation aller durchgeführten Module nutzbar.

Zusätzlich zu den in Moodle vorhandenen Applikationen wurden eine Reihe weiterer digitaler Werkzeuge für Q&A-Sessions, kleinere Quizze und (spontane) Abstimmungen während den virtuellen Präsenzphasen verwendet, unter anderem in Zoom selbst, aber auch von anderen Anbietern wie Slido[12], Mentimeter[13], Pingo[14] und Kahoot[15]. Alle genannten Werkzeuge haben den Vorteil, dass die Lehrenden die entsprechenden Interaktionen mit den Studierenden entweder im Vorfeld vorbereiten oder kurzfristig während der Lehre erstellen können. Zudem ist ihre Nutzung niedrigschwellig und ohne umfassende Einführungen möglich.

[11] https://www.zotero.org/groups/2235292/digitales_datenmanagement_-_weiterbildender_masterstudiengang/library.
[12] https://www.sli.do/.
[13] https://www.mentimeter.com/.
[14] https://trypingo.com/.
[15] https://kahoot.com.

4 Best Practices

Allgemein ist festzuhalten, dass grundsätzlich auf eine sehr gute Tonqualität (Headset, Mikrofon, Nebengeräusche durch konsequente Stummschaltung aller Zuhörenden vermeiden etc.) zu achten ist. Die Bildqualität ist nur von sekundärer Bedeutung [KWR97].

Sofern die oben genannten Rahmenbedingungen eingehalten werden, unterscheiden sich die präferierten synchronen virtuellen Veranstaltungen nur graduell von einer Präsenzlehre. Die Lernwirksamkeit synchroner virtueller Lehre ist vergleichbar zu Live-Settings, Engagement und Zufriedenheit im letzteren Fall allerdings höher, was seitens der Lehrenden mehr Überlegungen für virtuelle, aktivierende, partizipative Elemente erfordert [LM09] [Sk09]. Allerdings lassen sie deutlich weniger Raum für Improvisationen und erfordern dadurch eine stärkere Auseinandersetzung mit didaktischen Konzepten wie z. B. der Bloomschen Taxonomie [Ju16:114*ff*] sowie Constructive Alignment [KHK20: 26*ff*]. Die Sessions müssen gut geplant sein (Thema, Methode und Sozialform, Werkzeug, Zeit) und es ist im Besonderen auf eine nachvollziehbare, klare und die Gruppe einbeziehende Kommunikation zu achten. Aber dies sind letztlich Aspekte, die auch bei einer guten Präsenzlehre zu berücksichtigen sind. Es müssen längere Pausen als in der Präsenzlehre eingeplant und zudem extra Zeit für das Hin- und Herschalten zwischen verschiedenen Systemen (Moodle, Zoom etc.) berücksichtigt werden. Neben der oben genannten Videoausstattung sind zwei Bildschirme für die Dozierenden sowie die Studierenden ideal.

Neben der inhaltlichen Aufbereitung der Themen muss darauf geachtet werden, dass Erwartungshaltungen moderiert und permanent abgefragt werden. Dies erfolgte in den Präsenzveranstaltungen über die verfügbaren Abfragewerkzeuge oft auch anonym, so dass Studierende offen auch soziale bzw. andere Probleme (mehr Pausen benötigt) ausdrücken konnten. Diese diversen Möglichkeiten zur (anonymen) Meinungsbekundung stellten gegenüber physischen Präsenzveranstaltungen sogar einen Vorteil dar, insbesondere in einer Gruppe, die sich vorher nicht kannte.

4.1 Best Practices für die Zeit vor dem Kursbeginn

Vor Kursbeginn sollten folgende Punkte berücksichtigt werden:

- Bereitstellung der Kursmaterialien über Moodle mindestens vier Wochen vor Beginn der Kursveranstaltung, um Planbarkeit herzustellen.
- Möglichst viele für die Präsenzveranstaltungen relevanten Materialien bereits im Vorfeld rechtzeitig freischalten. Hierzu zählen besonders Foliensätze und Präsentationen, beispielsweise bei thematischen Einführungen, damit die Studierenden diese im Vorfeld rezipieren und zugleich parallel während des Lehrvortrages nutzen können.

- Willkommensnachrichten mit Vorstellung und Ermunterung zur Teilnahme an vorbereitenden Klärungsmöglichkeiten (Erwartungshaltung, z. B. über ein Etherpad oder eine Mentimeter-Abfrage) über das Moodle-Forum.
- Abfrage der Erwartungshaltungen der Teilnehmenden beispielsweise über anonymisierte „Fragerunden“ mittels der benannten Werkzeuge (vor und während der Präsenzveranstaltungen).
- Für kleinere Aufgaben, die im Kurs bearbeitet werden sollen (stille Lektüre, Tandem- oder Gruppenarbeit), sollten die „Aufgabenblätter“ und ggf. weitere Materialien vorher unsichtbar in Moodle eingestellt und erst zum Kursbeginn bzw. vor dem Stellen der Aufgabe sichtbar geschaltet werden. Dadurch ist für die Studierenden ein Zugriff auf eine schriftliche Form möglich und die Aufgabenstellungen können mehrfach rezipiert werden. Insbesondere ist auf eindeutige Aufgabenstellungen mit klaren Anweisungen und Rollenverteilungen zu achten [MAB17].
- Im Moodlekurs ein Etherpad für gemeinsame Notizen während der Präsenzphase anlegen.

4.2 Best Practices für Durchführung des Kurses

Während des Kurses ist auf folgende Punkte zu achten:

- Am Anfang der virtuellen Sitzung viel Zeit einplanen, um sich gegenseitig kennenzulernen (Schaffung einer kommunikativen Atmosphäre) und die Rahmenbedingungen zu klären, z. B.:
  - Vorstellung Dozent*in (und ggf. der Studierenden)
  - Icebreaker (Quiz, Statements zur Erwartungshaltung, Umfrage etc.)
  - Code of Conduct (wie gehen wir miteinander um, anonyme Abstimmung zu Duzen oder Siezen)
  - Vorstellung des virtuellen Klassenraums mit Funktionen
  - Vorstellung Netiquette
- Permanent auf das Etherpad für die gemeinsamen Notizen hinweisen, da sonst alle Informationen im Chat landen (und nach Ende der Sessions verloren gehen). Auch hat es sich als hilfreich erwiesen, für einzelne Sessions „Protokollant*innen“ zu benennen.
- Zu einer aktiveren Gestaltung des Ablaufs, um eine dauerhafte Aufmerksamkeit zu erhalten, können u. a. Formatwechsel (Vortrag, Übung), Wechsel der Gruppenstärken (Plenum, Arbeitsgruppen, Tandem) sowie gelegentliche Kanalwechsel (von der Videokonferenz zum Etherpad) neben den inhaltlichen Themenwechseln (und ausreichend Pausen) beitragen.

- Für kleinere digitale Aktivierungen [Sa20] hat sich die Nutzung von Werzeugen zur Durchführung von Quizzen, Wissenstests, Umfragen und Abstimmungen, sogenannte Audience Response Systems, als sehr wichtig erwiesen [KL09].
- Es sollte eine Pufferzeit für unerwartete technische Störungen bzw. das Umschalten zwischen technischen Umgebungen (Moodle, Zoom, Mentimeter) eingeplant werden.

4.3 Best Practices für die Zeit zwischen Modulkursen

Zwischen den Kursen ist auf folgende Punkte einzugehen:

- Dozierenden-Konsultationstermine: Regelmäßiges Feedback der Dozierenden, die bereits einen Kurs abgehalten haben, mit Dozierenden, die erst noch ihre Lehreinheit vor sich haben. So gelingt ein Wissenstransfer durch Lessons-Learned.
- Studierenden-Konsultationstermine: Regelmäßige digitale Konsultationsrunden mit den Studierenden. Diese können in loser Reihenfolge am Beginn des Semesters zu bestimmten Themen erfolgen (z. B. zur Nutzung von Moodle oder Zoom, Erwartungen an Prüfungen) und im Laufe des Semesters als generelle Feedback-Möglichkeit zur digitalen Lehre weitergeführt werden. Durch die regelmäßigen Konsultationen mit den Studierenden war es im DDM-Master unter anderem möglich, direkt auf Kritikpunkte zu reagieren, (technische) Hemmnisse zeitnah zu beseitigen, Verbesserungsvorschläge für die nachfolgenden Kurse aufzugreifen und in den oben genannten Dozierenden-Konsultationsterminen zu besprechen.
- Lehrevaluationen zu jedem Kurs direkt im Rahmen der jeweiligen Lehreinheit: Diese Evaluationen wurden sofort ausgewertet, den Dozierenden und Studierenden zurückgemeldet und bei Bedarf in den regelmäßigen Dozierenden- und Studierendenrunden besprochen. Der Rücklauf liegt bisher bei über 90% und bisher sind alle Kurse mit sehr gut bewertet worden.
- Erstsemester-Evaluation: In der Mitte des Sommersemesters ist ebenfalls die Erstsemester-Evaluation erfolgt und hat sich insbesondere mit dem Erwartungsmanagement des gesamten Studiengangs auseinandergesetzt. Auch wurde nach der Akzeptanz der digitalen Lehre gefragt. Obwohl die Vorteile der Präsenzlehre – insbes. der bessere Austausch - erkannt werden, haben im DDM-Studiengang zwei Drittel der Studierenden signalisiert, digitale Lehrformate zu bevorzugen, wenn die Lehre keine Präsenzen erforderlich macht. Da der Lebensmittelpunkt der Studierenden für diesen Studiengang in ganz Deutschland verteilt ist und diese teilweise nur für die Präsenztermine an die Hochschulstandorte kämen, ist die Ersparnis des Transits (aus ökonomischen, Umwelt- und Zeitersparnisgründen) ein nicht zu unterschätzender Faktor. Die Antworten lieferten ebenfalls wertvolles Feedback für die weitere digitale Gestaltung des Wintersemesters.

5 Zusammenfassung

Zusammenfassend ist festzuhalten, dass die Durchführung der einzelnen Lehrveranstaltungen des Studiengangs DDM als digitale synchrone Veranstaltungen im Vergleich zu traditionellen Live-Lehrveranstaltungen sowohl zu Mehraufwand bei den Dozierenden als auch bei den Studierenden geführt hat. Während die Dozierenden vor allem ihre Sitzungen kleinteiliger vorbereiten mussten und auf Detailebene im Vorfeld das Zusammenspiel von Thematik und Didaktik auf die Möglichkeiten sowie Grenzen des Zusammenspiels der digitalen Systeme abstimmen mussten – und dies auch oftmals im Vorfeld der Lehrveranstaltungen erstmalig evaluieren und erproben mussten – ergab sich bei den Studierenden insbesondere ein Mehraufwand in der Vor- und auch Nachbereitung der einzelnen Sitzungen, aber auch der Gruppen- und Tandemarbeiten, die ebenfalls online abgestimmt werden mussten. Insbesondere der Grad der Eigenverantwortlichkeit liegt hier wesentlich höher, da digitale Lehre nur dann gelingen kann, wenn diese durch einen selbständigen, durch die Studierenden aktiv gestalteten und zugleich problemorientierten Lernprozess begleitet wird. Aber faktisch unterscheidet sich dies nicht wirklich strukturell von der klassischen Lehre im kritischen Diskurs im „physischem" Angesicht zu Angesicht, sondern gewinnt nur an gradueller Bedeutung. Ein weiterer hier anzuführender Aspekt ist sicherlich der konstant notwendige Austausch über die Studiengangsorganisation, die ebenfalls einen höheren Grad an Verbindlichkeit benötigt.

Zugleich stellte sich heraus, dass digitale Lehre eine Vielzahl an neuen und höchst vielfältigen didaktischen und methodischen Ansätzen bietet, die durch die o.g. digitalen Systeme ermöglicht werden und die durchaus die Chance haben, universitäre Lehre und das gegenseitige Lernen miteinander neu zu denken bzw. zu überdenken. Dies gilt insbesondere für den schnellen Wechsel zwischen Vorlesung, Einzelübungen, Team- und Gruppenarbeiten etc., die im physischen Raum in dieser Geschwindigkeit nur bedingt realisierbar sind und die zu einer neuen und ungewohnten Dynamik führen. Aber auch die reflektierte und zielgerichtete Nutzung digitaler Systeme, die als digitale didaktische „Ermöglicher" zielgruppenadäquat Wissen und vor allem Methoden vermitteln können, führte zu einem fruchtbaren Austausch zwischen den Dozierenden und einem höheren Variantenreichtum in der didaktischen Vermittlung, welches positiv von den Studierenden wahrgenommen wurde. Die Corona-Pandemie trägt auf diesem Wege sicherlich dazu bei, dass universitäre Lehre, die faktisch bis zum Beginn des Jahres 2020 oftmals näher am idealisierten Humboldt'schen Ideal des beginnenden 19. Jahrhunderts als an den Möglichkeiten des 21. Jahrhunderts verhaftet war, einen Innovationsschub durchlebt, die neue digitale Wege des Lehrens und Lernens ermöglicht.

Literaturverzeichnis

[BADI20] Berliner Beauftragte für Datenschutz und Informationsfreiheit: Hinweise für Berliner Verantwortliche zu Anbietern von Videokonferenz-Diensten. Version 1.0 vom 3. Juli 2020, https://www.datenschutz-berlin.de/fileadmin/user_upload/pdf/orientierungshilfen/2020-BlnBDI-Hinweise_Berliner_Verantwortliche_zu_Anbietern_Videokonferenz-Dienste.pdf, Stand: 14.07.2020.

[DB20] Deutscher Bundestag (19. Wahlperiode): Entwurf eines Gesetzes zum Schutz der Bevölkerung bei einer epidemischen Lage von nationaler Tragweite. Drucksache 19/18111, 24.03.2020, https://dipbt.bundestag.de/doc/btd/19/181/1918111.pdf, Stand: 14.07.2020.

[Ju16] Jung, E.: Kompetenzerwerb: Grundlagen, Didaktik, Überprüfbarkeit. Oldenbourg, München, 2010.

[KL09] Kay, R.H. & LeSage, A.: Examining the benefits and challenges of using audience response systems: A review of literature. Computers & Education 53:3, S. 819-827, 2009. https://doi.org/10.1016/j.compedu.2009.05.001.

[KHK20] Kergel, D.; Heidkamp-Kergel, B.: E-Learning, E-Didaktik und digitales Lernen. Springer, Berlin, Wiesbaden, 2020.

[KWR97] Kies, J.K., Williges, R.C. & Rosson, M.B.: Evaluating desktop video conferencing for distance learning. Computers & Education 28:2, S. 79-91, 1997.

[LM09] Lietzau, J.A. & Mann, B.J.: Breaking out of the Asynchronous Box: Using Web Conferencing in Distance Learning. Journal of Library & Information Services in Distance Learning 3:3-4, S. 108-119, 2009. https://doi.org/10.1080/15332900903375291.

[MAB17] Martin, F., Ahlgrim-Delzell, L. & Budhrani, K.: Systematic Review of Two Decades (1995 to 2014) of Research on Synchronous Online Learning., American Journal of Distance Education 31:1, S. 3-19, 2017. https://doi.org/10.1080/08923647.2017.1264807.

[NRPK19] Neuroth, H.; Rothfritz, L.; Petras, V.; Kindling, M.: Digitales Datenmanagement als neue Aufgabe für wissenschaftliche Bibliotheken. Bibliothek Forschung und Praxis 43:3, S. 421-431, 2019, Preprint: https://doi.org/10.18452/20680.

[PKNR19] Petras, V.; Kindling, M.; Neuroth, H.; Rothfritz, L.: Digitales Datenmanagement als Berufsfeld im Kontext Data Literacy. ABI Technik 39:1, S. 26-33, 2019, https://doi.org/10.1515/abitech-2019-1005.

[Ri20] Ries, U.: Videokonferenz-Software: Ist Zoom ein Sicherheitsalptraum? Heise Online, 02.04.2020, https://www.hei-se.de/security/meldung/Videokonferenz-Software-Ist-Zoom-ein-Sicherheitsalptraum-4695000.html, Stand: 14.07.2020.

[Sa20] Salmon, G.: E-tivities, https://www.gillysalmon.com/e-tivities.html, Stand: 14.07.2020.

[Sch20] Schmidt, J.: Kommentar: „Warum prügelt ihr alle auf Zoom ein?“ Heise Online, 02.06.2020, https://www.heise.de/meinung/Kommentar-Warum-pruegelt-ihr-alle-auf-Zoom-ein-4771225.html, Stand: 14.07.2020.

[Sk09] Skylar, A.A.: A Comparison of Asynchronous Online Text-Based Lectures and Synchronous Interactive Web Conferencing Lectures. Issues in Teacher Education 18, S. 69-84, 2009.

neue Szenarien und Formate des mediengestützten Lehrens und Lernens,

Werkzeuge, Architekturen und Infrastrukturen für innovative Lehr-/Lernszenarien

Das Stud.IP ePortfolio-Plugin als digitaler Lern- und Prüfungsort in der Lehrer*innenbildung

Yvette Völschow[1], Julia-Nadine Warrelmann[2], Stefanie Brunner[3]

Abstract: In diesem Beitrag werden neben Gemeinsamkeiten und Unterschieden von digitalen gegenüber Papier-Portfolios in der Lehrer*innenbildung Einsatzmöglichkeiten, Herausforderungen und Chancen im Zusammenhang mit der Entwicklung und Nutzung eines ePortfolio-Plugins beschrieben und diskutiert. Zudem wird sich mit technischen und hochschuldidaktischen Komponenten der Implementierung befasst. Neben theoretischen Befunden werden hierfür eigene Erfahrungen aus dem von der Qualitätsoffensive Lehrerbildung (QLB) des BMBF geförderten Projekt BRIDGES der Universität Vechta herangezogen und Handlungsempfehlungen hieraus abgeleitet.

Keywords: digitales Portfolio; ePortfolio; Portfolio-Plugin; Lehrer*innenbildung; Kompetenzentwicklung; Reflexivität; Reflexionsförderung

1 Einleitung

Die Covid-19-Pandemie stellt die Hochschullandschaft in Deutschland seit März 2020 vor besondere Herausforderungen, auf die u.a. mit einer zuvor kaum vorstellbar zügig voran getriebenen Digitalisierung reagiert wurde und wird. Einen entsprechenden Beitrag hierzu lieferte bereits vor der aktuellen Pandemie-Situation auch das von der BMBF Qualitätsoffensive Lehrerbildung geförderte und an der Universität Vechta durchgeführte Projekt „BRIDGES – Brücken bauen, Zusammenarbeit initiieren und gestalten“. Im Rahmen dieses Projektes wird seit 2016 an der Entwicklung und bis 2023 an der Implementierung eines ePortfolio-Tools als Plugin gearbeitet. Das „elektronisch gestützte Kompetenzentwicklungsportfolio“, im Folgenden eKEP genannt, wurde für das Informations-, Campus-, Projekt- und Lernmanagementsystem Stud.IP entwickelt, das an fast jeder dritten Universität Deutschlands genutzt wird.

Im vorliegenden Beitrag wird das eKEP vorgestellt und aufgezeigt, was es aus technischer und hochschuldidaktischer Perspektive leisten kann. Die Entwicklung und Implementation

[1] Universität Vechta, Arbeitsstelle für Reflexive Person- und Organisationsentwicklung, Driverstr. 22, 49377 Vechta, yvette.voelschow@uni-vechta.de
[2] Universität Vechta, QLB Projekt „BRIDGES“, Driverstr. 22, 49377 Vechta, julia-nadine.warrelmann@uni-vechta.de
[3] Universität Vechta, QLB Projekt „BRIDGES“, Driverstr. 22, 49377 Vechta, stefanie.brunner@uni-vechta.de

doi:10.18420/inf2020_51

werden ebenso dargestellt wie die konkrete Anwendung in ausgewählten Modulen der lehramtsrelevanten Fächer Designpädagogik und Mathematik. Über entsprechende empirische Ergebnisse werden Erfahrungen der Nutzer*innen und hieraus resultierende weitere Bedarfe und Erkenntnisse erläutert sowie ein Fazit für ePortfolioarbeit im Allgemeinen und die technische Weiterentwicklung des Stud.IP-Plugins im Speziellen gezogen.

Da der Projektabschluss erst auf das Jahr 2023 datiert ist, wird der Beitrag als „Work in progress“ im Themenbereich neue Szenarien und Formate des mediengestützten Lehrens und Lernens sowie Werkzeuge, Architekturen und Infrastrukturen für innovative Lehr-/Lernszenarien verankert.

2 Hintergrund und theoretischer Rahmen

2.1 Das Projekt BRIDGES

„Brücken bauen. Zusammenarbeit initiieren und gestalten“ ist der Name und zugleich das übergeordnete Ziel des Projekts BRIDGES[4]. Durch unterschiedliche inhaltlich und strukturell verankerte, fächer-, institutionen- und phasenübergreifende Brücken sollen Strukturen geschaffen werden, um interdisziplinäre und praxisbezogene Forschung in der Lehrerbildung zu fördern, den Professionalisierungsprozess auch mit Blick auf die Reflexionsförderung angehender Lehrkräfte zu unterstützen und dadurch die Qualität der Lehrerbildung – gerade im Hinblick auf aktuelle bildungspolitische Aufgaben – zu stärken.

In der ersten Förderphase (2016-2019) wurde in zwei BRIDGES-Teilprojekte differenziert. Im Teilprojekt I "Forschungswerkstatt Inklusion" arbeiteten Wissenschaftler*innen, Promovierende und Studierende fächerübergreifend und in Kooperation mit regionalen Vertreter*innen aus Schulen und Fördereinrichtungen zusammen. Ziel war die Erforschung, Entwicklung und Dokumentation von Lernumgebungen und didaktischen Konzepten für einen inklusiven (Fach-)Unterricht. Von der Unterrichtspraxis ausgehend wurden im Rahmen verschiedener Promotionsprojekte inklusive und heterogene Lernumgebungen thematisiert, didaktische Konzepte entwickelt, erprobt und publiziert (vgl. z.B. Herkenhoff, 2020; Gummels, 2020).

Im Rahmen des Teilprojekts II wurde die Förderung von Beratungs- und Selbstreflexionskompetenzen angehender Lehrkräfte u.a. über die Entwicklung verschiedener themenspezifischer Lehr- und Fortbildungsangebote begleitet und unterstützt. Um Studierenden während ihres gesamten universitären Professionalisierungsprozesses eine nicht zuletzt reflexionsfördernde Begleitung anbieten zu können, fand in dem Zusammenhang auch die Konzeption des elektronischen Kompetenzentwicklungsportfolios (eKEP) statt, dessen Entwicklung und Implementation in Kapitel 3 ausführlicher beschrieben wird.

[4] „BRIDGES – Brücken bauen“ der Universität Vechta wird im Rahmen der gemeinsamen ‚Qualitätsoffensive Lehrerbildung‘ von Bund und Ländern mit Mitteln des Bundesministeriums für Bildung und Forschung gefördert.

Seit 2019 werden in einer zweiten Förderphase bis 2023 die geschaffenen Brücken, Strukturen und Inhalte für eine Erweiterung der Aufgaben genutzt und gefestigt. In dem Zusammenhang soll das eKEP als studienphasenübergreifendes und -begleitendes Element auf verschiedene Module in der Bachelor- und Masterebene der Lehramtsausbildung ausgeweitet und konzeptionell weiterentwickelt werden.

2.2 Vom Papier- zum digitalen Portfolio: Das ePortfolio als digitaler Lehr-Lern- und Prüfungsort

Einer der Ursprünge des eKEP liegt in der Portfolioarbeit. Ein Portfolio – ob in Papier- oder in digitaler Form – soll den Lernprozess durch Dokumentation und Reflexion unterstützen (Häcker, 2005). Es gilt ferner als eine Methode zur Professionalisierung, z.B. von pädagogischen Fachkräften angesichts wachsender Heterogenität in Lehr-Lern-Kontexten (Hofmann et al., 2016). Dabei handelt es sich bei der Portfolioarbeit generell um „eine zielgerichtete und systematische Sammlung von Arbeiten unterschiedlicher Art, die die individuellen Bemühungen, Fortschritte und Leistungen des Lernenden in einem oder mehreren Lernbereichen darstellt und reflektiert“ (Häcker, 2006: 36). Übergeordnete Ziele der Portfolioarbeit sind neben dem inhaltlichen Kompetenzerwerb z.B. die Förderung selbstregulierten Lernens (Albert, 2008) und die Reflexion der eigenen Lern- und Entwicklungsprozesse (Garner, 2006). Sie wird daher auch in der Fort- und Weiterbildung genutzt (Bessot, 2006).

Portfoliokonzepte sind dabei sehr vielfältig und können z.B. chronologisch orientiert sein wie beim Portfolio über das gesamte Studium, aber auch bei Jahres- oder Kursportfolios. Sie können bestimmten Veranstaltungsformaten wie z.B. Projekten angegliedert sein, einem bestimmten Zweck folgen (Bewerbungs- oder Entwicklungsportfolio) bzw. einer Qualifikationsorientierung, also z.B. der Kompetenzentwicklung dienen. Dabei sind auch Kombinationen verschiedener ePortfoliotypen (Vorzeige-, Arbeits-, Beurteilungs-, Bewerbungsportfolios etc.) möglich (vgl. Klampfer 2013: 17ff). Gleichzeitig kann das digitale Portfolio als virtueller Prüfungsort gesehen werden.

Besonders hilfreich stellt es sich als Lern- und Reflexionsort, also zur Bearbeitung von reflexionsfördernden Aufgaben und als erfahrungsbasiertes oder retrospektives Portfolio dar (vgl. z.B. Bräuer, 2014; Heid, 2011; Karpa, Kempf & Bosse, 2013). Des Weiteren ist das Portfolio insbesondere für die Bearbeitung pädagogischer Praxiserfahrungen, bei denen die Studierenden als Expert*innen ihrer selbst bzw. ihres inzwischen im Studium erfahrenen Wissens und Könnens gefragt sind, geeignet. So kann bspw. über die Portfolioarbeit auch eine transparente Abbildung von Lernprozessen im Rahmen der Praxisphase erfolgen – eine Funktion des Portfolios, die auch im hier beschriebenen Projekt genutzt wird.

Bereits ältere empirische Befunde (u.a. von Klenowski, 2000) zeigen umfassend positive Effekte von Portfolioarbeiten für die pädagogische und reflexive Praxis von Lehrer*innen im Vorbereitungsdienst. Konkret werden selbstevaluative Kompetenzen und die eigenständige Reflexion von beruflichen Erfahrungen gestärkt, sowie die Anwendung dieser Erkenntnisse

auf neue Unterrichtssituationen unterstützt. Zudem wird die Entwicklung von Verantwortung für den eigenen Lernprozess sowie eine Verbesserung des reflexiven Lernens im Sinne einer Veränderung der Wahrnehmung des eigenen Denkens gefördert. Auch Kompetenzen bezüglich der Unterrichtsplanung und Präsentationsfähigkeiten, dem Umgang mit Schüler*innenverhalten sowie die Kritikfähigkeit nehmen durch entsprechend ausgerichtete Portfolioarbeit zu. Lehrende und Referendar*innen profitieren von der Portfolio-Methode z.B. dahingehend, dass die Verwendung auf Kommunikation ausgelegter Methoden (Feedbackgespräche, Gruppenarbeiten etc.) unterstützt wird. Nötig für ein derart profitables Gelingen ist jedoch immer eine umfassende Einführung in die Portfolio-Methode sowie eine intensive, kontinuierliche Begleitung (Klenowski, 2000).

Für den deutschsprachigen Raum beschäftigten sich Feder & Cramer (2019) mit der Wirksamkeit von Portfolioarbeit im Allgemeinen im Kontext von Lehrer*innenbildung und erstellten ein Review unter Einbezug von 25 „forschend-empirischen" (S. 1230) Studien aus dem Veröffentlichungszeitraum 2005-2016. Hier zeigten sich insgesamt betrachtet bislang eher verhalten positive Effekte über Portfolioarbeit. Im englischsprachigen Raum hingegen belegen bereits umfangreiche Forschungen den Nutzen der Arbeit mit ePortfolios (vgl. z.B. Dune et al., 2018).

Muckel et al. führen in Bezug auf digitale Portfolios an, sie seien „nice to have, but they do cause inconvenience" (ebd.: 238). Sie legen dar, dass die Arbeit mit ePortfolios von Studierenden verständlicherweise eine größere Selbststeuerungs- und Selbstlernkompetenz und dementsprechend eine lerner*innen-orientiertere didaktische Perspektive erfordere. Das sei anstrengend, befördere jedoch den Lernparadigmenwechsel gemäß des sogenannten „Shift from Teaching to Learning" (z.B. Barr & Tagg, 1995), der im Zuge der Digitalisierung zurzeit ohnehin stattfindet (Ehlers, 2011, 2020). Insofern lässt sich aus der bisher berichteten eher geringen Akzeptanz Studierender nicht schließen, dass das Tool „ePortfolio" nicht nützlich sei. Die Begründungen könnten stattdessen eine Ermutigung sein, solche Tools auch deshalb einzusetzen, um die Eigenverantwortung der Studierenden für ihren Lernprozess weiter zu stärken.

Ein ePortfolio weist dabei gegenüber der Papierform einige Vorteile auf und gilt als digitale Sammlung von Arbeiten „einer Person, die dadurch das Produkt (Lernergebnisse) und den Prozess (Lernpfad/Wachstum) ihrer Kompetenzentwicklung in einer bestimmten Zeitspanne und für bestimmte Zwecke dokumentieren und veranschaulichen möchte. Die betreffende Person hat die Auswahl (. . .) selbstständig getroffen, und diese in Bezug auf das Lernziel selbst organisiert. Sie (Er) hat als Eigentümer(in) die komplette Kontrolle darüber, wer, wann und wie viel Information aus dem Portfolio einsehen darf" (Hornung-Prähauser et al. 2007: 14).

Die Nutzer*innen und Dozent*innen haben bei der ePortfolionutzung durch die vielfältigen zur Verfügung stehenden, einzufügenden digitalen Portfoliobestandteile und diverse Kommunikations- und Vernetzungsoptionen einen deutlich größeren Interaktions- und Spielraum. So ist eine Nutzung des ePortfolios als kreativer und gut strukturierbarer Doku-

mentationsort, also zur Sammlung von verfassten Texten, audio- und visuellen Arbeitsproben und diversen Aufgabenformaten zu einem fachwissenschaftlichen oder fachdidaktischen Thema aber auch zur reflexionsfördernden Nachbereitung von Lehrveranstaltungen möglich.

Das digitale Portfolio weist neben der Einbeziehung interaktiver Kommunikations- und Vernetzungsoptionen im Vergleich zu den papierbasierten Portfolios also eine Vielzahl an Vorteilen auf wie z.B.:

1. die Integration von vielfältigen digitalen Portfoliobausteinen (Bilder, Videos etc.),
2. die flexible Zugänglichkeit,
3. den flexiblen Austausch und Ergänzungsmöglichkeiten,
4. die Förderung von Medienkompetenz,
5. die elektronische Speicherung und Verwaltung von Daten,
6. die Erleichterung der Verteilung von Arbeitsmaterialien und Präsentation (vgl. Völschow, Israel, Warrelmann, 2019).

Dabei gehört das erstellte Portfolio den Studierenden wie bereits angedeutet selbst und jede*r Erstellende entscheidet, wann welche Teile des ePortfolios welchen Dritten über das Erteilen von Leserechten einsehbar gemacht werden. Somit können hier auch hinführende Vorarbeiten aufgenommen werden, die den Lehrenden noch nicht zur Ansicht freigegeben werden sollen. Diese Überlegungen wurden in der Konzeptentwicklung zum digitalen Portfolio aufgegriffen und weitergedacht.

3 Das eKEP: Entwicklung und Implementation

Entscheidung für ein ePortfolio-System

Eine Motivation für die eKEP-Entwicklung war es, Angebote zu entwickeln, die Lehramtsstudierende in ihrer pädagogischen und reflexiven Praxis unterstützen. U.a. auf Basis der oben vorgestellten Befunde (Klenowski, 2000; Hornung-Prähauser, 2007) und unter Berücksichtigung von damals bereits ausgemachten Digitalisierungserfordernissen fiel die Entscheidung für die Entwicklung eines ePortfolio-Tools. Die nächste Entscheidung stand dann im Zusammenhang mit der Auswahl einer hierfür nötigen ePortfolio-Software. Zunächst wurde dazu ein Überblick ggf. infrage kommender ePortfolio-Systeme erstellt. Deutlich wurde, dass eine Vielzahl von Systemen existiert (z.B. Drupal, Exabis, Mahara), die jeweils über gewisse Vor- und Nachteile verfügen. Diese beziehen sich z.B. auf die Funktionen des Sammelns von Artefakten, des Reflektierens, des Publizierens, in der Verwaltung und der sogenannten Usability (vgl. Klampfer 2013:38 ff.). Publikationen zur Evaluation von ePortfolio-Software gaben weitere Hinweise zu relevanten, zu beachtenden

Merkmalen. So nennen Himpsl & Baumgartner (2009) die folgenden fünf zu vergleichenden Kategorien:

- Usability;
- Administration;
- Representing Publishing;
- Reflecting, Testing, Verifying, Planning;
- Collecting, Organizing, Selecting (vgl. ebd.).

Vorteile der Stud.IP-Verankerung

Da die Universität Vechta mit Stud.IP bereits über ein etabliertes Campus-Management-System verfügte, lag die Prüfung seiner Nutzbarkeit für den ePortfolio-Einsatz nahe, um nicht zuletzt über die Vertrautheit mit der Oberfläche die Nutzungsschwelle gering zu gestalten. Damit fiel die Entscheidung auf die Eigenentwicklung eines ePortfolio-Plugins für Stud.IP.

Die Vorteile der Nutzung des bereits implementierten Campusmanagementsystems betreffen u.a. die Administration: Alle potentiellen Nutzer*innen sind bereits registriert und es müssen daher keine zusätzlichen Logins vergeben werden. Auch die benötigten Daten zu den Lehrveranstaltungen sind bereits vorhanden. Außerdem verfügt die Universität Vechta über einen eigenen Administrator, der sich um alle Stud.IP-Belange der Universität kümmert.

Mit Blick auf die Kategorie Usability müssen sich Studierende und Lehrende kein weiteres Login und auch keine neue URL merken, da sie über Stud.IP einen Zugang nutzen, mit dem sie ohnehin täglich arbeiten. Da bei der Entwicklung des ePortfolio-Tools auf viele bereits in Stud.IP vorhandene Funktionen zurückgegriffen werden kann, müssen sich die Studierenden nicht in ein weiteres technisches Tool bzw. eine weitere technische Oberfläche einarbeiten.

Bezogen auf die technische Entwicklung reduzierte sich der Aufwand für neue Programmierungen durch die Möglichkeit, auf bereits existierende Tools und Funktionalitäten (wie z.B. die Courseware) zurückzugreifen. Die Software steht bereits auf einer Basis mit umfangreichem bundesweitem Netzwerk, Support und einer breit aufgestellten Entwickler*innenkompetenz. So gibt es zwar Entwicklungskosten für die erste Erstellung; wenn sich weitere Standorte für das Tool interessieren, existiert jedoch eine größere Nutzer*innengemeinschaft, die sich die einmaligen Entwicklungskosten teilen kann. Wird das Plugin in den Kern des CMS implementiert, wird es routinemäßig bei entsprechenden neuen CMS Releases zentral aktualisiert. Hiermit entfallen gegenüber einer eingekauften Fremdsoftware fortlaufende Folgekosten für Zugänge, Aktualisierungen und Support.

Außerdem ist mit der Nutzung in einem bestehenden CMS wie Stud.IP das Anforderungsmerkmal der Datensicherheit bzw. des Datenschutzes synchron zum bestehenden System und damit leichter über einen Schritt zu prüfen und umzusetzen (vgl. Kammerl, 2011: 149).

eKEP Plugin-Entwicklung (2016 – 2020)

Die Entwicklung des Plugins wurde durch ein Team des Zentrums für Digitale Lehre, Campus-Management und Hochschuldidaktik (virtUOS) der kooperierenden Universität Osnabrück übernommen und durch einen interdisziplinär aufgestellten Beirat beratend begleitet. In einer ersten Version konnte das eKEP im Wintersemester 2018/19 eingesetzt werden. Die Arbeit mit dem ePortfolio ersetzte in verschiedenen Seminaren sukzessive die bisherige Portfolioarbeit in Papierform.

Zu den wichtigsten Funktionen des Tools gehören:

- die Erstellung von Arbeits- und Aufgabenvorlagen,
- die Verteilung der Vorlagen an die Studierenden,
- die Erstellung von multimedialen und interaktiven Inhalten sowohl in der Vorlage (Lehrende) als auch in den ePortfolios der Studierenden,
- die Erteilung von Zugriffsberechtigungen durch die Studierenden,
- die Erstellung von Feedback- und weiteren Kommunikationselementen.

Der Funktionsumfang nahm im Projektverlauf sukzessive zu. Ab dem Sommersemester 2020 konnte das eKEP in nahezu vollständiger Funktionalität genutzt werden.

Implementation an der Universität Vechta

Erste Studierendenbefragungen zeigten, dass das eKEP gut angenommen und in der Bearbeitung der damals inhaltlich im Zentrum stehenden, fortlaufenden Reflexionsaufgaben ein Gewinn gesehen wurde. Zum Sommersemester 2020 nutzten dann aufgrund der Corona-Sondersituation und der hiermit im Zusammenhang stehenden Anforderung, das Semester komplett online umsetzen zu müssen, sehr viel mehr Lehrende eigeninitiativ das eKEP, als vom Projekt aus zu diesem Zeitpunkt intendiert. Daher liegen zum jetzigen Zeitpunkt schon einige erste vorzeitige Evaluationsergebnisse vor, die im folgenden Kapitel vorgestellt werden.

4 Ergebnisse der Pilotierung

Im Folgenden werden die Ergebnisse der nutzungsbegleitenden Befragungen vorgestellt. Die Befragungen fanden seit dem Wintersemester 2019/2020 statt, da das eKEP zu diesem Zeitpunkt technisch bereits relativ weit entwickelt war.

4.1 Erfahrungen Studierender

Zwischen 2016 und 2019 wurden Studierende aus drei Kohorten zu mehreren Zeitpunkten zu ihren Erfahrungen mit der neu implementierten Portfolioarbeit befragt, die, wie oben erwähnt, in den ersten beiden Kohorten in Papierform und sukzessive als ePortfolioarbeit erfolgte. Die Ergebnisse wurden formativ für die Weiterentwicklung des Moduls genutzt. Die Befragungen wurden mittels teilstandardisierter Fragebögen und ergänzender Interviews durchgeführt.

Die dritte der genannten Kohorten arbeitete als erste Gruppe komplett online mit dem fertiggestellten ePortfolio. Aufgrund des Pilotprojekt-Charakters handelt es sich hier noch um eine relativ kleine Stichprobe von neun Teilnehmer*innen (TN), so dass diese Ergebnisse nur als erste Tendenz verstanden werden konnten: Vorweg genommen sei, dass am Ende des Moduls alle TN die Sinnhaftigkeit der ePortfolioarbeit mit Blick auf den Gewinn für die Reflexion ihrer Berufswahl betonten. Sechs TN stimmten der selbstklärenden und reflexionsfördernden Funktion (u. a. hinsichtlich der Berufswahl) der ePortfolioarbeit vollständig zu und drei teilweise. Aus den vertiefenden Interviews mit Seminarteilnehmer*innen wurde deutlich, dass die Aussage letzterer drei TN vor allem in technischen Problemen bei der ePortfolioerstellung begründet lagen und eine vollumfängliche Zustimmung daher ausblieb. Kritisiert wurde vor allem ein umständlicher Dateienupload (z.B. Interview mit Teilnehmer 1, 17.07.2018). Diese Aussage bestätigt sich im Kontrast zu den Rückmeldungen in einem früheren, nicht mit einem digitalen sondern papierbasierten Portfolio arbeitenden Durchgang des Seminars, bei dem alle Teilnehmer*innen der Aussage zur reflexionsfördernden Funktion vollständig zustimmten. Ebenfalls als gewinnbringend mit Blick auf eine tiefergehende Reflexion wurde die ergänzende Unter-stützung durch besonders qualifizierte Peer-Coaches erlebt (N=7). Alle TN (N=9) äußerten zudem, das ePortfoliomodul anderen Studierenden bedenkenlos weiterempfehlen zu können.

Ergänzend zur Fragebogenerhebung wurden vertiefende Einzelinterviews mit drei Studierenden durchgeführt, die diese Tendenzen bestätigten. Hier wurde der Mehrwert für die Lehramtsausbildung sowohl bezüglich der eigenen Professionalitätsentwicklung als auch des Studiums betont und die Empfehlung geäußert, das Modul verpflichtend zu verankern. Die einzigen Kritikpunkte bezogen sich auch hier auf zum Befragungszeitpunkt vorhandene technische Mängel des ePortfolio-Plugins, das in der Pilotphase eben noch nicht komplett entwickelt bzw. technisch ausgereift war.

Bei der ausschließlich digitalen Nutzung des eKEP im Wintersemester 2019/20 zeigte sich durch die fortlaufenden Gespräche sowie über die Auswertung der semesterbegleitenden E-Mail-Kommunikation mit den Studierenden, dass sich die Nutzer*innen in drei Gruppen aufteilen ließen: Die Studierenden der ersten Gruppe waren begeistert und nutzten das Tool ganz selbstverständlich. Sie probierten aus und testeten selbständig verschiedene Funktionen.

Die zweite Gruppe äußerte sich der eKEP Arbeit gegenüber eher neutral. Diese Studierenden nutzten das Tool eben, weil die Nutzung vorgegeben war.

Die Studierenden der dritten Gruppe hingegen zeigten sich kritisch und hinterfragten den Nutzen mit Blick auf die Zuverlässigkeit des Tools. Sie verhielten sich besorgt und schickten oft parallel ihre erarbeiteten Ergebnisse zusätzlich per E-Mail an die Dozierenden, um sicherzugehen, dass ihre Leistungen auch wirklich bei den Lehrenden ankamen.

4.2 Erfahrungen Lehrender

Zu den Erfahrungen Lehrender mit dem eKEP liegen erste Rückmeldungen aus dem aktuell allerdings noch nicht abgeschlossenen Sommersemester 2020 vor.

Vorweggenommen sei, dass die drei bislang befragten Lehrenden so überzeugt von dem Tool sind, dass sie ihren Kolleg*innen die Nutzung des eKEPs bedingungslos anraten würden. In informellen Gesprächen sowie drei Einzelinterviews zeigte sich, dass die Lehrenden durchweg von positiven Erfahrungen berichten und die Nutzung des eKEPs gerne in den folgenden Semestern fortsetzen möchten. Analog zu den Rückmeldungen der Studierenden waren die einzigen negativen Einschätzungen technischen Mängeln zuzuschreiben, die jedoch den Nutzen des Tools an sich nicht infrage stellten.

Mehrere Lehrende aus dem Bereich der Mathematikdidaktik nutzten das eKEP im Sommersemester 2020 erstmalig in einem gemeinsam verantworteten Modul, das bisher mit Papierportfolios durchgeführt wurde. Eine Lehrende berichtet, dass die e-Variante den großen Vorteil besäße, dass die Studierenden nun unkompliziert während des Semesters zu jedem beliebigen Zeitpunkt Aufgaben und Reflexionen einreichen und um Feedback bitten könnten und auf diese Weise auch ein Kompetenzerwerbsverlauf abbildbar werde. Eigentlich seien feste Termine für die Einreichungen vorgesehen gewesen, aber es habe sich so entwickelt, dass Studierende eigeninitiativ zusätzlich über die Freischaltung im Plugin um Feedback gebeten hätten. Der kommunikative, interaktive Charakter des eKEP kann daher als digitale Brücke in der durch die COVID-19 pandemiebedingten Einschränkungen der face-to-face Kommunikation gesehen werden. Er dürfte aber mit Blick auf die niedrigschwelligen, zeit- und ortsunabhängigen Nutzungsoptionen auch unabhängig davon als Kommunikationsbrücke bzw. -verstärker in Frage kommen.

Zudem wurde von Lehrenden auch der Aspekt der Förderung der Selbständigkeit und Eigenverantwortlichkeit Studierender erwähnt: Didaktisch wird die ePortfolioarbeit vor allem deswegen als empfehlenswert eingeschätzt, da die Studierenden sich Kompetenzen nicht nur passiv aneignen, indem sie einem vorgegebenen Pfad folgen, sondern aktiv den Vorgang der Kompetenzaneignung mitbestimmen können. Sie entschieden selbstständig, wann sie welche Aufgaben bearbeiten, an welchen Stellen evt. auch nachträglich im Kompetenzerwerbsprozess noch Inhalte hinzufügt werden oder weitere Erläuterungen notwendig sind. Auf der anderen Seite, gab ein*e Lehrende*r zu bedenken, dass das

ePortfolio für Lehrende und Studierende erst einmal u.a. auch einarbeitungsintensiver sei, da die Nutzenden mehr Bearbeitungsmöglichkeiten hätten. Die Ermöglichung der Einbringung mehr eigener Kreativität spräche trotzdem für eine Nutzung (Lehrende*r 1).

Eine weitere Lehrperson berichtet über ihre Beobachtung bzgl. des Zuwachses studentischer Ideenreichtums. Über das ePortfolio bringen sich die Studierenden demnach stärker und kreativer selbst ein als zuvor gewohnt (Lehrende*r 2).

Ein*e dritte*r Lehrende*r bewertet es sehr positiv, dass das ePortfolio Lehrenden einen guten Einblick über studentische Aktivitätsverläufe gibt. Die kontinuierliche Information, wer zu welchem Zeitpunkt welche Inhalte eingestellt hat, wird – verglichen mit einer punktuell abgegebenen Leistung am Semesterende – wurde als hilfreiche Kontrollunterstützung beschrieben. Zudem liefere das ePortfolio einen effizienten Überblick sowohl über die Pünktlichkeit bei der Bearbeitung der Aufgaben als auch über das Aktivitätsniveau und den Umfang. So wird z.B. zwischendurch schnell deutlich, ob Studierende eher wenige Sätze einreichen oder gar einige Seiten. Besonders hervorzuheben ist die rasche Rückmeldemöglichkeit durch die Feedbackfunktion des Tools. Das ePortfolio wird von den bisher befragten Lehrenden als insgesamt leistungsstarkes Tool beschrieben.

Bezüglich der Usability bzw. Nutzerfreundlichkeit wurde von den befragten Lehrenden erwartungsgemäß berichtet, dass es den Studierenden leicht fiel, mit dem eKEP zu arbeiten, da es bereits aus Stud.IP bekannte Courseware-Komponenten enthält. Nützlich sei auch die im ePortfolio enthaltene Semesterübersicht über alle Aufgaben, die zu erledigen sind. Wenn zu Beginn alle Aufgaben eingestellt werden, sowohl die sofort bearbeitbaren als auch jene, die ggf. noch nicht freigeschaltet sind und erst später bearbeitet werden sollen, könnte es aber auch passieren, dass sich einige Studierende durch den anfänglichen Umfang der gesamten Aufgaben unter Druck gesetzt fühlten, wenn Dozent*innen diese gleich alle freischalten. Andere Studierende jedoch hätten die Transparenz der Aufgabenübersicht geschätzt und zuweilen auch entsprechend vorgearbeitet. Die Vermutung des Lehrenden, der diesen Effekt beobachtet hat, ist daher, dass die Studierenden durch das aufgabentransparente ePortfolio nicht mehr – wie häufig zuvor in den Lehrveranstaltungen erlebt – die gesamte Prüfungsleistung wenige Tage vor dem vorgegebenen Abgabestichtag erarbeiten, sondern frühzeitig mit der sukzessiven Bearbeitung beginnen und dass somit die Ergebnisse am Ende besser sein werden (Lehrende*r 3).

5 Zusammenfassung und Ausblick

Das ePortfolio-Plugin für Stud.IP wird an der Modelluniversität Vechta seit 2016 entwickelt und sukzessive implementiert. Bereits im Wintersemester 2018/2019 konnte das eKEP, das damals für ein neu geschaffenes Modul zur Reflexionsförderung konzipiert wurde, eingesetzt werden und wird seitdem kontinuierlich auch für den Einsatz in anderen Themenfeldern der Lehramtsausbildung weiterentwickelt und evaluiert. Da COVID-19-pandemiebedingt das gesamte Sommersemester 2020 digital zu gestalten war, kam es zu einer verstärkten

vorgezogenen Nutzung des Plugins an der Universität Vechta: Lehrende aus zu diesem Zeitpunkt noch nicht für die Implementation des eKEP anvisierter Fächer stiegen situationsbedingt bereits in die eKEP-Nutzung ein, auch wenn ursprünglich nur eine Vortestung in ausgewählten Pilotmodulen angedacht war.

Die Befragungen zu den ersten Durchführungen des Pilotsystems zeigen durchweg positive Erfahrungen mit der ePortfolioarbeit. Das trifft sowohl für die Seite der Studierenden als auch auf die der Lehrenden zu und ermutigt, den eingeschlagenen Weg weiterzuverfolgen. Aus den Erkenntnissen verschiedener Vorgängerstudien des deutschen und englischsprachigen Raumes (s. Kapitel 2.2) lassen sich noch vielfältige Forschungsdesiderata ableiten. Das betrifft z.B. konkret die Untersuchung von ePortfolioarbeit in ihrer Wirkung auf Reflexionsförderung, für die es zumindest im deutschsprachigen Raum wenig belastbare Erkenntnisse gibt.

Besonders wichtig ist aber – insbesondere auch im Zuge der ungeplanten intensiveren vorzeitigen Nutzung des Plugins – die Erkenntnis, dass die technische Wartung des Systems von hoher Bedeutsamkeit ist: Die Nutzer*innen-Befragungen v.a. von Studierenden, aber auch die Befragungen von Lehrenden, verdeutlichen, dass bereits kleinste technische Defekte zu einer enormen Unsicherheit führen, die dann die gesamte digitale Portfolio-Arbeit in Frage stellen lässt.

Aufbauend auf den bisherigen Modellprojekterfahrungen werden das Konzept und die Technik für das Stud.IP ePortfolio-Plugin kontinuierlich weiterentwickelt. Dabei werden aktuell bereits im Austausch mit anderen Universitäten auch neue Bedarfe mitgedacht, wie z.B. die gemeinsame Arbeit mehrerer Studierender an einem Gruppenportfolio. Des Weiteren wird – noch prüfungsrechtlich abzusichern – angestrebt, mit dem ePortfolio auch eine ausschließlich digitale Prüfungsleistung bereitzustellen. Aktuell müssen aufgrund des modellhochschulischen Prüfungsrechts alle Portfolio-Leistungen auch noch einmal in Papierform abgegeben werden. Das führt zu Herausforderungen, wenn Video- oder Audiodateien im ePortfolio enthalten sind.

Insgesamt wird das Stud.IP ePortfolio-Plugin von den Lehrenden, die es nutzen, wie auch von den meisten Studierenden als leistungsstarkes Tool angesehen, dessen Ausbau und Weiterverbreitung vielversprechende Vorteile bietet.

Literaturverzeichnis

[Al08] Albert, U.: Portfolio im Kontext von Evaluation. In (Brüsemeister, T. & Eubel, K.-D., Hrsg.): Evaluation, Wissen und Nichtwissen. VS Verlag, Wiesbaden, S. 275-294, 2008.

[BT95] Barr, R. B. & Tagg, J.: From Teaching to Learning — A New Paradigm For Undergraduate Education. Change: The Magazine of Higher Learning, 27:6, S. 12-26, DOI: 10.1080/00091383.1995.10544672, 1995.

[Be06] Bessot, R.: Portfolios und Dossiers von Lehrpersonen. Pädagogische Einführung, 4, S. 229-234, 2006.

[Br14] Bräuer, G.: Das Portfolio als Reflexionsmedium für Lehrende und Studierende. 2. Aufl. UTB: Regensburg, 2014.

[Du18] Dune, T., Crnek-Georgeson, K., Bidewell, J., Firdaus, R., John, J., Arora, A.: Undergraduate health science students' development of reflective practice on communication skills via e-Portfolios, Journal of University Teaching & Learning Practice, 15(3), 2018, https://ro.uow.edu.au/jutlp/vol15/iss3/5, Stand: 13.02.2020.

[Eh11] Ehlers, U.-D.: Qualität für digitale Lernwelten: Von der Kontrolle zur Partizipation und Reflexion. In (Hugger, U., Walber, M., Hrsg.): Digitale Lernwelten. Konzepte, Beispiele und Perspektiven. VS Springer, Wiesbaden, S. 59-74, 2011.

[Eh20] Ehlers, U.-D.: Be Prepared for the shift: From Teaching to Learning. 2020. https://ulf-ehlers.net/2020/04/03/beprepared-for-the-shift-from-teaching-to-learning, Stand: 05.09.2020.

[FC19] Feder, L., Cramer, C.: Portfolioarbeit in der Lehrerbildung. Ein systematischer Forschungsüberblick. Zeitschrift für Erziehungswissenschaften 22, 1225-1245, 2019.

[Ga06] Garner, B.: Portfolios: Portraits guten Unterrichtens. Das Lehrportfolio als Instrument professioneller Entwicklung. In (Brunner, I., Häcker, T. & Winter, F., Hrsg.): Das Handbuch Portfolioarbeit. Konzepte, Anregungen und Erfahrungen aus Schule und Lehrerbildung Kallmeyer, Seelze-Velber, S. 249-254, 2006.

[Gu20] Gummels, I.: Wie kooperatives Lernen im inklusiven Unterricht gelingt – Entwicklung und Evaluation einer Lernumgebung für den Mathematikunterricht. Springer, 2020.

[Hä05] Häcker, T.: Portfolio als Instrument der Kompetenzdarstellung und reflexiven Lernprozesssteuerung, Berufs- und Wirtschaftspädagogik, 8, 2005.

[Hä06] Häcker, T.: Vielfalt der Portfoliobegriffe. Annäherung an ein schwer fassbares Konzept. In (Brunner, I., Häcker, T. & Winter, F., Hrsg.): Das Handbuch Portfolioarbeit. Konzepte, Anregungen und Erfahrungen aus Schule und Lehrerbildung Kallmeyer, Seelze-Velber, S. 33-39, 2006.

[He11] Heid, M.: Arbeit am pädagogischen Selbst – das Portfolio-Konzept in der Lehrerinnen- und Lehrerbildung. BIOS, 1/11, S. 98-118, 2011.

[He20] Herkenhoff, J.: Inklusiver Mathematikunterricht - Entwicklung eines Instruments zur Planung von Mathematikunterricht in einem inklusiven Setting. Springer, 2020.

[HB09] Himpsl, K. & Baumgartner, P.: Evaluation of E-Portfolio Software. International Journal of Emerging Technologies in Learning (iJET), S. 16-22, 2009, https://www.learntechlib.org/p/45093, Stand: 03.09.2020.

[Ho16] Hofmann, F., Wolf, N., Klaß, S., Grassmé, I. & Gläser-Zikuda, M.: Portfolios in der LehrerInnenbildung. Ein aktueller Überblick zur empirischen Befundlage. In (Boos, M., Krämer, A. & Kricke, M., Hrsg.): Portfolioarbeit phasenübergreifend gestalten. Konzepte, Ideen und Anregungen aus der LehrerInnenbildung. Waxmann, Münster, S. 23-39, 2016.

[Ho07] Hornung-Prähauser, V., Geser, G., Hilzensauer, W. & Schaffert, S.: Didaktische, organisatorische und technologische Grundlagen von E-Portfolios und Analyse internationaler Beispiele und Erfahrungen mit E- Portfolio-Implementierungen an Hochschulen, Research Forschungsgesellschaft, Salzburg, 2007.

[Ka11] Kammerl, R.: Integrierte E-Portfoliofunktionen in Stud.IP -das Projekt »InteLeC - Integrierter eLearning Campus«. In (Meyer, T., Mayrberger, K., Münte-Goussar, S., Schwalbe, C., Hrsg.): Kontrolle und Selbstkontrolle. Zur Ambivalenz von E-Portfolios in Bildungsprozessen. Verlag für Sozialwissenschaften, Wiesbaden, S. 145-150, 2011.

[Ka13] Karpa, D., Kempf, J. & Bosse, D.: Das E-Portfolio in der Lehrerbildung aus Perspektive von Studierenden. Digitale Medien und Schule, 7/13, S. 1-14, 2013.

[Kl13] Klampfer, A.: E-Portfolios als Instrument zur Professionalisierung in der Lehrer- und Lehrerinnenausbildung. Bewertung technologischer und motivationaler Faktoren der Nutzung durch Studierende. vwh-Verlag, Glücksstadt, 2013

[Kl00] Klenowski, V.: Portfolios. Promoting Teaching. Assessment in Education: Principles, Policy & Practice, S. 215-236, 2000.

[MHB12] Muckel, P., Heidkamp, B., Brunner, S.: Learning Scenarios with Integrated ePortfolios. EPortfolios are nice to have but do cause inconvenience. . . In (S. Ravet et al., Hrsg.): ePIC - ePortfolio and Identity Conference-Proceedings 2012. S. 238 – 244, 2012.

[Vö19] Völschow, Y., Israel, S. & Warrelmann, J.: Das elektronische Kompetenzentwicklungsportfolio. Ein Instrument zur Reflexionsförderung in Lehramt. In (Safi, N., Bauer, C. & Kocher, M., Hrsg.): Lehrberuf: Vorbereitung, Berufseinstieg, Perspektiven. Beiträge aus der Professionsforschung. Hep, Bern, S. 61-70, 2019.

6. Workshop zum Stand, den Herausforderungen und Impulsen des Geschäftsprozessmanagements

Stand, Herausforderungen und Impulse des Geschäftsprozessmanagements

Michael Fellmann[1] Ralf Laue[2] Birger Lantow[3] Jana-Rebecca Rehse[4]

Der Workshop zum Stand, den Herausforderungen und Impulsen des Geschäftsprozessmanagements findet 2020 zum sechsten Male im Rahmen der GI-Jahrestagung INFORMATIK statt. Wie schon in den Vorjahren, möchte der Workshop auch in diesem Jahr wieder eine Diskussionsplattform sein und dazu beitragen, neue Impulse und Anregungen für die zukünftige Forschung zu liefern.

Zum Workshop eingereicht werden konnten Beiträge, die entweder neue Erkenntnisse beinhalten, auf bereits veröffentlichten Arbeiten basieren oder ein interaktives Format jenseits der klassischen Publikation mit Vortrag vorsehen. Als Ergebnis des Begutachtungsprozesses wurden schließlich neun Beiträge zur Veröffentlichung im Tagungsband akzeptiert.

Wie in den Vorjahren auch, umfassen die Beiträge wieder ein breites Themenspektrum, das von modellierungsnahen und technischen Themen bis hin zu Themen der Organisation, des Umgangs mit der Modellierung und auch der Bewertung von Modellen reicht. Thomas Bauer und Ralf Laue untersuchen die Repräsentation der organisatorischen Perspektive bei der Modellierung von Geschäftsprozessen. Die Autoren gehen dabei der Frage nach, warum und in welchen Bereichen die Modellierung hierbei immer noch Defizite aufweist. Die Analyse beinhaltet den Vergleich von vier kommerziellen Prozessmanagementsystemen und führt schließlich zur Ableitung von Forschungsfragen und Herausforderungen.

Ebenfalls mit einem bisher noch wenig beachteten Aspekt befassen sich Thorsten Schoormann und Kristin Kutzner in ihrem Beitrag zu sozialen Geschäftsprozessmustern. Die Autoren konstatieren dabei zunächst ein Defizit hinsichtlich der Berücksichtigung der sozialen Nachhaltigkeit bei Prozessverbesserungen etwa in Bezug auf integrative Arbeitsumgebungen, Gesundheit der Mitarbeiter, Beteiligung oder faire Wertschöpfungsketten. Anhand von bereits existierenden sozialen Geschäftsprozessmustern werden Nachhaltigkeitsberichte verschiedener Organisationen mit der Mustersammlung analysiert, um diese schließlich zu verfeinern und zu erweitern. Die Autoren zeigen somit auf, wie Prozesse im Hinblick auf soziale Belange untersucht werden können und die ihre Eigenschaften in sozialer Hinsicht verbessert werden können.

[1] Universität Rostock, Institut für Informatik, michael.fellmann@uni-rostock.de
[2] Westsächsische Hochschule Zwickau, Fachgruppe Informatik, ralf.laue@fh-zwickau.de
[3] Universität Rostock, Institut für Informatik, birger.lantow@uni-rostock.de
[4] Universität Mannheim, Fakultät für Betriebswirtschaftslehre, rehse@uni-mannheim.de

doi:10.18420/inf2020_52

Auch die kompetente Modellierung und deren Vermittlung in der Lehre ist Gegenstand aktueller Forschung und wird von Michael Striewe, Constantin Houy, Jana-Rebecca Rehse, Meike Ullrich, Peter Fettke, Niclas Schaper und Andreas Oberweis untersucht. Im vorliegenden Beitrag zu Möglichkeiten einer automatisierten Bewertung der Modellierungskompetenzen untersuchen die Autoren schwerpunktmäßig, wie die bisher zahlreichen Konzepte und Werkzeuge, die das Lernen und Lehren der grafischen Modellierung unterstützen, integriert werden könnten. Hierzu werden sowohl didaktische als auch technische Herausforderungen bei der Integration von Lern- und Lehransätzen für die grafische Modellierung erörtert und ein Lösungsansatz skizziert, der derzeit in einem laufenden Forschungsprojekt verfolgt wird.

Über die Modellierung hinaus werden auch technische und ausführungsbezogene Aspekte von den Autoren der Workshop-Beiträge aufgegriffen. Mit Herausforderungen im Bereich der Analyse und Optimierung von Fertigungsprozessen mit Process-Mining befassen sich dabei Simon Dreher, Peter Reimann und Christoph Gröger. In ihrem Beitrag beleuchten die Autoren die gegenwärtige Anwendbarkeit von Process-Mining in klassischen Prozessen der Produktion. Anhand eines systematischen Literaturüberblicks identifizieren die Autoren domänenspezifische Herausforderungen und damit verbundene Forschungslücken bezüglich unstrukturierter, kaskadierter und nichtlinearer Prozesse oder heterogener Datenquellen. Auch werden einige Anregungen erarbeitet, welche fertigungsbezogenen Prozesse durch Process-Mining-Techniken analysiert werden können.

Die Spezifikation, Ausführung und das Monitoring von Workflows in verteilten Wissensgraphen wird von Tobias Käfer und Andreas Harth untersucht. Die Autoren untersuchen, wie webbasierte Datenrepräsentationsformate wie das Resource Description Framework (RDF) und REST-basierte Protokolle verwendet werden können, um Workflows zu beschreiben und auszuführen. Im Kern stellen die Autoren hierzu eine Ontologie zur Beschreibung ausführbarer Workflows vor. Zur Ausführung wird eine operationale Semantik angegeben, die auf einem regelbasierten Rechnermodell für Read-Write-Linked-Data basiert. Die Anwendbarkeit des Ansatzes wird über ein prototypisches System demonstriert und die lineare Skalierfähigkeit anhand eines Benchmarks festgestellt.

Dem wohl zunehmend bedeutsamen „Faktor Mensch“ im Geschäftsprozessmanagement wie auch managementbezogenen Fragen widmen sich drei weitere Beiträge. So untersuchen Aleksandra Dzepina und Franz Lehner in ihrer Studie den Umgang mit der Modellierung an sich. Dies geht von der Beobachtung aus, dass insbesondere kleine und mittelständische Unternehmen nach wie vor Schwierigkeiten bei der erfolgreichen Umsetzung des Geschäftsprozessmanagements haben. Vor diesem Hintergrund untersuchen die Autoren die heutigen Strukturen des Geschäftsprozessmanagements in deutschsprachigen Unternehmen und gewinnen durch Experteninterviews einen Einblick in die Praxis. Die Autoren folgern, dass sich nach wie vor keine Sprache eindeutig als Modellierungsstandard etablieren konnte, eine Beurteilung der Qualität der Prozessmodelle zweitrangig bleibt und bis heute eine standardisierte Lösung für eine personen- und anforderungsspezifische Navigation in der Praxis fehlt. Im Rahmen ihrer Studie werden fünf Gestaltungsfaktoren beschrieben, die für eine erfolgreiche Umsetzung des Geschäftsprozessmanagements von Bedeutung sein können.

Neben dem Umgang mit der Modellierung ist auch die Akzeptanz ihrer Resultate durch die relevanten Personen von erheblicher Bedeutung, um eine korrekte Ausführung der Prozesse zu erreichen. In diesem Zusammenhang untersuchen Lars Drewes und Volker Nissen anhand einer experimentellen Studie den Einfluss der Prozesslänge und -dauer auf die Akzeptanz eines Prozesses. Den theoretischen Unterbau hierzu liefert die Prozessakzeptanztheorie, die einen Einfluss der Akzeptanz auf die korrekte Ausführung postuliert. Die Studie beinhaltet einen generischen Einkaufsprozess, der online in verschiedenen Varianten ausgeführt wird. Die Prozessakzeptanz wird mit Hilfe eines Fragebogens gemessen. Die Analyse zeigt, dass es einen signifikanten Unterschied in der Akzeptanz von Prozessen in Abhängigkeit von deren Prozessdurchlaufzeit als Maß für die Prozesslänge gibt.

Dem agilen Management von Geschäftsprozessen schließlich widmen sich Janek Ziehmann und Birger Lantow. Die Autoren argumentieren hierbei, dass Agilität alleine in der IT-Entwicklung zu kurz greift. Vielmehr sollte auch das Management der Geschäftsprozesse agil werden, um insbesondere der durch die Digitalisierung induzierten Dynamik Rechnung zu tragen. Die Autoren beantworten anhand einer systematischen Literaturanalyse die Frage, welche Ansätze zur Umsetzung von Agilität im Geschäftsprozessmanagement bisher vorgeschlagen wurden. Abschließend wird der Bedarf für zukünftige Forschung und Entwicklung in diesem Bereich abgeleitet.

Abgerundet wird der diesjährige Workshop durch einen stark interaktiven Part, in der die Teilnehmer zur Mitarbeit eingeladen sind. Dieser „Workshop im Workshop“ korrespondiert mit dem Beitrag von Arno Mueller, Hinrich Schroeder und Lars von Thienen zur Prozessdigitalisierung in der Praxis. In der Session soll der Weg vom Kundenproblem hin zu konkreten Anforderungen an die technische Implementierung am praktischen Beispiel erlebbar gemacht werden. Hierzu wird ein exemplarischer Serviceprozess aus Kunden- und Unternehmenssicht im Workshop von den Teilnehmern analysiert und optimiert. Durch das interaktive Durchspielen sollen Erkenntnisse bezüglich der Frage gewonnen werden, wie nicht nur die IT-Implementierung agil gestaltet werden kann, sondern alle Phasen eines BPM-Projektes. Die Frage der Agilität im Prozessmanagement wird damit in Ergänzung zum diesbezüglichen Literaturüberblicksbeitrag noch ein weiteres Mal aufgegriffen und diskutiert. Hintergrund zu dieser Session bilden die Erfahrungen der Moderatoren im Kontext eines praxisorientieren Forschungsprojekts mit ca. 50 IT- und Prozessexperten aus 13 Unternehmen.

Ein Workshop wie ZuGPM wäre nicht möglich ohne die tatkräftige Unterstützung vieler Beteiligter. Ein großer Dank geht an die Mitglieder des Programm-Komitees, die sehr aktiv und konstruktiv die Phase der Begutachtung unterstützt haben.

Wir wünschen allen Teilnehmerinnen und Teilnehmern des in digitaler Telepräsenz stattfindenden Workshops viele neuen Erkenntnisse und interessante Kontakte – auf dass sich inspirierende und fruchtbare Gespräche auch nach dem Workshop fortsetzen mögen.

September 2020

Michael Fellmann, Ralf Laue, Birger Lantow, Jana-Rebecca Rehse

Socially Business Process Patterns – A Sustainability Report-driven Demonstration and Refinement

Thorsten Schoormann[1], Kristin Kutzner[2]

Abstract: Business process patterns provide a well-accepted tool for analyzing and improving processes. Thus, patterns have the potential to leverage the sustainable transformation of organizations by considering societal and environmental concerns in addition to economic obligations. Although various process patterns have been proposed, there is a deficit regarding the social dimension of sustainability, which hinders improving processes in terms of issues such as inclusive working environments, employee health, participation, or fair value chains. By drawing on an initial set of socially business process patterns (SBPPs) that has been developed in previous work, this study seeks to examine sustainability reports from several organizations to evaluate, refine, and extend the current set of SBPPs. We contribute to social sustainability and process patterns by enabling researchers and practitioners to explore processes in respect of social concerns as well as providing orientation (i.e., examples) on how to improve social performance.

Keywords: Business Process Patterns; Social Sustainability; Sustainable Development; Evaluation

1 Introduction

A key booster for transforming businesses into more sustainable entities is given by business process management (BPM), which is a widely-accepted approach to design and/or change organizational structures [Ro06] [SBK17]. As businesses commonly adapt the triple bottom line-principle [El97] to contribute to sustainable development, it would seem a prerequisite that tools and approaches for analyzing and (re-)designing processes consider environmental and social sustainability, in addition to its economic obligations [But11] [Lo11]. However, business success is traditionally linked to criteria such as time, costs, quality, or flexibility, and thus, current BPM approaches and (software) tools are primarily designed to support economic-oriented efficiency [OKK14] [SSK19], which is problematic as it hinders taking a holistic and multidimensional view on sustainability.

Besides treating the economic dimension with priority, there has been some development in the field of sustainable BPM, Green BPM. Even though Green BPM aims at incorporating sustainability, it lays its focus on ecological concerns. In doing this, for instance, modeling notations have been extended (e.g., to visualize CO_2-emissions [Re12]) and process patterns

[1] Universität Hildesheim, Informationssysteme und Unternehmensmodellierung, Universitätsplatz 1, 31141 Hildesheim, thorsten.schoormann@uni-hildesheim.de
[2] Universität Hildesheim, Informationssysteme und Unternehmensmodellierung, Universitätsplatz 1, 31141 Hildesheim, kristin.kutzner@uni-hildesheim.de

doi:10.18420/inf2020_53

haven been derived to achieve ecological goals [No14] [LFL17]. In consequence, although information systems (IS)—and BPM in particular—have the potential to promote social concerns such as through enabling participation, promoting digital inclusion, contributing to security, ensuring employee health, or creating equity in working environments, social aspects are often neglected in IS [CG18] [KSK18] [Ol13] [SW19] [SBW12].

To bridge this gap, the overall goal of this research project is to build and evaluate 'socially business process patterns' (SBPP) to increase the social performance in organizations because those patterns provide proven and applicable solutions in the form of knowledge about best practices [SW00] and help to identify improvement potential [Be10] [Fe18] [Wi09]. Therefore, as a first step, we started to deduce patterns by means of an extensive literature review and initial expert interviews to reveal typical problems and already applied solutions from practitioners [Sc19]. Based on the previous results, this study follows the suggestion that "patterns should be backed by evaluations" [Fe18, p. 19] and therefore raises two research questions (RQ):

- *RQ1: Which solutions from practice support the current set of SBPPs?*
- *RQ2: How can new patterns be derived based on the solutions from practice?*

To achieve these goals, we aim at reviewing and coding sustainability reports (e.g., cooperate social responsibility (CSR) reports) from several companies that describe how social sustainability is adapted and promoted in their organizations. Doing so, we seek to disclose and learn from practical solutions and initiatives on social sustainability, which helps to evaluate the current patterns (e.g., is a pattern supported by a practical solution?) as well as to derive new patterns (e.g., how to translate a practical solution into an additional process pattern?). Hence, our study's contribution is twofold: first, evaluating the current set of SBPPs and, second, extending the current set. Overall, we aim at emphasizing the importance of the social dimension of sustainability in IS, and thus, reflect on the essential principles regarding social sustainability such as equity, diversity, interconnectedness, quality of life, and democracy [Mc04], which affect both internal practices (e.g., safety and healthy working conditions, and decent work) as well as external practices (e.g., equity along the supply chain) from organizations. For practitioners, our contribution can be employed for analyzing and (re-)designing business processes in terms of social sustainability and it illustrates alternative solutions in the form of general best practices. For researchers, our results extend the current body of knowledge in the field of sustainability and business processes as well as open consecutive research.

The paper is structured as follows: in Section 2, related work on business process patterns in general and for sustainability is summarized, including our previous work (i.e., the current set of SBPPs). In Section 3, we describe the research design and research process for building and refining/demonstrating the patterns. In Section 4, we present the refined and extended set of SBPPs as well as illustrate the solution scheme by representing abstract solution schemes for each pattern. In Section 5, we discuss our results, implications for research and practice, and the limitations of this study, as well as derive avenues for future research. Finally, in Section 6, we conclude with our paper.

2 Related Work

As introduced by Alexander [Al77b], a pattern formulates a relation between a certain context, a problem, and a solution. Drawing on this idea, [Ga95] proposed the concept of software design patterns, which poses a starting point for exploring patterns in further (sub-)disciplines like computer science and IS. In BPM, a process pattern can be defined as a "description of a proven solution to a recurring problem that is related to the creation or modification of business process models in a specific context" [Fe18, p. 3]. Patterns play a pivotal role in BPM [Wi09] to effectively and efficiently analyze and (re-)design processes and have been developed for diverse purposes such as enterprise architecture, workflow data, parallel computing, negotiation, collaboration, risk, and security, or compliance (e.g., [Fe17] [Sc20] [SW00]).

In terms of sustainability, mostly ecological-driven patterns have been proposed including, for example, 'green business process patterns' (e.g., compensating negative effects, green process automation for specific tasks, or green variant for providing ecologically-friendly process paths) [No14], 'ecological weakness patterns' (e.g., to identify weaknesses in terms of printing behavior or redundant document storage) [LFL17], 'ecological workflow patterns' (e.g., reducing energy consumption) [LFL16] aiming at disclosing ecological shortcomings or/and guiding the redesign of business processes.

However, there is only limited research on how to use the pattern-approach to contribute to social sustainability in processes. Therefore, as a first step, an initial set of seven SBPPs have been derived in previous work (see also Fig. 2): *social compensation* (i.e., improving social performance by compensating negative effects without changing the structure of a process), *social alternative* (i.e., improving social performance by offering an additional alternative process path to the customers without changing the (core) business processes), *social resource replacement* (i.e., improving social performance by replacing resources that have negative influence on employees (internal) and society (external) without changing the structure of a process), *social labeling* (i.e., improving the social image of an organization by labeling processes, products, and services with corresponding social certificates), *social sourcing* (i.e., improving social performance by (a) transferring activities to external organizations that are able to carry out an activity in a socially acceptable manner or by (b) transferring external activities into the own company), *social value chain* (i.e., improving social performance by defining social standards that have to be applied by every partner within a value chain), and *social-/human-centered individualization* (i.e., improving social performance by adjusting processes in terms of human-specific needs in order to allow various people to execute a certain process).

3 Research Design

In pursuing to achieve our overall goal (i.e., build and evaluate process patterns for social sustainability), we performed a two-stage research design (see Fig. 1).

In **Stage 1**, we conducted a literature review to obtain a sample of existing patterns that might be adaptable to the social dimension of sustainability (search: *("Green Business Process*" OR "Sustainable Business Process*" OR "Social Business Process*") AND pattern*)*. Furthermore, we performed several expert interviews to gather domain-specific knowledge that can be used to identify typical problems and best practices that are already applied in organizations (three sustainability manager, one sustainability consultant). Based on this, an initial set of seven SBPPs could be derived.[3]

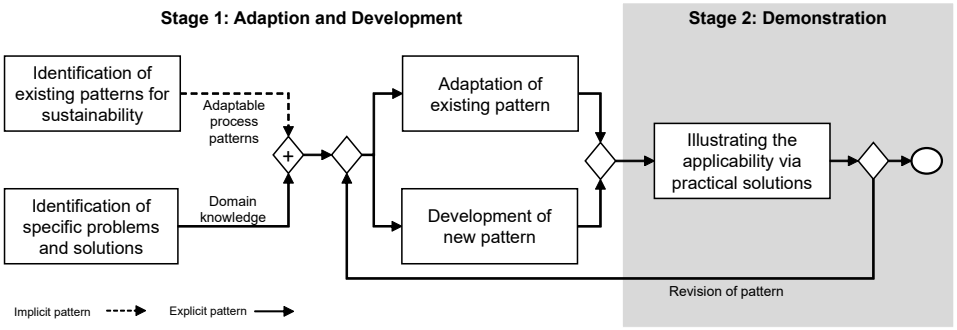


Fig. 1: Overall Research Design (Grey = Focus of this Study)

In **Stage 2**, we seek to demonstrate the applicability of the derived SBPPs by examining practical solutions and initiatives, here in the form of available sustainability reports[4] from different organizations. By reviewing such reports, the current set of SBPPs can be confirmed and complemented with examples as well as revised through (in case) additional solutions and initiatives. Therefore, we examined reports in diverse (and prominent, prevailing) industry sectors such as automobile, manufacturing, finance, commerce, etc. to ensure a representative set of units for the analysis—based on 'sales of the most important industrial sectors in Germany in the years from 2016 to 2018' [St20]. As a result, we obtained a total of 24 sustainability reports from larger companies across diverse various sectors (see Tab. 1 for more details).

Afterwards, one researcher coded the reports to extract text paragraphs that are relevant for this study´s purpose of social sustainability. In a next step, two researchers who were both involved in the development of the initial set of SPBBs analyzed the text snippets independently and assigned them to the existing SPBBs by highlighting if a text snippet (a) supports one of the existing patterns, (b) refines one of the existing patterns, or (c) provides insights that may lead to an additional pattern. Next, an in person-workshop was conducted in which the individual findings were collaboratively compared and discussed with both researchers. In this workshop, inter-coder agreements, especially regarding the refinement,

[3] See [Sc19] for more details of the initial set of patterns.

[4] To increase the transparency of the social and environmental performance the European Parliament and the member states of the EU adopted a so-called CSR directive to expand organization's reporting [CG17]. In Germany, a corresponding law has been established which says that environmental, social, and labor aspects as well as human rights, anti-corruption, and bribery need to be reported by larger organizations [Bu17].

could be disclosed (e.g., splitting the pattern for social compensation, differentiating between internal and external social alternatives, adding a sub-pattern for resource re-usage). Against this, some aspects had to be discussed in-depth, for example: 'should we rename the patterns for social compensation into more specific variants such as health for employees or social projects?'; 'how to handle concerns about inclusion and diversity?' To achieve consolidation, we decided to stick to the origin pattern titles and level of abstractions and subordinate most of the detailed concerns to one of the existing patterns (e.g., health as part of social compensation).

Tab. 1: Selection of Companies/Sustainability Reports

Industry	**Publisher of Report**	**Year of Report**	**Reference**
Automobile	BMW Group	2018	[BMW18]
	Volkswagen AG	2018	[VW18]
	Continental AG	2018	[Co18]
Manufacturing	Siemens AG	2018	[Si18]
	Robert Bosch GmbH	2018	[RB18]
Chemicals	BASF Group	2018	[BAS18]
	Bayer AG	2018	[Ba18]
Electronic	IBM Corporation	2018	[IBM18]
	Cisco Systems Inc.	2018	[Ci18]
	Canon Inc.	2018	[Ca18]
Nutrition (food)	Lidl Dienstl. GmbH & Co. KG	2016/2017	[Li16]
	Metro AG	2017/2018	[Me17]
Building	Hochtief AG	2018	[Ho18]
	Strabag SE	2018	[St18]
Finance	Deutsche Bank AG	2018	[DB18]
	Allianz Group	2018	[Al18]
Commerce	Otto (GmbH & Co KG)	2018/2019	[Ot18]
	Bertelsmann SE &Co. KGaA	2017	[Be17]
Energy	Uniper SE	2018	[Un18]
	E.ON SE	2018	[EON18]
Tourism	TUI Group	2018	[TUI18]
	Lufthansa Group	2018	[Lu18]
Telecommunication	Deutsche Telekom AG	2018	[DT18]
	Vodafone Group Plc.	2018	[Vo18]

4 A Refined Set of Socially Business Process Patterns

In line with our two research questions, we differentiate between existing patterns that could be confirmed through reviewing the sample of sustainability reports and new patterns that provide additional solutions for achieving social sustainability. The total set of SBPPs contains nine patterns from which four patterns are detailed through two sub-patterns (see Fig.2). In the following, each pattern is described in more detail by specifying a title, a short description, practical examples/references, and an icon.

Tab. 2: Confirmed/Revised and Additional Socially Business Process Patterns (SBPP)

Pattern title and description	
(1) Social compensation**	
Internal social compensation	
Improving internal social performance by compensating negative effects without changing the structure of a process. For example: health—preventive health care [TUI18, p. 70]; nutrition—information for nutrition [DT18, p. 82]; employment protection—guidelines for work safety [Si18, p. 20].	
External social compensation	
Improving external social performance by compensating negative effects without changing the structure of a process. For example: financial donations [BAS18, p. 40]; social projects—scholarships for women [IBM18, p. 36].	
(2) Social alternative	
Internal social alternative*	
Improving social performance by offering an additional alternative process path to the employees/partners without changing the (core) business processes. For example: flexible work—flexible working hours [Lu18, p. 68]; remote work—homeworking policies [Vo18, p. 17].	
External social alternative--	
Improving social performance by offering an additional alternative process path to the customers without changing the (core) business processes.	
(3) Social resource handling**	
Social resource replacement	
Improving social performance by replacing resources (i.e., working aids and materials) that have a negative influence on internal and external actors without changing the structure of a process. For example: hiring ergonomic-teams that ensure healthy working places [Co18, p. 33]; materials—avoiding so-called 'conflict materials' [Si18, p. 39].	
Social resource reuse*	
Improving social performance by reusing resources (i.e., circular-principle) that have a special value for society. For example: Effective management of water is important if projects are performed in regions that have water scarcity [Ho18, p. 138].	

Tab. 2: Confirmed/Revised and Additional Socially Business Process Patterns (SBPP)

Pattern title and description	
(4) Social labeling	
Improving the social image of an organization by labeling processes, products, and services with corresponding social certificates. For example: Fairtrade [Li16, p. 88]; Global Sustainable Tourism Council [TUI18, p. 11]; Blauer Engel [DT18, p. 83]; CmiA-certified cotton [Ot18, p. 83].	
(5) Social sourcing	
Improving social performance by (a) transferring activities to external organizations that can carry out an activity in a socially acceptable manner or by (b) transferring external activities into the own company.	
(6) Social value chain	
Improving social performance by defining social standards that have to be applied by every partner within a value chain. For example: New supplier needs to develop a social/ecological management system [IBM18, p. 18].	
(7) Social-/human-centered accessibility**	
Internal social accessibility Improving social performance by adjusting processes in terms of human-specific needs to allow employees/partners regardless of their background (e.g., gender, religion, physical health) to participate in a certain process. For example: Accessible working environments including lifts or subtitles that can be read aloud when using a computer [Si18, p. 18]. External social accessibility* Improving social performance by adjusting processes in terms of human-specific needs to allow customers/society regardless of their background (e.g., gender, religion, physical health) to participate in a certain process. For example: Accessible stores so that everyone can buy goods [Li16, p. 113].	
(8) Social training*	
Improving social performance by educating internal and external actors to enable them to participate in processes appropriately. For example: Competence management ensures that employees have the required knowledge to execute/participate in their tasks [RB18, p. 60].	
(9) Social togetherness*	
Improving social performance by creating a strong team-cohesion within an organization or company/department. For example: 'Walkathon Challenge' to support scholarships for women [IBM18, p. 36]; carpools [RB18, p. 58].	

* = new pattern; ** = revised pattern; -- = no practical solution obtained from the sample.

Following the notation for describing business process patterns as proposed by Nowak [No14]—who refined the pattern notation based on well-established approaches and guidelines such as from Alexander et al. [Al77a]—, we provide a graphical schema to visualize how the SBPPs can be employed for process improvement. By representing the solution scheme users are able to get a faster overview of the patterns [No14]. The representations consist of basic process model elements including activities, flows, gateways, resources, input/output objects, and external/internal actors. Whereas the lift side of the figure visualizes the original process, the right side of the figure highlights what changes occur from applying the patterns (see Fig. 3).

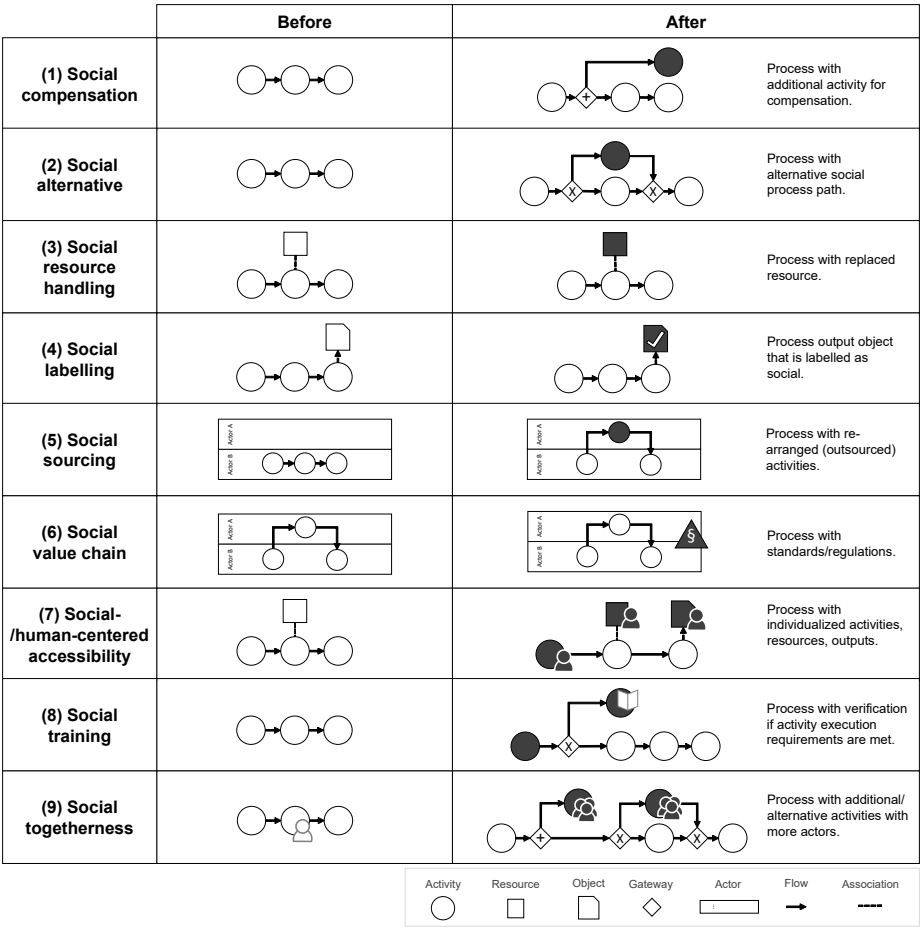


Fig. 2: Solution Scheme of SBPPs (notation adapted from [No14])

5 Discussion, Implications, and Avenues for Future Research

This study, as part of a larger research project aiming to design and evaluate business process patterns to leverage social sustainability, examines a sample of 24 sustainability reports from companies across diverse industry sectors and, based on that, derives example initiatives for validating, refining, and extending the initial set of patterns. Our **contribution** is threefold and has implications for research and practice:

First, as social concerns are often treated with secondary importance in IS research (e.g., [DLC11, SW19]), this study is a preliminary step to move IS towards considering all dimensions of sustainability. By focusing on the social dimension of sustainability, we complement existing research streams in the context of business process patterns and pose design-relevant knowledge that can be employed for process improvement/modeling. As a result, current collections of patterns such as the 'business process model patterns classification' (see www.bpmpatterns.org, [Sc20]) or the 'business process model pattern taxonomy' [Fe17] might be extended through social concerns—for example, by complementing the dimension for 'quality, compliance and risk' which already comprises approaches for the reduction of environmental impacts.

Second, from a practitioners' point of view, our refined set of SBPPs can be adapted to organizations that aim to improve their social performance in particular. The patterns (a) can help to create awareness of the rather vague concept of social sustainability that lacks a concise and well-accepted definition [Mc04] [MRB17], and thus, it is often hard to understand and implement in organizations. Moreover, the SBPPs (b) provide orientation and guidance on how to revise organizational processes by illustrating abstract solution schemes as well as specific examples from other companies.

Third, from a methodological view, as suggested by Nowak [No14], the development of process patterns should make use of existing domain-relevant knowledge. Therefore, this study presents an approach that extracts such knowledge from secondary data, here in the form of sustainability reports. In doing this, we seek to ensure both drawing on existing knowledge from the sustainability domain as well as ensuring practical relevance. Moreover, Fellmann et al. [Fe18] argued that the prevailing type of publications in the context of patterns is the *proposal of solution* (i.e., presenting a solution for a given problem by means of one or more patterns) and only 13% of their sample of publications fall into the category *evaluation research* (i.e., exploring the applicability and usefulness in practice). In pursuing to evaluate and demonstrate the SBPPs, this study itself can act as a demonstration case for pattern evaluation.

Even though this study provides some helpful attempts for incorporating social sustainability, it is not free of **limitations** that opens avenues for **future research**. We already mentioned in Section 3 that we found some dependencies between the patterns, for example, healthy working conditions can be achieved through 'social compensation' (e.g., offering additional activities), 'social alternative' (e.g., offering a completely new process path with fewer

health issues), or 'social training' (e.g., offering education so that employees can handle working materials appropriately). Accordingly, future research is required to investigate (a) whether a specific pattern can only be applied alone, (b) whether patterns can be employed together, and (c) whether patterns exclude a subset of other patterns. Therefore, a cluster analysis might be helpful in seeking to identify relationships.

Another limitation is the fact that the current set of SBPPs contains solely 'positive pattern' (i.e., description of a proven solution to a recurring problem that is related to the (re-)design of business process models, [Fe18]). In contrast to this type of pattern, there is research on 'anti-patterns' posing solutions that are known to have deficiencies and weaknesses [Be10] [KLF19]. This pattern type especially aims at identifying failures and shortcomings in current versions of business processes like, for instance, syntax errors, control-flow problems, understandability problems, or process-related defects including 'ecological impact' for highlighting negative impacts on the ecology [KLF19]. Accordingly, future endeavours might focus on creating anti-patterns for social sustainability in order to allow identifying issues in business process (e.g., activities that have a negative influence on employees and the society, or violation of social standards and regulations established by supply chains).

Although we demonstrate the applicability of the SBPPs, further research should deal with evaluating the set of patterns in a more naturalistic setting. Following the principles of design science research [GH13] [VPB16], the present study can be mapped as an *ex-post artificial evaluation* which "serves to initially demonstrate if and how well the artifact performs" [SB12, p. 395]. Consequently, *naturalistic evaluation* is required that "serves to ultimately show that an artifact is both applicable and useful in practice [and thus needs to include] real tasks, real systems, and real users" [SB12, p. 396]. Therefore, we plan to evaluate the artifact´s usefulness through conducting interviews with experts and a real-world case study that should disclose how the set of patterns can be employed during the (re-)design of business processes as well as which effects occur from that employment.

Furthermore, methodological limitations need to be considered such as (a) the bounded sample of sustainability reports that acts as the unit for analysis, (b) sustainability reports might naturally tend to reveal positive aspects, so, there is a risk of so-called 'green/social washing', (c) most of the reports are from larger organizations whereby solutions and initiatives from small and medium organizations are not included, and (d) the qualitative analysis is based on own decisions and interpretations.

6 Conclusion

Moving organizations towards more socially responsible entities is a challenging task, which requires appropriate tools and methods like those given by process patterns. In pursuing this ultimate goal, this study explores two research questions by obtaining uses cases from practice that support the initial set of SBPPs (RQ1) and refining the set of SBPPs through

additional findings from practice (RQ2). As a result, we propose nine different patterns leveraging social sustainability in business processes by examining solutions and initiatives from sustainability reports, which has implications for research and practice as well as opens different avenues for future research. Overall, we hope to shed light on the phenomenon of social sustainability in IS research.

Bibliography

[Al77a] Alexander, C.: A pattern language: towns, buildings, construction. Oxford university press, 1977.

[Al77b] Alexander, C.: The Timeless Way of Building. Oxford University Press, 1977.

[Al18] Allianz Group, https://www.allianz.com/content/dam/onemarketing/azcom/Allianz_com/investor-relations/en/results/2018_q4/en-190411-Sustainability-Report-2018.pdf, Access: 2020/05/11.

[Ba18] Bayer AG, https://www.investor.bayer.de/securedl/16789, Access: 2020/05/11.

[BAS18] BASF Group, https://bericht.basf.com/2018/de/serviceseiten/downloads/files/BASF_Bericht_2018.pdf, Access: 2020/05/11.

[Be10] Becker, J. et al.: Pattern-Based Semi-Automatic Analysis of Weaknesses in Semantic Business Process Models in the Banking Sector. In: Proceedings of the European Conference on Information Systems, Pretoria, South Africa, p. 156, 2010.

[Be17] Bertelsmann SE &Co. KGaA, https://www.bertelsmann.de/media/verantwortung/downloads/deutsch/bertelsmann-corporate-responsibility-bericht-2017-de.pdf, Access: 2020/05/11.

[BMW18] BMW Group, https://www.bmwgroup.com/content/dam/grpw/websites/bmwgroup_com/responsibility/downloads/de/2019/2019-BMW-Group-SVR-2018-Deutsch.pdf, Access: 2020/05/11.

[Bu11] Butler, T.: Compliance with institutional imperatives on environmental sustainability: Building theory on the role of Green IS. The Journal of Strategic Information Systems. 20, 1, 6–26 (2011).

[Bu17] Bundesgesetzblatt 2017 Teil I Nr. 20, Gesetz zur Stärkung der nichtfinanziellen Berichterstattung der Unternehmen in ihren Lage- und Konzernlageberichten (CSR-Richtlinie-Umsetzungsgesetz) vom 11. April 2017.

[Ca18] Canon Inc, https://global.canon/en/csr/report/pdf/canon-sus-2018-e.pdf, Access: 2020/05/11.

[Ci18] Cisco Systems Inc., https://www.cisco.com/c/dam/assets/csr/pdf/CSR-Report-2018.pdf, Access: 2020/05/11.

[CG17] CSR Europe & GRI (Global Reporting Initiative), Member State Implementation of Directive 2014/95/EU, www.globalreporting.org/resourcelibrary/NFRpublication%20online_version.pdf, Access: 2020/05/12.

[CG18] Carnahan, S., Greenwood, B.N.: Managers' Political Beliefs and Gender Inequality among Subordinates: Does His Ideology Matter More Than Hers? Administrative Science Quarterly 63/2, p. 287–322, 2018.

[Co18] Continental AG, https://www.continental.com/resource/blob/177190/37853d57eddfa3f46e64425352ebb552/nachhaltigkeitsbericht-2018-data.pdf, Access: 2020/05/11.

[DB18] Deutsche Bank AG, https://www.db.com/ir/de/download/Deutsche_Bank_Nichtfinanzieller_Bericht_2018.pdf, Access: 2020/05/11.

[DLC11] Dao, V., Langella, I., Carbo, J.: From green to sustainability: Information Technology and an integrated sustainability framework. The Journal of Strategic Information Systems 20/1, p. 63–79, 2011.

[DT18] Deutsche Telekom AG, https://bericht.telekom.com/geschaeftsbericht-2018/serviceseiten/downloads/files/entire_dtag_gb18.pdf, Access: 2020/05/11.

[El97] Elkington, J.: Cannibals with forks: The Triple Bottom Line of 21st Century Business. Capstone, Oxford, 1997.

[EON18] E.ON SE, https://www.eon.com/content/dam/eon/eon-com/Documents/en/sustainability-report/EON_Sustainability_Report_2018.pdf, Access: 2020/05/11.

[Fe17] Fellmann, M. et al.: A Taxonomy and Catalog of Business Process Model Patterns. In: Proceedings of the European Conference on Pattern Languages of Programs, Association for Computing Machinery, Irsee, Germany, pp. 1–10, 2017.

[Fe18] Fellmann, M. et al.: Business process model patterns: state-of-the-art, research classification and taxonomy. Business Process Management Journal, p. 972–994, 2018.

[Ga95] Gamma, E.: Design patterns: elements of reusable object-oriented software. Addison-Wesley Longman Publishing, Boston, MA, USA, 1995.

[GH13] Gregor, S., Hevner, A.: Positioning and Presenting Design Science Research for Maximum Impact. MIS Quarterly 37/2, p. 337–355, 2013.

[Ho18] Hochtief AG, https://www.hochtief.de/hochtief/mmdbdownload?id=173421&format=4, Access: 2020/05/11.

[IBM18] International Business Machines Corporation (IBM), https://www.ibm.org/static/responsibility/cr/pdfs/IBM-2018-CRR.pdf, Access: 2020/05/11.

[KLF19] Koschmider, A., Laue, R., Fellmann, M.: Business Process Model Anti-Patterns: A Bibliography and Taxonomy of Published Work. In: Proceedings of the European Conference in Information Systems, Stockholm-Uppsala, Sweden, 2019.

[KSK18] Kutzner, K., Schoormann, T., Knackstedt, R.: Digital Transformation in Information Systems Research: A Taxonomy-based Approach to Structure the Field. In: Proceedings of the European Conference on Information Systems, Portsmouth, UK.

[Li16] Lidl Dienstleistung GmbH & Co. KG, https://www.lidl-nachhaltigkeit.de/fileadmin/downloads/Lidl_Nachhaltigkeitsbericht_2016-2017.pdf, Access: 2020/05/11.

[Lo11] Loos, P. et al.: Green IT: A Matter of Business and Information Systems Engineering? Business & Information Systems Engineering 3/4, p. 245–252, 2011.

[Lu18] Lufthansa Group, https://www.lufthansagroup.com/media/downloads/en/responsibility/LH-sustainability-report-2018.pdf, Access: 2020/05/11.

[LFL17] Lübbecke, P., Fettke, P, Loos, P.: Sustainability Patterns for the Improvement of IT-Related Business Processes with Regard to Ecological Goals. In: Dumas, M. and Fantinato, M. (eds.) Business Process Management Workshops. Springer International Publishing, Cham, pp. 428–439, 2017.

[LFL16] Lübbecke, P., Fettke, P, Loos, P: Towards Ecological Workflow Patterns as an Instrument to Optimize Business Processes with Respect to Ecological Goals. In: Hawaii International Conference on System Sciences, Hawaii, USA, pp. 1049–1058, 2016.

[Mc04] McKenzie, S.: Social Sustainability: Towards some Definitions. Working Paper Series, Hawke Research Institute University of South Australia Magill, South Australia, 2004, accessed 2020/04/25.

[Me17] Metro AG, https://reports.metroag.de/corporate-responsibility-report/2017-2018/servicepages/downloads/files/entire_metro_crr1718.pdf, Access: 2020/05/11.

[MRB17] Missimer, M., Robèrt, K.H., Broman, G.: A strategic approach to social sustainability – Part 2: a principle-based definition. Journal of Cleaner Production 140, pp. 42–52, 2017.

[No14] Nowak, A.: Green Business Process Management: Methode und Realisierung. Dissertation, University of Stuttgart, Germany, 2014.

[Ol13] Olbrich, S. et al.: Inclusive Design-Theory: How to take advantage of diversity in Information Systems Design. In: Proceedings of the International Conference on Information Systems, Milan, Italy, 2013.

[OKK14] Opitz, N., Krüp, H., Kolbe, L.M.: Green Business Process Management – A Definition and Research Framework. In: Proceedings of the 47th Hawaii International Conference on System Science, Hawaii, Waikoloa, USA, pp. 3808–3817, 2014.

[Ot18] Otto Group, https://www.ottogroup.com/media/docs/de/geschaeftsbericht/Otto_Group_Geschaeftsbericht_2018_19_DE.pdf, Access: 2020/05/11.

[Re12] Recker, J. et al.: Modeling and Analyzing the Carbon Footprint of Business Processes. In: vom Brocke, J. et al. (eds.) Green Business Process Management. Springer, Berlin, Heidelberg, pp. 93–109, 2012.

[RB18] Robert Bosch GmbH, https://assets.bosch.com/media/global/sustainability/reporting_and_data/2018/bosch-nachhaltigkeitsbericht-2018-factbook.pdf, Access: 2020/05/11.

[Ro06] Rosemann, M.: Potential pitfalls of process modeling: part A. Business Process Management Journal 12/2, pp. 249–254, 2006)

[SBK17] Schoormann, T., Behrens, D., Knackstedt, R.: Sustainability in Business Process Models: A Taxonomy-Driven Approach to Synthesize Knowledge and Structure the Field. In: Proceedings of the International Conference on Information Systems, Seoul, Korea, 2017.

[SB12] Sonnenberg, C., vom Brocke, J.: Evaluations in the Science of the Artificial – Reconsidering the Build-Evaluate Pattern in Design Science Research. In: Design Science Research in Information Systems. Advances in Theory and Practice. Springer, Berlin, Heidelberg, pp. 381–397, 2012.

[SBW12] von Stetten, A., Beimborn, D., Weitzel, T.: Analyzing and Managing the Impact of Cultural Behavior Patterns on Social Capital in Multinational IT Project Teams. Business & Information Systems Engineering 4/3, pp. 137–151, 2012.

[Sc20] Schoknecht, A. et al.: Process Pattern, http://bpmpatterns.org/, Access: 2020/05/11.

[Sc19] Schoormann, T. et al.: Elevating Social Sustainability in Business Processes: A Pattern-Based Approach. In: Proceedings of the International Conference on Information Systems, Munich, Germany, 2019.

[Si18] Siemens AG, https://www.siemens.com/investor/pool/de/investor_relations/siemens_nachhaltigkeitsinformationen2018.pdf, Access: 2020/05/11.

[SSK19] Stadtländer, M., Schoormann, T., Knackstedt, R.: How Software Promotes the Integration of Sustainability in Business Process Management. In: Proceedings of the 14th Wirtschaftsinformatik, Siegen, Germany, pp. 1942–1953, 2019.

[St18] Strabag SE, https://www.strabag.com/databases/internet/_public/files.nsf/SearchView/6C0A043CAC95207FC12583ED0031FAD9/$File/STRABAG_SE_D_GB_2018_web.pdf?OpenElement, Access: 2020/05/11.

[St20] statista: Umsätze der wichtigsten Industriebranchen in Deutschland bis 2018, https://de.statista.com/statistik/daten/studie/241480/umfrage/umsaetze-der-wichtigsten-industriebranchen-in-deutschland/, Access: 2020/05/11.

[SW00] Smith, C.U., Williams, L.G.: Software Performance Antipatterns. In: Proceedings of the 2Nd International Workshop on Software and Performance. ACM, New York, NY, USA, pp. 127–136, 2000.

[SW19] Sadok, M., Welch, C.: Achieving Sustainable Business Systems Through Sociotechnical Perspectives. In: Proceedings of the European Conference in Information Systems, Stockholm & Uppsala, Sweden, 2019.

[TUI18] TUI Group, https://www.tuigroup.com/damfiles/default/tuigroup-15/de/nachhaltigkeit/berichterstattung-downloads/2019/nachhaltigkeitsbericht-de-en/TUI_CSR18_DE.pdf-91be99e0bc67220ef211280d05702d2a.pdf, Access: 2020/05/11.

[Un18] Uniper SE, https://ir.uniper.energy/download/companies/uniperag/Sustainability/Uniper_Nachhaltigkeitsbericht_2018.pdf, Access: 2020/05/11.

[Vo18] Vodafone Group Plc., https://www.vodafone.com/content/dam/vodcom/sustainability/pdfs/sustainablebusiness2018.pdf, Access: 2020/05/11.

[VPB16] Venable, J., Pries-Heje, J., Baskerville, R.: FEDS: A Framework for Evaluation in Design Science Research. European Journal of Information Systems 25/1, pp. 77–89, 2016.

[VW18] Volkswagen AG, https://www.volkswagenag.com/presence/nachhaltigkeit/documents/sustainability-report/2018/Nichtfinanzieller_Bericht_2018_d.pdf, Access: 2020/05/11.

[Wi09] Winter, R. et al.: Patterns in Business and Information Systems Engineering. Business & Information Systems Engineering 1/6, p. 468, 2009.

R. Reussner, A. Koziolek, R. Heinrich (Hrsg.): INFORMATIK 2020,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2021 605

Stand der anwendungsnahen Forschung und Technik für die organisatorische Perspektive von Geschäftsprozessen

Thomas Bauer[1], Ralf Laue[2]

Abstract: Damit die Diskussion über Geschäftsprozesse und die Erstellung von Prozessanwendungen einfach möglich ist, können Geschäftsprozesse graphisch modelliert werden. Allerdings ist dies für die organisatorische Perspektive häufig nicht möglich. Um organisatorische Anforderungen darzustellen, sind teilweise sogar Programmierkenntnisse erforderlich. Diese Arbeit untersucht, für welche Einzelaspekte und warum das so ist: Existieren hierzu keine geeigneten Forschungsansätze oder werden diese von kommerziellen Prozess-Management-Systemen (PMS) lediglich nicht umgesetzt? Um dies zu beantworten, wird der aktuelle Stand der anwendungs(system)nahen Forschung für die organisatorische Perspektive dargestellt und mit den von vier kommerziellen PMS angebotenen Funktionalitäten verglichen. Dabei werden mehrere Aspekte der organisatorischen Perspektive betrachtet, nicht nur die Zuordnung von Bearbeitern zu Aktivitäten. Als Ergebnis werden Forschungsfragen und Herausforderungen benannt, die zu lösen sind, um die Modellierung der organisatorischen Perspektive ähnlich komfortabel wie die Modellierung der Kontrollflussperspektive zu gestalten.

Keywords: Organisatorische Perspektive; Organisationsmodell; Bearbeiterzuordnung; Eskalation; Stellvertretung; Stand der Forschung; kommerzielle Systeme; PAIS; Forschungsfragen

1 Motivation

Prozessorientierte Anwendungen werden häufig basierend auf Prozess-Management-Systemen (PMS) erstellt. Hierbei müssen (zur Modellierungszeit) für die organisatorische Perspektive diverse Sachverhalte modelliert werden, die zur Ausführungszeit von den Endanwendern genutzt werden. So muss eine Aktivität den geeigneten (potentiellen) Bearbeitern angeboten werden, damit ein Benutzer diese reservieren, starten und bearbeiten kann. In Ausnahmefällen müssen Aktivitäten zurück- oder weitergegeben (delegiert) werden können. Das PMS sollte zudem Verzögerungen bei der Aktivitätenbearbeitung erkennen und ggf. Eskalation einleiten, sowie – bei längerfristiger Abwesenheit eines Benutzers – Aktivitäten an seine Stellvertreter weiterleiten können. Einige dieser Funktionalitäten werden in modernen PMS gut unterstützt und sind sogar einfach definierbar (z.B. Rückgabe oder Delegation einer Aktivität). Bei diesen ist ein ähnlicher Stand erreicht, wie auch für andere Perspektiven von Geschäftsprozessen (GP), d.h. ihre Festlegung erfordert keine Spezial- oder Programmierkenntnisse. Für die Festlegung diverser anderer Aspekte der

[1] Hochschule Neu-Ulm, Fakultät Informationsmanagement, Wileystr. 1, 89231 Neu-Ulm, thomas.bauer@hnu.de
[2] Westsächsische Hochschule Zwickau, Fakultät für Physikalische Technik / Informatik, Kornmarkt 1, 08056 Zwickau, ralf.laue@fh-zwickau.de

doi:10.18420/inf2020_54

organisatorischen Perspektive gilt das aber nicht: So wird z.B. für Bearbeiterzuordnungen nur eine sehr eingeschränkte Funktionalität angeboten oder deren Erstellung erfordert Programmierung (z.B. JavaScript).

Das Ziel dieses Beitrags ist es, zu untersuchen, warum die organisatorische Perspektive heutzutage in PMS teilweise noch unbefriedigend unterstützt wird. Hierzu werden folgende Fragestellungen betrachtet: Welche Aspekte werden unzureichend unterstützt? Gibt es hierfür keine geeigneten Ansätze aus der Forschung? Warum werden vorhandene Ansätze in PMS nicht eingesetzt?

Diese Fragestellungen wurden in der wissenschaftlichen Literatur bisher nur unzureichend betrachtet. [Ar18] betont zwar die Wichtigkeit der organisatorischen Perspektive und betrachtet sehr viele einschlägige wissenschaftliche Arbeiten, jedoch nur in Form einer Systematic Mapping Study (SMS). Der Inhalt dieser Arbeiten wird also nicht detailliert analysiert. Eine der Erkenntnisse ist jedoch, dass nur 14% der Veröffentlichungen in die Kategorie „Proposal of solution“ fallen. Die betroffenen Themengebiete werden aber nicht dargestellt. Diese Forschungslücke soll unser Beitrag verkleinern, um die im vorherigen Absatz genannten Fragestellungen zu beantworten und Themen mit aktuellem Forschungsbedarf zu identifizieren. Hierzu werden im Folgenden wissenschaftliche Arbeiten für die organisatorische Perspektive vorgestellt. Außerdem werden mehrere kommerzielle PMS untersucht. Schließlich wird der Stand der wissenschaftlichen Forschung mit den Funktionalitäten dieser Systeme verglichen, um zukünftige Forschungsbedarfe abzuleiten.

Der Fokus dieses Beitrags liegt auf anwendungsnaher Forschung, weil es das Ziel ist, dass die Funktionalität (kommerzieller) PMS in naher Zukunft besser wird. Betrachtet werden also Ansätze, die realistischerweise mit heutiger PMS-Technologie implementierbar wären, d.h. ohne das PMS weitgehend neu erstellen zu müssen. Deshalb werden die folgenden Forschungsfragen unter dieser Randbedingung betrachtet: Welche wissenschaftlichen Arbeiten gibt es zur Zuordnung von Personen zu GP-Aktivitäten (reguläre Bearbeiterzuordnungen, sowie Behandlung von Ausnahmefällen) und inwieweit sind diese in heutigen kommerziellen PMS umgesetzt? Existiert hierbei eine Diskrepanz und welche Forschungsbedarfe lassen sich hieraus ableiten?

Abschnitt 2 stellt den aktuellen Stand der Wissenschaft dar, Abschnitt 3 den einiger kommerzieller Systeme. Der Beitrag schließt mit daraus resultierenden Fragestellungen.

2 Stand der wissenschaftlichen Literatur

Im Folgenden wird der Stand der wissenschaftlichen Literatur[3] zur organisatorischen Perspektive von GP überblicksartig dargestellt. Wie erwähnt, liegt der Fokus hierbei auf anwendungsnaher Forschung. Hierbei werden alle zum Thema Bearbeiterzuordnung gehörenden Aspekte betrachtet, d.h. außer der reinen Festlegung der regulären Bearbeiter einer Aktivität werden auch Ausnahmebehandlungen durch Eskalationen und Stellvertretungen einbezogen.

Die für die Praxis durchaus relevanten Aspekte Delegation (an eine vom Benutzer direkt festgelegte Person), Zurückgeben und Pull-Mechanismen (z.B. Round-Robin) für Arbeitslisteneinträge sind nicht sehr kompliziert und in heutigen PMS oft gut umgesetzt [Cz19]. Deshalb werden sie im Folgenden nicht betrachtet. Außerdem wird nicht auf die Zuordnung anderer Ressourcentypen (außer Bearbeitern) eingegangen.

2.1 Literatur zum Organisationsmodell

Objekte des Organisationsmodells werden in Bearbeiterzuordnungen verwendet, um die potentiellen Bearbeiter einer Aktivität zu berechnen. Im Folgenden werden zuerst Arbeiten vorgestellt, die sich damit beschäftigen, welche Objekttypen ein Organisationsmodell enthalten soll. Danach wird betrachtet, wie prozessorientierte Organisationsmodelle auf Basis eines normalen Benutzer-Verzeichnisdienstes realisiert werden können.

Metamodell zur Speicherung organisatorischer Objekte: [Ru05] stellt diverse Muster für den organisatorischen Aspekt vor. Die Arbeit beinhaltet auch ein Metamodell für ein Organisationsmodell: Benutzer (Human Resources) haben eine Position in einer Organisationseinheit. Sie können temporär existierenden Gruppen (Organisational Teams) angehören. Außerdem sind ihnen Rollen, Fähigkeiten, eine organisatorische Ebene, vorgesetzte bzw. untergeordnete Personen und spezielle Charakteristika zugeordnet. Damit beschreibt dieses Metamodell eine Vielzahl an Objekttypen, die als Basis zur Definition zahlreicher Bearbeiterzuordnungen (vgl. Abschnitt 2.2) dienen können.

In [OS10] wird ein Metamodell für die organisatorische Perspektive vorgestellt, das insbesondere Kompetenzen, Fertigkeiten und Wissen betrachtet. Außer Benutzern sind darin auch andere Arten von Ressourcen vorgesehen.

Der Fokus von [Aw09] ist die Integration von Bearbeiterzuordnungen in BPMN. Es wird auch ein Organisationsmetamodell vorgestellt, das allerdings lediglich organisationale und funktionale Rollen inkl. einer Hierarchie enthält. Andere Objekttypen (z.B. Fähigkeiten) werden nicht berücksichtigt.

[3] Die Literaturrecherche erfolgte hauptsächlich in Google-Scholar unter anderem mit den Suchbegriffen Organizational Model / Resource / Staff / Actor Assignment / Escalation / Substitution jeweils kombiniert mit Business Process / Workflow. Außerdem wurden in den ermittelten Publikationen enthaltene Literaturreferenzen untersucht. Schließlich wurden alle in [Ar18] der Kategorie „Proposal of solution“ zugeordneten Arbeiten genau betrachtet und auf Relevanz geprüft.

Auch [SCV15] erweitert BPMN um eine organisatorische Perspektive. Hierzu wird für einige wenige organisatorische Objekttypen eine graphische Visualisierung vorgestellt. Kern der Arbeit ist jedoch deren Integration in BPMN.

Externer Benutzer-Verzeichnisdienst: Es gibt keine wissenschaftlichen Arbeiten, die explizit untersuchen, wie ein Organisationsmodell für ein PMS in einem kommerziellen Verzeichnisdienst (z.B. LDAP, Active Directory) realisiert werden sollte, d.h. wie die Prozessobjekte (Rollen, Fähigkeiten, etc.) auf diese Struktur abgebildet werden können. Auch gibt es keine Literatur, welche die Anbindung eines PMS an einen solchen Dienst betrachtet. Es findet sich lediglich die Aussage, dass Benutzer und Rollen in Unternehmen in kommerziellen Verzeichnisdiensten wie z.B. Microsoft Active Directory [KRS13] oder einem LDAP-Verzeichnis definiert sind.

[LSR14] betrachtet das Problem, dass Organisationsmodelle in Unternehmen redundant für jede Applikation erstellt und gepflegt werden. Deshalb wird C-Org als zentraler Server für die Speicherung des Organisationsmodells vorgestellt. Jedoch basiert C-Org nicht auf einem kommerziellen Verzeichnisdienst, sondern verwendet ein im Projekt selbst entwickeltes Datenmodell, das die Modellierung beliebiger Beziehungen ermöglicht.

[KRS13] realisiert ein subjektorientiertes PMS. Das Organisationsmodell hierfür wird in einem Active Directory von Microsoft realisiert, das ohnehin schon in der IT-Infrastruktur des betroffenen Unternehmens vorhanden ist. Auf dessen Aufbau wird nicht näher eingegangen. Es wird nur erwähnt, dass dort Benutzer und Rollen gespeichert sind. Weitere organisatorische Objekte (Kompetenzen, Abteilungen, etc.) werden nicht betrachtet.

In einigen Arbeiten werden kommerzielle Verzeichnisdienste als Randthema erwähnt: In [BB01] stehen PMS nicht im Fokus (sondern Model-Driven Architecture), werden aber als eine mögliche Architekturvariante erwähnt. Die vorgestellte Architektur enthält ein Organisationsmodell, das auf ein LDAP-Verzeichnis abgebildet wird. Die zur Ausführungszeit benötigten LDAP-Anfragen werden aus dem Prozessmodell generiert. Das Kernthema von [KLW09] ist eine Kategorisierung von BPM-Standards. Beim Thema Konfiguration des PMS wird erwähnt, dass eine Synchronisation mit dem Active Directory des Unternehmens erfolgen muss, um Rollen und Benutzer-Accounts zu ermitteln. In [EP99] wird ein verteiltes LDAP-Verzeichnis verwendet, um Information über die GP, aber auch organisatorische Information wie z.B. Rollen, zu speichern.

2.2 Literatur zu Bearbeiterzuordnungen

Zur Modellierungszeit werden Bearbeiterzuordnungen modelliert, um die potentiellen Bearbeiter einer Aktivität zur Ausführungszeit ermitteln zu können. In diesen Bearbeiterzuordnungen wird hierzu auf Objekte des Organisationsmodells verwiesen. Im Folgenden werden Arbeiten vorgestellt, die dieses Thema betrachten.

Unabhängige Bearbeiterzuordnungen: Im einfachsten Fall ist die Menge der potentiellen Bearbeiter einer Aktivität nicht von Daten der betroffenen Prozessinstanz abhängig, sondern wird ausschließlich auf Basis des zugrundeliegenden Organisationsmodells berechnet (z.B. Rolle = Designer). Es ergeben sich also bei allen Prozessinstanzen dieselben Bearbeiter. Die Existenz solcher Bearbeiterzuordnung wird (zumindest implizit) in allen Arbeiten angenommen. So werden in [Ru05] die erwähnten Organisationsmodellobjekte mit dem Ziel eingeführt, darauf basierende Bearbeiterzuordnungen definieren zu können. Dies wird z.B. in [CRR11] aufgegriffen und detailliert.

Abhängige Bearbeiterzuordnungen: Diese ermöglichen eine flexiblere Festlegung der potentiellen Bearbeiter, weil in solchen Regeln Prozessvariablen, Bearbeiter früherer Aktivitäten, deren Rolle, Abteilung, etc. verwendet werden können. So kann z.B. festgelegt werden, dass ein Prüfschritt von einem Mitarbeiter derselben Abteilung wie Aktivität X, aber von einer anderen Person als Aktivität X (4-Augen-Prinzip) durchgeführt werden soll.

In [Ru05] werden einige wenige abhängige Bearbeiterzuordnungen erwähnt. Hierbei handelt es sich um das bereits erläuterte 4-Augen Prinzip (Separation of Duties) und die Möglichkeit, eine Aktivität demselben Bearbeiter zuzuordnen, der bereits eine bestimmte Vorgängeraktivität durchgeführt hat.

[BFA99] definiert eine Sprache zur Definition von zusätzlichen Einschränkungen (Constraints) innerhalb von Bearbeiterzuordnungen. Mit diesen kann z.B. das 4-Augen-Prinzip umgesetzt werden. Außerdem werden Algorithmen vorgestellt, welche die Konsistenz der Constraints prüfen und die Benutzer den Aktivitäten so zuordnen, dass keine Constraints verletzt werden.

In [KR09] werden einzelne Anforderungen an abhängige Bearbeiterzuordnungen erwähnt, wie z.B. die Abhängigkeit der Bearbeiter von Prozessinstanzdaten. Einige andere Arbeiten (z.B. [SKR14], [WS07]) verwenden einzelne abhängige Bearbeiterzuordnungen, ohne weiter auf dieses Thema einzugehen.

Art der Festlegung von Bearbeiterzuordnungen: Außer der Mächtigkeit von Bearbeiterzuordnungen ist entscheidend, wie (einfach) diese definiert werden können. In PMS reicht dies von der Befüllung eines Eingabefelds mit z.B. einem Gruppennamen bis hin zur eigenen Programmierung einer Regel z.B. mit JavaScript (vgl. Abschnitt 3).

Der BPMN-Standard [OMG11] ermöglicht die Festlegung einer Bearbeiterzuordnung mittels XPath. Allerdings muss der erforderliche XPath-Ausdruck durch den Modellierer selbst definiert werden, was eine ähnliche Schwierigkeit darstellt, wie das erwähnte Programmieren einer Regel.

[Aw09] schlägt eine Erweiterung des BPMN-Metamodells um Bearbeiterzuordnungen vor. Diese werden mittels Constraints der Object Constraint Language (OCL) definiert, sind also ebenfalls vom Modellierer textuell zu erstellen.

[CRR11] entwickelt die Sprache RAL zur Ressourcenzuordnung. In dieser werden typisierte organisatorische Objekte (Role, Capability . . .) verwendet. Vorteil dieser Sprache ist, dass sie sehr mächtig und gut lesbar (d.h. einfach) ist. Dieser Ansatz wird in [Ca15] um die graphische Notation RALph erweitert. Diese ist leicht verständlich und deckt alle Ressource-Muster [Ru05] ab. Regeln können auch mit booleschen Operation kombiniert werden, so dass mit diesem Ansatz wohl alle in der Praxis üblicherweise auftretenden Anforderungen an Bearbeiterzuordnung realisierbar sind. Allerdings berücksichtigt er keine Eskalationen und Stellvertretungen (vgl. Abschnitt 2.3). Stattdessen wird im Ausblick lediglich erwähnt, dass es auch noch andere Arten von Verantwortlichkeiten gibt.

Erweiterungen von GP-Modellierungssprachen: Einige Arbeiten beschäftigen sich mit der Erweiterung von standardisierten GP-Modellierungssprachen wie BPMN [OMG11] und UML, da diese die organisatorische Perspektive unzureichend unterstützen. [Gr08] erweitert das BPMN-Datenmodell (BPDM), um so alle in den Ressource-Muster [Ru05] aufgelisteten Bearbeiterzuordnungen zu unterstützen. In [SCV15] werden Metamodelle (UML-Klassen) zur Speicherung der organisatorischen Perspektive entwickelt. Sie erweitern ebenfalls das BPMN-Metamodell, enthalten aber keine konkreten Vorschläge für zusätzliche Objekttypen (wie z.B. Rollen, Abteilungen, Fähigkeiten). [WS07] beschreibt eine BPMN-Erweiterung, mit der zusätzliche Constraints an Bearbeiterzuordnungen definiert werden können. Dies reicht von einfachen (z.B. selber Bearbeiter, 4-Augen-Prinzip) bis zu sehr weitreichende Regeln (max. 5 Aktivitäten durch denselben Benutzer). Die Modellierung erfolgt durch die Festlegung von Schwellwerten (d.h. Zahlen), was für Geschäftsprozessmodellierer zumindest ungewohnt ist.

Andere Arbeiten behandeln eine Erweiterung von UML: [SM11] erweitert UML- Aktivitätsdiagramme um die organisatorische Perspektive. Neu entwickelte „Business Activities“ realisieren hierzu eine Rollen-basierte Zugriffskontrolle (RBAC). In [Li08] wird das UML-Anwendungsfalldiagramm erweitert, um z.B. Bearbeiterzuordnungen, Delegation und Eskalation zu ermöglichen. Das Konzept ermöglicht die Transformation des Platform Independent Models (PIM) in ein Platform Specific Model (PSM), welches dann zur GP-Steuerung durch ein PMS verwendet wird.

Constraints zur Sicherstellung der Compliance: Compliance-Richtlinien werden in Unternehmen üblicherweise unabhängig von den GP definiert, betreffen aber auch die Zuordnung von Bearbeitern zu Aktivitäten. Entsprechende Regeln definiert [NS07] als sog. Internal Controls. Eine zusätzliche Schicht löst Recovery-Aktionen aus, falls eine der Internal Controls bei der GP-Ausführung verletzt wird. [SKR14] erlaubt eine graphische Modellierung der zusätzlichen Constraints. Außerdem wird automatisch geprüft, ob erfolgte Bearbeiterzuordnungen korrekt waren. [KRK15] realisiert ein Monitoring der Prozessausführung inkl. einer graphischen Visualisierung von Constraint-Verletzungen, damit die Benutzer diese einfacher nachvollziehen und darauf reagieren können.

Weitergehende Bearbeiterzuordnungen: Im Folgenden werden einige Arbeiten vorgestellt, deren Konzepte auf Basis heutiger PMS schwer umsetzbar sind, bzw. die eher seltene

Anforderungen betrachten. In [BE01] wird das 4-Augen-Prinzip um sehr weitgehende Anforderungen erweitert, indem Konflikte zwischen Rollen oder Benutzern berücksichtigt werden. So darf z.B. ein Familienmitglied des Antragstellers eine Genehmigungsaktivität nicht ausführen. [AK01] erweitert das Organisationsmodell um ein Team-Konzept. Eine einzelne Aktivität wird dann von einem Team mit mehreren Bearbeitern ausgeführt anstatt einem einzelnen Bearbeiter. In [CRR12b] werden einer Aktivität außer ihrem (normalen) Bearbeiter zusätzlich Personen zur Kontrolle, Information und Support der Ausführung zugeordnet (vgl. RACI-Matrix). [CRR12a] erweitert die Sprache RAL, um Abhängigkeiten zwischen den Bearbeitern unterschiedlicher Prozessinstanzen definieren zu können. Bei [SRS08] darf einer Bearbeiterzuordnung auch Soft-Constraints enthalten. Das sind Regelteile, die eingehalten werden sollen, aber nicht müssen.

2.3 Behandlung von Ausnahmesituationen

Im Folgenden werden Eskalationen und Stellvertretungen betrachtet. Diese haben gemeinsam, dass das PMS auf eine zuvor festgelegte Art und Weise auf Ausnahmesituationen reagiert, nämlich auf eine zu große Verzögerung bei einer Aktivitätenbearbeitung bzw. auf die (längerfristige) Abwesenheit der regulären Bearbeiter.

Eskalationen: Wird eine Aktivität von einem Benutzer nicht rechtzeitig gestartet oder beendet, kann das PMS eine Nachricht versenden oder die Aktivität automatisch an eine andere Person delegieren. Bei der GP-Modellierung ist festzulegen, wer hierbei die Zielperson sein soll. Eine solche Festlegung ist komplexer als „normale Bearbeiterzuordnungen“, weil diese Zielperson nicht nur von der betroffenen Aktivität und Prozessinstanzdaten abhängig ist, sondern sich zudem für unterschiedliche „Original-Bearbeiter“ unterscheiden kann (z.B. dessen Vorgesetzter). Eskalationen werden in der wissenschaftlichen Literatur zu PMS zwar erwähnt (z.B. [Gr08], [Ru05]), es gibt aber keine Arbeiten, die sich speziell mit dieser Fragestellung befassen.

Im Folgenden werden Arbeiten vorgestellt, die sich mit speziellen Arten von Eskalationen befassen: Bei [ARD07] wird nicht nur auf erfolgte Zeitüberschreitungen reagiert, sondern es werden auch erwartete Zeiten für die Beendigung von Aktivitäten und Prozessinstanzen berücksichtigt. Ziel hierbei ist, rechtzeitig reagieren zu können. Außerdem kann eine Eskalation mehrere Aktivitäten oder Prozessinstanzen betreffen. Die Art der Eskalation wird automatisch oder auch von einem Menschen gewählt. Zudem werden unterschiedliche Eskalationsstufen unterschieden, die nacheinander durchlaufen werden, und zu verschiedenen Eskalationen führen. Es werden mehrere Eskalationsstrategien vorgestellt, wie z.B. Ausführung einer alternativen Aktivität, Parallelisierung, Bereitstellung zusätzlicher Ressourcen. Die Arbeit entwickelt also einen sehr mächtigen Eskalationsmechanismus, zugeschnitten auf das betrachtete Szenario der Überlastung eines Call-Centers. Die oben erwähnte Fragestellung der Modellierung einer Zielperson für eine Eskalation wird nicht betrachtet.

Ein Ziel von [PR98] ist ein besseres Management von Eskalationen, um die Anzahl der Eskalationen minimieren. Hierzu passt ein Algorithmus Deadlines an, um Verzögerungen auszugleichen und damit Eskalationen zu vermeiden. Außerdem werden die durch nicht vermeidbare Eskalationen entstehenden Kosten minimiert. Hierzu wird vorgesagt, ob Eskalation zu erwarten sind. In einem solchen Fall erfolgen diese Eskalationen möglichst früh (im GP), weil Kosten für frühe Eskalationen meist niedriger sind.

Stellvertretungen: Sind Benutzer längerfristig abwesend, so verbleiben für bestimmte Aktivitäten evtl. keine oder zu wenig Bearbeiter. Betroffene Aktivitäten sollten dann vom PMS automatisch Stellvertretern zugeordnet werden.

[Ba09] definiert diverse Anforderungen an Stellvertreterregelungen. So sollten diese abhängig von der abwesenden Person oder von der betroffenen Aktivität modellierbar sein. Zudem muss festgelegt werden, ob die Regelung aktiviert wird, wenn ein bzw. alle potentiellen Bearbeiter abwesend sind, oder explizit ihre Stellvertretung aktiviert haben, ob mehrstufige Stellvertretungen gewünscht sind (falls Stellvertreter selbst abwesend sind), und ob Aktivitäten den Stellvertretern entzogen werden sollen, wenn reguläre Bearbeiter zurückkehren. Außerdem werden Algorithmen zur Berechnung der Stellvertreter vorgestellt. Die Arbeit enthält jedoch kein Konzept, das eine „einfache Modellierung“ solcher Stellvertreterregelungen ermöglicht. Dies kann jedoch ohne eine geeignete Vorgehensweise sehr aufwendig werden, wenn viele unterschiedliche Stellvertreterregelungen abhängig vom Originalbearbeiter, der betroffenen Aktivität und dem Prozesskontext (Daten) erstellt werden müssen. Die Arbeit enthält lediglich den Hinweis, dass diese Anzahl reduziert werden kann, indem Regelungen für eine gesamte Klasse von Aktivitäten oder Prozessvorlagen definiert werden.

Auch andere Arbeiten zum Thema Stellvertretungen machen hierzu keine Vorschläge: In [Mu04] wird ein einfacher Stellvertreter-Mechanismus beschrieben. Hierzu wird ein (ggf. eingeschränkter) Zugriff auf die Arbeitsliste der zu vertretenden Person erlaubt und die Stellvertreter können mittels Rollen definiert werden. In [Ru05] werden Stellvertretungen nicht gesondert betrachtet, sondern als vom PMS durchgeführte Delegation. [HD05] erwähnt lediglich die Notwendigkeit von Stellvertretungen, es werden aber keine Einzel-Anforderungen genannt oder Lösungskonzepte vorstellt. [RM98] beschreibt ein Metamodell für die organisatorische Perspektive, die auch Stellvertretungen enthält. Es sieht jedoch keine Abhängigkeit entsprechender Regeln von der betroffenen Aktivität vor. Die Notwendigkeit einer „kontextbezogenen Stellvertretung“ wird zwar erwähnt, aber kein Konzept für die Modellierung solcher Stellvertreterregelungen vorgestellt.

In einigen Arbeiten wird die Nicht-Verfügbarkeit beliebiger Ressourcentypen diskutiert: [HS99] betrachtet verschiedene Arten von Policies beim Ressourcen-Management. Diese Policies, auch die für Stellvertretungen, werden mit einer SQL-artigen Sprache definiert. Hiermit kann festgelegt werden, dass eine bestimmte Menge von Stellvertretern X bestimmte reguläre Bearbeiter Y bei bestimmten Aktivitäten Z vertreten können. Die Mengen X, Y, Z werden über Rollen- bzw. Aktivitätennamen definiert und können mit Where-Klauseln weiter eingeschränkt werden. Auch bei [DW15] sind Stellvertreter nur ein Teilaspekt. Es

wird nicht nur die Abwesenheit von Bearbeitern betrachtet, sondern generell der Ausfall von Ressourcen. Dann soll eine automatische Optimierung, basierend auf Daten aus einer Prozess-Protokolldatei, mittels Process-Mining die optimale Ersatzressource ermitteln.

2.4 Sonstige Themen

Einige Arbeiten (siehe [Ar18]) beschäftigen sich mit Themen, die für die GP-Steuerung durch PMS weniger relevant sind. Deshalb wird auf diese nun nicht detailliert eingegangen. Betrachtete Themen sind z.B. das Mining von Bearbeiterzuordnungen aus Protokolldateien, die Simulation der Ressourcenauslastung oder die Optimierung der Ressourcenauswahl durch das PMS (mittels Heuristiken oder zusätzlicher Anforderungen).

3 Organisatorische Perspektive kommerzieller PMS

In einer Bachelorarbeit [Cz19] wurde die organisatorische Perspektive von 4 kommerziellen PMS untersucht und durch Recherchen der Autoren ergänzt. Im Folgenden werden die wichtigsten Erkenntnisse für die PMS Bizagi Studio V.11.2.3 [Bi20], IBM Business Process Designer V.8.0.1 und V.8.6.0 [IBM17], K2 Cloud V.4.0 [K20] und Signavio Workflow Accelerator V.13.6.0 [Si20] dargestellt. Dabei werden aus Platzgründen und, weil sie sich teilweise sehr ähnlich verhalten, nicht bei jedem Thema alle PMS betrachtet.

Organisationsmodell: Für das Metamodell bietet IBM am wenigsten Strukturierungsmöglichkeiten, da nur sog. Teilnehmergruppen angeboten werden. Signavio ermöglicht zusätzlich die Verwendung von Rollen, was K2 und Bizagi noch um Kompetenzen und Organisationseinheiten inkl. einer Hierarchiebildung (z.B. Team, Abteilung, Direktion) erweitern. Alle betrachteten Produkte ermöglichen die Anbindung eines externen Verzeichnisdienstes in Form eines Active-Directory, LDAP- oder Sharepoint-Servers, um auf Benutzerdaten und deren Gruppenzugehörigkeiten etc. zuzugreifen.

Bearbeiterzuordnungen: Beim Produkt *K2 Cloud V.4.0* kann eine Bearbeiterzuordnung mittels Bedingungen aus mehrere Regeln kombiniert werden. Es sind jedoch keine abhängigen Bearbeiterzuordnungen (vgl. Abschnitt 2.2) möglich. Außerdem können keine Bearbeiterzuordnungen selbst programmiert werden (z.B. mittels JavaScript), sondern sie werden stets durch das Befüllen des Formulars „Empfängerregel“ definiert. Beim *Workflow Accelerator* von Signavio werden Bearbeiterzuordnungen ebenfalls in einem Formular erstellt, es sind hierbei jedoch auch zwei Arten von abhängigen Bearbeiterzuordnungen möglich: Man kann fordern, dass eine Aktivität denselben Bearbeiter wie eine Vorgängeraktivität hat und das 4-Augen-Prinzip modellieren. Werden sonstige Bearbeiterzuordnungen benötigt, so können diese ausschließlich mittels JavaScript erstellt werden. *Bizagi Studio* ermöglicht die Erstellung komplexer Regeln (mit AND/OR-Verknüpfungen) in einem graphischen Editor (siehe Abb. 1a) und zusätzlich die Definition beliebiger Regeln mittels XPath. Beim

Business Process Designer von IBM kann einer Aktivität eine Gruppe, der Bearbeiter einer Vorgängeraktivität oder der Starter der Prozessinstanz zugeordnet werden. Außerdem können mittels selbst erstellbarer Routing-Policys komplexe Regeln mit von Prozessvariablen abhängigen Bedingungen und booleschen Verknüpfungen erstellt werden (vgl. Abb. 1b). Allerdings ermöglichen diese Art von Regeln keine abhängigen Bearbeiterzuordnungen und keine beliebigen booleschen Kombinationen von Bedingungen. Zudem wird dieser Mechanismus bei neueren Produktversionen als „Deprecated“ (veraltet) gekennzeichnet [IBM17], so dass er in zukünftigen Versionen wegfallen wird. Dann können komplexe Regeln nur noch mit JavaScript erstellt werden oder sie müssen in Aufrufe externer Services ausgelagert werden.

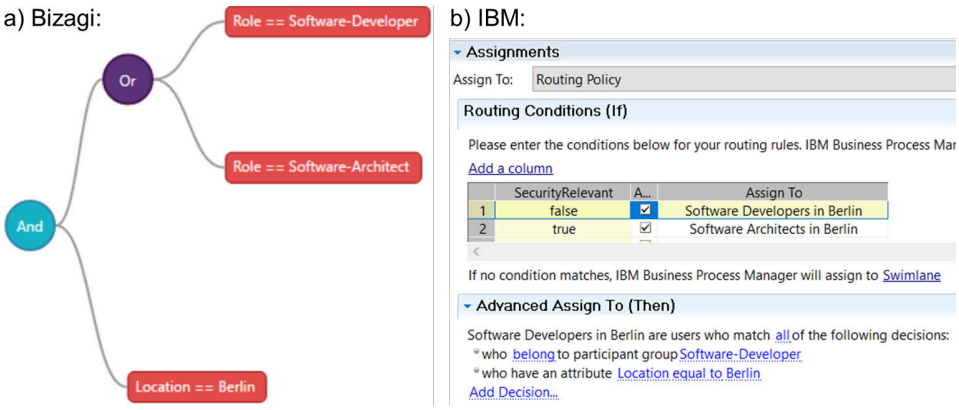


Abb. 1: a) Im Bizagi-Editor erstellte boolesche Verknüpfung mehrerer Einzelbedingungen. Der Bearbeiter der Akt. „Check Program Code“ muss die Rolle Software-Developer oder Software-Architect haben und außerdem dem Standort Berlin zugeordnet sein. b) Bei der von IBM angebotenen Routing-Richtlinie können zusätzlich Wenn-Dann-Regeln definiert werden. So wird die Aktivität bei nicht sicherheitsrelevanten Programmteilen (If) von einer als Software-Developer eingestuften Person aus Berlin durchgeführt (*Then: Location* ist hierbei ein benutzerdefiniertes Attribut einer Person) und bei sicherheitsrelevanten von einer als Software-Architect eingestuften Person in Berlin (die Definition dieser Regel ist in Abb. 1b nicht sichtbar, weil aktuell der erste Teil der Routing-Richtlinie definiert wird).

Behandlung von Ausnahmesituationen: Bizagi unterstützt keine Eskalationen. Signavio ermöglicht, dass bei einer Zeitüberschreitung eine E-Mail an Benutzer gesendet wird, die abhängig von der betroffenen Aktivität gestaltet werden kann. Bei K2 ist außer einer E-Mail auch die Weitergabe der Aktivität möglich, beides jedoch nur an einen einzelnen Benutzer (d.h. keine Gruppe). IBM ermöglicht ebenfalls die Weitergabe der Aktivität und Notifikationen per E-Mail. Eskalationen können an mehrere Personen gerichtet sein, bei verspäteter Reservierung oder Beendigung der Aktivität ausgelöst werden und es sind auch mehrere Eskalationen (parallel und nacheinander) definierbar.

Bizagi ermöglicht keine Stellvertretungen. Bei K2 können für unterschiedliche Aktivitäten verschiedene Stellvertreter definiert werden. Bei IBM ist für jede Aktivität wählbar, ob sie

an einen Stellvertreter weitergeben werden soll. Eine entsprechende Liste muss aber zuvor fest für den Benutzer konfiguriert werden, d.h. Stellvertreter können nicht abhängig von der betroffenen Aktivität oder Prozessdaten gewählt werden, und die Aktivität wird stets nur an den ersten anwesenden Stellvertreter dieser Liste weitergegeben.

4 Fazit und existierende Herausforderungen

Als Erkenntnis des letzten Abschnitts ist festzustellen, dass die Organisationsmodelle der verschiedenen Systeme nicht einheitlich sind und teilweise nur sehr wenige Objekttypen unterstützt werden (z.B. ausschließlich Gruppen). Abhängige Bearbeiterzuordnungen werden teilweise gar nicht unterstützt oder es sind nur sehr wenige und einfache Regeln vorgesehen. Auch die Kombination von Teilausdrücken mit booleschen Operationen ist teilweise gar nicht möglich und teilweise wenig komfortabel. Kein PMS erlaubt eine einfache Modellierung (z.B. graphisch, vgl. RALph [Ca15]) zusammengesetzter Regeln aus beliebigen (auch abhängigen) Teil-Bearbeiterzuordnungen, wie z.B. „Rolle = Sachbearbeiter AND Abteilung wie bei Akt. X AND NOT selber Bearbeiter wie Akt. X“. Auf Ausnahmesituationen kann teilweise nur unzureichend reagiert werden. Deshalb sollten zukünftig folgende Forschungsziele[4] verfolgt werden:

1. Evaluation, ob ein einheitliches Metamodell für Organisationsmodelle von PMS generell sinnvoll ist, wenn man unterschiedliche Anwendungen und Fachdomänen einbezieht: Es ist also zu klären, ob deren Anforderungen eher ähnlich oder extrem unterschiedlich sind. In ersterem Fall folgt daraus, dass für das Metamodell organisatorischer Daten ein Standard entwickelt werden sollte (ähnlich wie BPMN für den Kontrollfluss).
2. Evaluation, ob auch in unterschiedlichen Domänen üblicherweise dieselben Arten von Bearbeiterzuordnungen verwendet werden: Nur in diesem Fall ist es sinnvoll, in einem PMS eine solche (einfach verwendbare) Funktionalität zu implementieren. Sind die Anforderungen hingegen derart uneinheitlich, dass für sehr viele Bearbeiterzuordnungen Spezialfunktionalitäten erforderlich sind, dann ist deren Programmierung (z.B. in Java-Script) ohnehin in den meisten Fällen unvermeidbar.
3. Falls 2. mit ja beantwortet wurde, kann evaluiert werden, ob für Bearbeiterzuordnungen eine graphische Modellierung (vgl. RALph) oder eine textuelle (RAL, OCL) besser geeignet ist. Da hierbei die Verständlichkeit für GP-Modellierer entscheidend ist, müssen evtl. die fachliche (semantische) GP-Modellierung und technische GP-Implementierung getrennt betrachtet werden, weil sich die (IT-)Kenntnisse der jeweiligen Modellierer stark unterscheiden. Um die Benutzung von PMS unterschiedlicher Hersteller zu erleichtern, sollte auch für diese Funktionalität ein

[4] Die Autoren planen nicht, diese alle selbst zu bearbeiten. Zweck dieser Liste ist auch, anderen Forschungsgruppen Anregungen zu liefern.

Standard entwickelt werden. Dieser muss insb. beliebige (auch abhängige) Bearbeiterzuordnungen und die Kombination von Teilausdrücken mit booleschen Operationen beinhalten.

4. Erstellen eines Konzepts, um ein prozessorientiertes Organisationsmodell in einem „gewöhnlichen“ (kommerziellen) Verzeichnisdienst (z.B. LDAP, Active Directory) zu realisieren: Ein solcher existiert bereits in den meisten Organisationen und enthält für PMS relevante Informationen (z.B. Abteilungszugehörigkeiten, Vorgesetzte, Gruppen), die nicht redundant gepflegt werden sollen. Das Konzept muss klären, wie ein solcher Verzeichnisdienst um prozessorientierte Aspekte (z.B. Rollen, Kompetenzen) erweitert werden kann. Hierbei darf die Übersichtlichkeit nicht verloren gehen, damit er „leicht pflegbar“ bleibt. Auch die Schnittstelle zum PMS sollte betrachtet werden, da diese die effiziente Berechnung der aus komplexen Bearbeiterzuordnungen resultierenden Bearbeiter unterstützen muss.

5. Ein wissenschaftliches Konzept zur Modellierung von Eskalationen[5] für GP-Aktivitäten: Bei einer Eskalation ist die Menge der Zielpersonen nicht nur von der betroffenen Aktivität und dem Prozesskontext abhängig, sondern zudem von dem Bearbeiter, der diese Aktivität ursprünglich reserviert oder gestartet hat. Da es jedoch zu aufwendig wäre, separate Regeln für jede Kombination von Aktivität und Originalbearbeiter zu definieren, ist ein Konzept erforderlich, das sich deutlich von normalen Bearbeiterzuordnungen unterscheidet.

6. Ein Konzept für Stellvertretungen: Da der Stellvertreter auch hier abhängig vom Originalbearbeiter festgelegt werden kann, ergeben sich ähnliche Problemstellungen wie bei 5. Zusätzlich müssen noch Aspekte, wie mehrstufige Stellvertretungen oder der Entzug einer Stellvertretung bei Rückkehr des Originalbearbeiters berücksichtigt werden.

Literaturverzeichnis

[AK01] Aalst, W. M. van der; Kumar, A.: A Reference Model for Team-Enabled Workflow Management Systems. In Data & Knowledge Engineering, 2001; S. 335–363.

[Ar18] Arias, M. et al.: Human Resource Allocation in Business Process Management and Process Mining. In Management Decision, 2018, 56; S. 376–405.

[ARD07] Aalst, W. M. van der; Rosemann, M.; Dumas, M.: Deadline-based Escalation in Process-Aware Information Systems. In Decision Support Systems, 2007; S. 492–511.

[Aw09] Awad, A. et al.: Enabling Resource Assignment Constraints in BPMN, Hasso Plattner Institute, Potsdam, 2009.

[5] Hier sind „normale“ Eskalationen gemeint, wie sie in [Ru05] beschrieben und von heutigen PMS teilweise unterstützt werden (z.B. Benachrichtigung per E-Mail, Weitergabe an andere Bearbeiter). Das Thema beinhaltet nicht solch weitgehende Aktionen wie in [ARD07] beschrieben.

[Ba09] Bauer, T.: Stellvertreterregelungen für Task-Bearbeiter in prozessorientierten Applikationen. In Datenbank-Spektrum, 2009, 9; S. 40–51.

[BB01] Breton, E.; Bézivin, J.: Model-Driven Process Engineering. In Proc. 25th Annual International Computer Software and Applications Conference, 2001; S. 225–230.

[BE01] Botha, R. A.; Eloff, J.H.P.: Separation of Duties for Access Control Enforcement in Workflow Environments. In IBM Systems Journal, 2001; S. 666–682.

[BFA99] Bertino, E.; Ferrari, E.; Atluri, V.: The Specification and Enforcement of Authorization Constraints in Workflow Management Systems. In ACM Transactions on Information and System Security, 1999; S. 65–104.

[Bi20] Bizagi: Bizagi 11.2.3 BPM Suite User Guide. http://help.bizagi.com/bpm-suite, Zugriff: 7.2.2020.

[Ca15] Cabanillas, C. et al.: RALph: A Graphical Notation for Resource Assignments in Business Processes. In Proc. Int. Conf. on Advanced Information Systems Engineering, 2015; S. 53–68.

[CRR11] Cabanillas, C.; Resinas, M.; Ruiz-Cortés, A.: RAL: A High-Level User-Oriented Resource Assignment Language for Business Processes. In Proc. Int. Conf. on Business Process Management, 2011; S. 50–61.

[CRR12a] Cabanillas, C.; Resinas, M.; Ruiz-Cortés, A.: Designing Business Processes with History-Aware Resource Assignments. In Proc. Int. Conf. on Business Process Management, 2012; S. 101–112.

[CRR12b] Cabanillas, C.; Resinas, M.; Ruiz-Cortés, A.: Automated Resource Assignment in BPMN Models Using RACI Matrices. In Proc. Int. Conf. on Cooperative Information Systems, 2012; S. 56–73.

[Cz19] Czastka, J.: Organisatorischer Aspekt von Prozess-Management-Systemen - Anforderung und Analyse kommerzieller Produkte. Bachelorarbeit, Hochschule Neu-Ulm, 2019.

[DW15] Dulai, T.; Werner-Stark, A.: A Database-Oriented Workflow Scheduler with Historical Data and Resource Substitution Possibilities. In Proc. 4th International Conference on Operations Research and Enterprise Systems, 2015; S. 325–330.

[EP99] Eder, J.; Panagos, E.: Towards Distributed Workflow Process Management. In Proc. Workshop on Cross-Organisational Workflow Management and Co-ordination, 1999.

[Gr08] Großkopf, A.: An Extended Resource Information Layer for BPMN, Hasso-Plattner-Institute for IT Systems Engineering, Potsdam, 2008.

[HD05] Hochmüller, E.; Dobrovnik, M.: Flexibility Issues in Workflow Management Systems. In Proceedings Business Process Modeling, Development and Support, 2005, 5.

[HS99] Huang, Y. N.; Shan, M. C.: Policies in a Resource Manager of Workflow Systems: Modeling, Enforcement and Management. In Proc. 15th Int. Conf. on Data Engineering, 1999.

[IBM17] IBM: Business Process Manager, 2017. https://www.ibm.com/support/knowledge center/en/SSFPJS_8.6.0, Zugriff: 7.2.2020.

[K20] K2: K2 Cloud: Low Code Digital Process Automation. https://www.k2.com/platform/k2-cloud, Zugriff: 7.2.2020.

[KLW09] Ko, R.K.L.; Lee, S.S.G.; Wah Lee, E.: Business Process Management (BPM) Standards: A Survey. In Business Process Management Journal, 2009, 15; S. 744–791.

[KR09] Künzle, V.; Reichert, M.: Integrating Users in Object-aware Process Management Systems: Issues and Challenges. In Proc. Int. Conf. on Business Process Management, 2009; S. 29–41.

[KRK15] Knuplesch, D.; Reichert, M.; Kumar, A.: Towards Visually Monitoring Multiple Perspectives of Business Process Compliance. In CAiSE Forum, CEUR Workshop Proceedings 1367, 2015.

[KRS13] Kotremba, J.; Raß, S.; Singer, R.: Distributed Business Processes - A Framework for Modeling and Execution. In CoRR, 2013.

[Li08] Link, S. et al.: Model-Driven Development of Human Tasks for Workflows. In Proc. 3rd Int. Conf. on Software Engineering Advances, 2008; S. 329–335.

[LSR14] Lawall, A.; Schaller, T.; Reichelt, D.: Enterprise Architecture: A Formalism for Modeling Organizational Structures in Information Systems. In Proc. Workshop on Enterprise and Organizational Modeling and Simulation, 2014; S. 77–95.

[Mu04] Zur Muehlen, M.: Organizational Management in Workflow Applications – Issues and Perspectives. In Information Technology and Management Journal, 2004; S. 271–291.

[NS07] Namiri, K.; Stojanovic, N.: Pattern-Based Design and Validation of Business Process Compliance. In Proc. Int. Conf. on Cooperative Information Systems, 2007; S. 59–76.

[OMG11] Object Management Group: Business Process Model and Notation (BPMN) 2.0, 2011.

[OS10] Oberweis, A.; Schuster, T.: A Meta-model based Approach to the Description of Resources and Skills. In Proc. Americas Conference on Information Systems, 2010.

[PR98] Panagos, E.; Rabinovich, M.: Reducing Escalation-Related Costs in WFMSs. In Workflow Management Systems and Interoperability, 1998; S. 107–128.

[RM98] Rosemann, M.; Zur Mühlen, M.: Modellierung der Aufbauorganisation in Workflow-Management-Systemen: Kritische Bestandsaufnahme und Gestaltungsvorschläge. In EMISA-Forum, 1998; S. 78–86.

[Ru05] Russell, N. et al.: Workflow Resource Patterns: Identification, Representation and Tool Support. In Proc. Int. Conf. on Advanced Information Systems Engineering, 2005; S. 216–232.

[SCV15] Stroppi, L. J. R.; Chiotti, O.; Villarreal, P. D.: Defining the Resource Perspective in the Development of Processes-aware Information Systems. In Information and Software Technology, 2015, 59; S. 86–108.

[Si20] Signavio: Signavio Workflow Accelerator. https://www.signavio.com/products/work flow-accelerator, Zugriff: 7.2.2020.

[SKR14] Semmelrodt, F.; Knuplesch, D.; Reichert, M.: Modeling the Resource Perspective of Business Process Compliance Rules with the Extended Compliance Rule Graph. In Proc. 15th Int. Working Conf. on Business Process Modeling, Development, and Support, 2014; S. 48–63.

[SM11] Strembeck, M.; Mendling, J.: Modeling Process-related RBAC Models with Extended UML Activity Models. In Information and Software Technology, 2011, 53; S. 456–483.

[SRS08] Stefansen, C.; Rajamani, S.; Seshan, P.: SOFTALLOC: A Work Allocation Language with Soft Constraints. In Prov. IEEE Int. Conf. on Web Services, 2008; S. 441–448.

[WS07] Wolter, C.; Schaad, A.: Modeling of Task-based Authorization Constraints in BPMN. In Proc. Int. Conf. on Business Process Management, 2007; S. 64–79.

R. Reussner, A. Koziolek, R. Heinrich (Hrsg.): INFORMATIK 2020,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2021 621

Application Fields and Research Gaps of Process Mining in Manufacturing Companies

Simon Dreher[1] Peter Reimann[1,2] Christoph Gröger[2]

Abstract: To survive in global competition with increasing cost pressure, manufacturing companies must continuously optimize their manufacturing-related processes. Thereby, process mining constitutes an important data-driven approach to gain a profound understanding of the actual processes and to identify optimization potentials by applying data mining and machine learning techniques on event data. However, there is little knowledge about the feasibility and usefulness of process mining specifically in manufacturing companies. Hence, this paper provides an overview of potential applications of process mining for the analysis of manufacturing-related processes. We conduct a systematic literature review, classify relevant articles according to the Supply-Chain-Operations-Reference-Model (SCOR-model), identify research gaps, such as domain-specific challenges regarding unstructured, cascaded and non-linear processes or heterogeneous data sources, and give practitioners inspiration which manufacturing-related processes can be analyzed by process mining techniques.

Keywords: Process Mining; Application; Production; Manufacturing; SCOR; Literature Review

1 Introduction

Manufacturing companies are facing global competition and increasing cost pressure due to new competitors in the proceeding globalization of markets. In order to ensure future competitiveness, companies need to optimize their costly manufacturing-related processes regarding effectiveness and efficiency [BPR16]. This process optimization first requires a profound understanding of one's own manufacturing processes. However, this knowledge and transparency is often not sufficiently available in companies as defined process models in manufacturing often only represent an idealized image of reality or are not transparent [EAW15]. Hence, process owners, e.g., production managers, cannot implement process optimizations to be able to meet targets set by the management of the company [EAP15]. Companies are increasingly using information systems in manufacturing such as Enterprise Resource Planning (ERP) or Manufacturing Execution Systems (MES) to plan and control processes and resources. As a result, large amounts of process-related data are collected and stored in database systems, data warehouses, or data lakes [Gr16]. However, data are available in such large quantities that the results of conventional methods (e.g., Reporting, Online Analytical Processing (OLAP), Value Stream Mapping), which work with aggregated

[1] Graduate School of Excellence advanced Manufacturing Engineering, University of Stuttgart, Nobelstr. 12, 70569 Stuttgart, Germany, simon.dreher@gsame.uni-stuttgart.de, peter.reimann@gsame.uni-stuttgart.de
[2] Institute for Parallel and Distributed Systems, University of Stuttgart, Universitätsstraße 38, 70569 Stuttgart, Germany, christoph.groeger@ipvs.uni-stuttgart.de

doi:10.18420/inf2020_55

data, are often not as detailed and precise as necessary. Hence, these methods are not able to provide the highly needed process knowledge and transparency [VW04, Va16] or require a high amount of manual effort for data integration and analysis [KRP19].

In this context, process mining has developed in recent years and has established itself as an independent research discipline [Va16]. By a process-oriented view on raw data, process mining constitutes an important approach to gain a profound understanding of the actual process execution. This process-oriented view is created by applying data mining and machine learning techniques on event data. Using process mining to explore the event logs related to manufacturing processes is a promising way to gain the necessary process knowledge and transparency in order to pave the way for process optimizations and future competitiveness [Va16]. In recent years, the number of publications on process mining in the academic field has increased significantly. This is underpinned by literature reviews on the application of process mining in various domains [Da18, Ga19]. These especially show that process mining has so far been less researched for the manufacturing industry compared to other domains. However, the ongoing digitalization in line with Industry 4.0 has significantly improved the data basis in manufacturing and makes this domain along with its process-oriented characteristics predestined for the use of process mining.

Hence, this paper provides a systematic literature review to contribute to the stream of literature focusing on the domain-specific application of process mining. We survey various use cases for analyzing manufacturing-related processes using process mining and identify research gaps for future directions in this relatively new field of study. Note that we do not want to make a statement about the general importance of process mining for manufacturing and especially not about its importance compared to other, specific analytical tools from manufacturing literature, such as lean production methods or value stream mapping. Therefore, our contribution is intended mainly as an analysis of the application of process mining in manufacturing and of associated research gaps. In order to classify the identified use cases, we propose the SCOR-model [HSW04]. By an operational process perspective, the SCOR-model integrates concepts of business process re-engineering, benchmarking and process measurement into one framework. It has become the de-facto standard for defining process types in operations management [HSW04] and is consequently adopted for this review. As the SCOR-model is implemented in many companies, our classification of use cases gives practitioners inspiration which manufacturing-related processes can be analyzed by process mining techniques.

The remainder of the paper is structured as follows: Section 2 describes theoretical background. In Section 3, the literature selection process is clarified. Section 4 outlines the identified use cases, while Section 5 discusses research gaps. Section 6 summarizes with a conclusion.

2 Theoretical Basis

This section outlines the fundamental basics of the SCOR-model and process mining.

2.1 SCOR-model

The SCOR-model was first developed by the Supply Chain Council in 1996 and has been revised continuously [HSW04, Zh11]. The SCOR-model basically differentiates six level 1 process types: Plan, Source, Make, Deliver, Return and Enable [HSW04]. These process types represent potential application fields of process mining for our literature review. So, we give a brief overview on these process types in the following. The Plan process contains all activities that balance demand and supply in order to develop a course of action which best meets sourcing, manufacturing and delivery requirements [Zh11]. The Source process describes activities that procure and issue materials needed to produce the planned demand. The subsequent Make process includes all activities that transform material to finished products and is considered to be the core process of the model [Zh11]. The activities of the Deliver process provide the finished products to retailers and/or end-consumers [HSW04]. The Return process encompasses activities managing the reverse flow of used products and materials back to the manufacturing company [Zh11]. The Enable processes support the realization and governance of the other process types. Hence, they interact with the HR, IT and Financial department [HSW04].

2.2 Process Mining

According to van der Aalst, „the idea of process mining is to discover, monitor and improve real processes [. . .] by extracting knowledge from event logs readily available in today´s systems.“ [Va16] It constitutes an important data-driven approach applying data analytics and machine learning on event data. Process mining can be distinguished from traditional business process management (BPM) by the fact that BPM methods usually allow for identifying a process model by means of expert interviews, and not based on data. Nevertheless, the application of process mining methods and BPM is not an either/or decision, but process mining establishes a connection between data science and BPM or process science [Va16]. Process mining can basically be applied for all kinds of processes. In doing so, data from one or more IT systems (e.g., ERP, MES), is extracted into an event log representing the history of process executions [Va16]. To analyze the event log data, various algorithms can be used [Va16]: e.g., α-algorithm, heuristic miner, genetic miner, inductive miner and the fuzzy miner. However, the α-algorithm has a lot of shortcomings, like problems with noise or complex routing constructs. It is therefore not seen as one of the main algorithms used to analyze event data [Va16]. Instead, the inductive miner is mainly used and seems to provide the best results [NE19, Va16]. The first academically developed process mining software tool “ProM” was introduced in 2005 and has been continuously

developed since then [Va16]. In the past years, commercial software vendors added process mining functionality to their tools as well [Va12a]. According to van der Aalst [Va16], three categories of process mining may be distinguished: process discovery, process conformance and process enhancement. The aim of process discovery is to convert event log data into an initial process model describing how actual processes have been executed [Va12b]. So, this initial process model represents the order of individual process steps, including possible branches and loops between the steps. The aim of process conformance is to compare the actual process execution with a predefined process model in order to check whether defined process steps are carried out properly [Va12b]. Finally, the goal of process enhancement is to extend a previously defined process model or to optimize the process, e.g., with regard to bottlenecks and resource utilization [Va12b].

3 Methodology

To provide an overview of applications of process mining in the manufacturing field, we conducted a systematic literature review following the steps proposed by Thomé et al. [TSS16]. We defined keywords according to the main objective of the paper and then combined these keywords in the search string ("process mining" AND ("manufacturing" OR "production")). We then used this search string to identify relevant articles and conference papers in six different databases: SpringerLink, IEEE Xplore, ScienceDirect, Web of Science, Emerald Insight and Academic and Business Source Premier. We retrieved 361 articles by searching titles, abstracts and keywords of all publications within these databases with the search string. Subsequently, we removed duplicates (to reject 53 articles) and then screened titles, abstracts (to reject 281 articles) and finally also full texts (to reject 4 articles) based on the following inclusion and exclusion criteria.

Inclusion criteria:

- Full text of the paper is electronically available
- Paper focuses on the application of process mining for process analysis by conducting a case study

Exclusion criteria:

- Paper is not written in English or German
- Paper is published outside peer-reviewed journals or conference proceedings
- Paper references process mining solely in its introduction or only as future research directions
- Paper is not focused on process mining, e.g., mining metal

Through this procedure, we identified 23 relevant articles and conference papers.

Note that not every paper that analyzes process mining explicitly uses the keywords we used in our search string mentioned above. So, there might be a few relevant articles that we possibly did not identify, as they use other related keywords. The keyword "process mining" also limits the covered time period, since this term for process-oriented mining approaches had not been used until van der Aalst coined this term in 1998 [VW04]. Nevertheless, we decided to use the keyword "process mining" in the search string, since process mining has meanwhile established itself as an independent research stream. So, the keyword "process mining" is sufficient to identify a vast amount and also the most important papers regarding the application of process mining in manufacturing.

4 Results

The distribution of the publications over years ranges from 2009 to 2020 and shows a strong increase in 2017. This supports the proposition that digitalization and Industry 4.0 improved the data basis in manufacturing in recent years making the use of data-driven approaches like process mining valuable. In the following, we classify the identified literature that represent potential application fields for process mining according to the six process types of the SCOR-model. Table 1 provides an overview. As the Make process is considered to be the core process, most studies have been identified here. No articles can be assigned to the Return process.

Tab. 1: Literature assigned to the SCOR-process-types and to the process mining categories

	Process Discovery	**Process Conformance**	**Process Enhancement**
Plan	[Er18], [NE19]	[Ji18]	[Ji18]
Source	[BLP17], [RC09], [RC17]	[EAP15], [EAW15]	[EAW15], [RC09], [RC17]
Make	[BPR16], [DSK17], [IB18], [Me17], [NWD19], [RC09], [Ro09], [RAB18], [Ru18], [TS16]	[Na17], [Pi17], [RAB18], [Ro09]	[AB20], [Pa15], [RAB18], [RC09], [Ro09]
Deliver	[RC09]	[Pa13]	[RC09]
Return	-	-	-
Enable	[Ro19]	-	-

Process mining in the Plan process

Three studies can be assigned to the Plan process [Er18, Ji18, NE19]. In order to efficiently generate manufacturing plans, companies make use of IT systems (e.g., ERP systems), which create corresponding planning drafts based on capacities and demand. High uncertainty

about the future forces companies to frequently modify the plans [Mu06]. This so-called "nervousness syndrome" causes the "bullwhip effect" in supply chains, which has been analyzed predominantly using system dynamics or mathematical models [Mu06]. Er et al. aim to investigate the effects of this syndrome on process level by applying process discovery on ERP data [Er18]. The results show amongst others that the use case company changed 31% of its plans. Thereby, the "nervousness syndrome" could be proven at process level by using process mining [Er18]. As manufacturing plans are generated for different purposes, planning processes also show diverse characteristics. Consequently, each kind of planning process may require the usage of different existing process mining algorithms. Nuritha/Er [NE19] apply five algorithms to the use cases of two different planning processes. By extracting ERP data for each process, Nuritha/Er suggest due to their performance comparison that the genetic and inductive miner are suitable for the considered use cases of planning-to-stock processes, and that the inductive miner is most appropriate for the use case of planning-to-export processes. However, this claim also needs further validation [NE19]. Since the concepts of self-organizing manufacturing systems indicate reductions in planning efforts, little knowledge exists on the quality of the self-planned and self-executed manufacturing processes [Ji18]. Jimenez et al. [Ji18] apply conformance checking and enhancement for the diagnosis of self-organizing manufacturing systems. They show improvements in the function of those systems by permitting efficient and smooth reactions to perturbation events. So, process mining can support evaluations of plans generated by self-organizing manufacturing systems [Ji18].

Process mining in the Source process

Overall, four studies may be assigned to the Source process [RC17, EAP15, EAW15, BLP16]. R'bigui/Cho [RC17] demonstrate the usability of process discovery and enhancement for sourcing processes in heavy manufacturing industries. Er et al. [EAP15] address the problem of disruptions and downtimes in manufacturing due to defected or missing material. They apply conformance checking to the process handling incoming material at warehouses based on ERP data. The results show that the actual process highly conforms to the predefined standard. However, in case of failed quality checks, additional manual tests need to be done delaying the entire process. The analysis also found the single sourcing strategy to be critical [EAP15]. The storing and issuing of material to the manufacturing line is considered as one of the critical processes in sourcing. Er et al. [EAW15] analyze material movements in warehouses to carry out process conformance checking and enhancement. The conformity check indicates that additional, previously not defined process steps are carried out, while others are skipped presumably for time reasons. Some materials cause disruptions due to quality defects, even though the quality was checked before. The analysis reveals that materials stored in high racks are particularly prone to quality issues. Hence, this process step can cause damage and has to be optimized. Finally, a dotted chart analysis of individual process steps states that the First-In-First-Out (FIFO) rule is not maintained [EAW15]. As sourcing processes are characterized by high frequency of changes and fluctuations, defined process models need to be updated continuously. Becker et al. [BLP17]

propose a process maintenance concept based on process mining. Process mining acts as an enabler to automatically create process models and this way replaces manual paper and pen modeling. However, challenges exist as standard software cannot handle the heterogeneous data sources [BLP17].

Process mining in the Make process

Thirteen and hence the most amount of articles may be assigned to the Make process [Ro09, Pa15, BPR16, TS16, DSK17, Na17, Me17, Pi17, IB18, RAB18, Ru18, NWD19, AB20]. Rozinat et al. [Ro09] show one of the first applications of all three process mining categories by analyzing the test phase of wafer steppers. Park et al. [Pa15] demonstrate the use of process enhancement for workload and delay analysis in make-to-order manufacturing. Dišek et al. [DSK17] discover the control flow and other parameters of the manufacturing process of transmission parts in the automotive industry. Ribeiro et al. [RAB18] use process conformance and enhancement to analyze unsatisfactory performance levels approached in the referred process. Based on the analysis results, they propose adjustments to the "L*life-cycle model" methodology developed by van der Aalst [Va16]. Bettacchi et al. [BPR16] compare the application of five algorithms on an interlinked manufacturing process. They show that the inductive miner is best-suited for process discovery in their use case. Semi- and unstructured processes are addressed by Meincheim et al. [Me17], who combine the inductive miner with trace clustering to discover process variants based on MES data. This approach seems promising for analyzing performance issues in unstructured processes. Intayoad/Becker [IB18] and Natschläger et al. [Na17] show that unstructured processes are often accompanied with heterogeneous data sources. Hence, Intayoad/Becker apply a Markov chain as a sequence clustering technique for the data processing steps based on MES data [IB18]. A comparison of the results with and without Markov chain shows an improvement of the discovered process model quality by the indicator of replay fitness. They conclude that involving experts with domain knowledge is crucial for successful process mining in manufacturing [IB18]. Natschläger et al. check the conformance of two manufacturing processes using process mining based on ERP data [Na17]. They propose a new procedure to extract, load and transform data from heterogeneous sources, which can also be applied in other application domains [Na17].

Most mentioned studies use ERP data. In contrast, Altan/Birgün [AB20] apply process enhancement to the manufacturing process of propeller shafts using machinery data. They use process mining to evaluate lead times and to improve the machining process [AB20]. Others studied the use of process mining for Make support processes and side aspects. Tu/Song [TS16] propose a concept to analyze and predict manufacturing process costs based on process enhancement. Therefore, they extend the event log by the costs of each activity in a process. They demonstrate the concept by analyzing a manufacturing processes of jeans [TS16]. Pika et al. [Pi17] provide insights on the usability of process mining for checking conformance of safety processes in the manufacturing area [Pi17]. Ruschel et al. [Ru18] apply process discovery for maintenance inspections of machines and equipment providing better support to managers in scheduling activities [Ru18]. Nagy et al. [Na17] use

real-time process mining to detect early deviations in a manufacturing process. They show that real-time process mining can speed up the process of detecting potential sources of defects and thereby reduce the number of faulty products.

Process mining in other process types

The study of Paszkiewicz [Pa13] may be assigned to the Deliver process. Paszkiewicz uses conformance checking to analyze the outbound logistic process of delivering products to the company´s customer. The results show that rules, e.g., FIFO, are not obeyed by employees, and further analyzes indicate an ineffective configuration of the warehouse [Pa13]. The study of Roldán et al. [Ro19] can be assigned to the Enable process. The authors investigate how a training system for industrial operators in assembly tasks may benefit from process discovery and virtual reality. Their results show that the automatically retrieved process models can help in teaching new employed machine operators in an efficient way [Ro19]. Finally, R'bigui/Cho [RC09] examine a customer order fulfillment process of a heavy manufacturing company using process discovery and enhancement. This study spans several process types (Source, Make and Deliver processes) and, thereby, constitutes the only study, which analyzes an end-to-end process.

5 Research Gaps

The conducted literature review surfaces a variety of process mining applications in manufacturing companies. However, there still exist research gaps, which we derived from the results of our study and which we explain in the following.

Consideration of process types and mining categories

As shown in Table 1, most studies explore the process type Make, while the Plan and Source processes receive some attention. However, the Deliver and Enable processes are considered too rarely so far, while the Return process is even not covered at all. This Return process is however of particular interest, since it is becoming crucial in the course of an increasing circular economy in manufacturing. So, one research gap is to identify and analyze additional possibilities to use process mining for Deliver, Enable, and especially Return processes. Furthermore, only one study analyzes end-to-end processes [RC09]. As the optimization of end-to-end processes has a greater impact on efficiency improvements, future research should investigate the usefulness of process mining on a larger scale for these end-to-end processes. Moreover, the results show that most studies aim to discover process models (14 of 23 articles), while less studies check conformance (8 of 23 articles) or enhance (8 of 23 articles) existing models. Only 6 of 23 articles combine those approaches [EAW15, Ji18, RAB18, RC09, RC17, Ro09]. Yet, conformance checks and process enhancement are of utmost importance for companies in order to keep their manufacturing processes compliant and to optimize them. Hence, future research should especially focus on conformance checking and process enhancement.

Selection of process mining algorithms for given use cases

There is little knowledge on how to select the right process mining algorithm for a given manufacturing use case [NE19]. For instance, it would be valuable to know which process mining algorithm is suitable for each SCOR-process-type. Thereby, future work is to develop a selection framework which assigns algorithms to a SCOR-process-type or to concrete use cases. The studies of Bettacchi et al. and Nuritha/Er indicate that the inductive miner shows good performance when applying discovery techniques to certain use cases of a Make process [BPR16, NE19]. However, other algorithms may recognize additional properties of the processes that go beyond the pure workflow of the process model and that are of relevance as well, e.g., properties describing which different organizational units and which machine resources are involved in a manufacturing process. Further research is needed to validate these first assumptions, to provide more robust findings and to provide a selection framework for process mining algorithms. The more general framework to guide the selection and configuration of machine learning solutions in manufacturing proposed by Villanueva et al. may be a starting point for this research [VRM18]. Nevertheless, it has to be adapted to the specific needs of process mining.

Unstructured, cascaded, and non-linear processes

Since each application domain has its own characteristics, this also comes along with specific application problems that only occur in this domain [NE19]. However, there is little knowledge on resulting domain- and application-specific challenges. Only very few studies have addressed some manufacturing-specific challenges such as unstructured, cascaded, and non-linear manufacturing processes [BLP17, EAP15, EAW15, IB18, Ro19]. For instance, in case a product does not pass a test in a quality control gate, the processing order of this product does not follow the predefined order of the manufacturing process. The processes then become non-linear [Ch17, Wi20]. Here, the processing order typically includes loops from one step back to a preceding step, or even self-loops within a single manufacturing step. This means that data samples of a specific product occur multiple times in the related data set, but with different timestamps for one and the same process step. This makes it challenging to clearly associate the process instances identified by process mining with the real process steps and products. Although such unstructured, cascaded, and non-linear processes are typical for the manufacturing domain, no adequate solution exists to obtain the event log [EAP15, EAW15, Ro19] and to avoid “spaghetti-like” process models as a result [IB18, BLP17]. Thereby, it can be difficult to differentiate between expected and undesired process outcomes. Also, there is little knowledge how discovery techniques can deal with varying processes, i.e., where process steps may vary depending on the product variant being produced. The product variety that is inherent in manufacturing also increases the variety of underlying event data. It leads to complex data relations, to high data dimensionalities, and to a complicated interpretability of event data [Wi20, Wu16]. Hence, Intayoad/Becker mention that experts with domain knowledge have to be properly involved when planning an application project in manufacturing [IB18].

Integration and fusion of heterogenous data sources

Another main challenge for implementing process mining in manufacturing is a huge disconnect between physical flow of materials and the digital information flow. As complex, unstructured, and non-linear processes are often accompanied with heterogeneous data sources and IT systems, data quality and data integration of various data sources are of great importance in manufacturing [BLP17, NA17]. Here, a problem is the assumption of process mining that a process instance and its event data can be traced and clearly assigned through the entire process. In manufacturing, however, this can often only be the case over a limited segment of the process, e.g., over a single manufacturing line. A process change destroying the uniqueness of a process instance, can cause problems for process mining applications. This problem is even intensified by the unstructured, cascading, and non-linear nature of manufacturing process, as well as by the high variety of underlying products mentioned above.

Furthermore, the increasing use of sensor technology in the shop floor area provides new data sources for process analysis [AB20]. As a result of Industry 4.0 and the ongoing digitization, more and more sensors are being installed on machines and measuring stations [Gr16]. Hence, the generated data from those sensors are highly relevant for process mining in manufacturing as they contain additional and unique information. However, the integration of sensor data with machinery or ERP data sources is challenging, as sensor data are usually not linked to discrete process steps. In fact, no explicit or only incomplete key relationships (e.g., primary and foreign key relations) exist between both data sources. This makes the integration of such data sources difficult and can significantly increase the effort in terms of time and resources.

Hence, significant need for research exists especially for integrating and fusing heterogeneous data sources in process mining [BLP17, IB18, Na17]. This is underpinned by the fact that almost all studies apply process mining exclusively on ERP or MES data (except for [AB20]). So, they do not integrate data from various sources into one transformed event log. However, a combination of ERP, machinery and sensor data seems promising as this enables both an overall view on and a deep dive into end-to-end processes [Gr16]. Hence, existing process mining tools need to be enhanced with data integration features or be combined with other tools that deliver these features (e.g., Talend Open Studio for Data Integration).

Inter-organizational process mining

As manufacturing companies are interconnected in global supply chains, more and more of such inter-organizational processes are supported by IT systems. However, all identified studies analyze intra-organizational manufacturing processes and only use data from one company. Hence, the application of process mining for analysis of such inter-organizational processes should be in the scope of further studies [EA14]. However, this endeavor will most likely face a lot of problems. Not only heterogenous data sources within one company, but multiple and even more diverse data sources from other involved companies need

to be integrated into consistent event logs. This will intensify already existing problems. Furthermore, it requires the involved parties to share confidential data. Nevertheless, this seems to be a promising way to optimize inter-organizational manufacturing processes, reducing the often-cited bullwhip effect in supply chains.

6 Conclusion

Process mining constitutes an important data-driven approach to gain a profound process knowledge and to pave the way for process optimizations. Process mining in manufacturing constitutes a novel and thus rather unexplored research field. Therefore, this paper provides an overview of various use cases for analyzing manufacturing-related processes. Our SCOR-based classification of the use cases gives an outline where process mining can be an alternative to existing tools. It especially helps practitioners identify suitable use cases for process mining, as the SCOR-model is implemented in many companies. Furthermore, we identify research gaps that need to be filled to ensure a broad adoption of process mining in the manufacturing domain. These research gaps are summarized in Table 2.

Tab. 2: Research agenda summarizing the identified research gaps for the application of process mining in manufacturing

Research Gap	Description
Consideration of process types and mining categories	Stronger focus on Deliver, Return and Enable process types, as well as on the mining categories conformance checking and process enhancement
Selection of algorithms for given use cases	Develop a selection framework which assigns algorithms to a process type or to concrete use cases
Handling of unstructured, cascaded, and non-linear processes	Develop adequate solutions to obtain event logs from such complex and dynamic processes and to prevent “spaghetti-like” process models
Integration and fusion of heterogenous data sources	Implement appropriate methods for data integration and fusion
Inter-organizational process mining	Examine and validate the use of process mining for inter-organizational process analysis

Here, we see addressing unstructured, cascaded, and non-linear processes as well as the integration and fusion of heterogenous data sources as most important. Those challenges are present in almost every manufacturing environment [Wi20] and are also linked to other gaps, e.g., to inter-organizational process mining. Also, the development of a use-case-related selection framework of algorithms seems valuable, especially for practitioners.

Bibliography

[AB20] Altan, Z.; Birgün, S.: Using Process Mining Approach for Machining Operations. In: (Durakbasa, N. M.; Gençyılmaz, M. G. Hrsg.): Proceedings of the International Symposium for Production Research, Springer, Cham, S. 452-464, 2020.

[BLP17] Becker, T.; Lütjen, M.; Porzel, R.: Process Maintenance of Heterogeneous Logistic Systems – A Process Mining Approach. In (Freitag, M.; Kotzab, H.; Pannek, J. Hrsg.): Dynamics in Logistics, Springer, Cham, S. 77–86, 2017.

[BPR16] Bettacchi, A.; Polzonetti, A.; Re, B.: Understanding Production Chain Business Process Using Process Mining: A Case Study in the Manufacturing Scenario. In (Krogstie, J.; Mouratidis, H.; Su, J. Hrsg.): Proc. 26th Int. Conf. on Advanced Information Systems Engineering CAiSE 2016, Springer, Cham, S. 193-203, 2016.

[Ch17] Cheng, Y. et al.: Data and Knowledge Mining with Big Data Towards Smart Production. Journal of Industrial Information Integration, 9/17, S. 1-13, 2017.

[Da18] Dakic, D. et al.: Business Process Mining Application: A Literature Review. In (Katalinic, B. Hrsg.): Proc. 29th Int. Symposium, DAAAM Int., S. 866-875, 2018.

[DSK17] Dišek, M.; Šperka, R.; Kolesár, J.: Conversion of Real Data from Production Process of Automotive Company for Process Mining Analysis. In (Jezic, G. et al. Hrsg.): Agents and Multi-Agent Systems Technologies and Applications, Springer, Cham, S. 223-233, 2017.

[EA14] Er, M.; Astuti, H. M.: A Case Study on Process Mining Implementation in Modeling Supply Chain Business Process – A Lesson Learnt. In (Pujawan, I. N.; Vanany, I.; Baihaqi, I. Hrsg.): Proc. 6th Int. Conf. on Operations and Supply Chain Management, Department of Industrial Engineering, Institut Teknologi Sepuluh Nopember (ITS), Indonesia, S.808-819, 2014.

[EAP15] Er, M.; Astuti, H. M.; Pramitasari, D.: Modeling and Analysis of Incoming Raw Materials Business Process: A Process Mining Approach. International Journal of Computer and Communication Engineering, 4/15, S. 196-203, 2015.

[EAW15] Er, M.; Astuti, H. M.; Wardhani, I. R. K.: Material Movement Analysis for Warehouse Business Process Improvement with Process Mining: A Case Study. In (Bae, J.; Suriadi, S.; Wen, L. Hrsg.): Asia Pacific Business Process Management, Springer, Cham, S.115-127, 2015.

[Er18] Er, M. et al.: Analysis of Production Planning in a Global Manufacturing Company with Process Mining. Journal of Enterprise Information Management, 31/18, S. 317-337, 2018.

[Ga19] Garcia, C. et al.: Process Mining Techniques and Applications – A Systematic Mapping Study. Expert Systems with Applications, 133/19, S. 260-295, 2019.

[Gr16] Gröger, C. et al.: A Mobile Dashboard for Analytics-based Information Provisioning on the Shop Floor. International Journal of Computer Integrated Manufacturing 29/16, S. 1335-1354, 2016.

[HSW04] Huan, S. H.; Sheoran, S. K.; Wang, G.: A Review and Analysis of Supply Chain Operations Reference (SCOR) Model. Supply Chain Management: An International Journal, 9/04, S. 23-29, 2004.

[IB18] Intayoad, W.; Becker, T.: Applying Process Mining in Manufacturing and Logistic for Large Transaction Data. In (Freitag, M.; Kotzab, H.; Pannek, J. Hrsg.): Dynamics in Logistics, Springer, Cham, S. 378-388, 2017.

[Ji18] Jimenez, J.-F. et al.: Using Process-mining for Understating the Emergence of Self-organizing Manufacturing Systems. IFAC-PapersOnLine 51/18, S. 1618-1623, 2018.

[KRP19] Knoll, D.; Reinhart, G.; Prüglmeier, M.: Enabling Value Stream Mapping for Internal Logistics Using Multidimensional Process Mining. Expert Systems with Applications 124/19, S. 130-142, 2019.

[Me17] Meincheim, A. et al.: Combining Process Mining with Trace Clustering: Manufacturing Shop Floor Process - An Applied Case. In (IEEE Hrsg.): Proc. of the 29th IEEE Int. Conf. on Tools with Artificial Intelligence, IEEE, S. 498–505, 2017.

[Mu06] Mula, J. et al.: Models for Production Planning Under Uncertainty: A review. International Journal of Production Economics, 103/06, S. 271-285, 2006.

[NWD19] Nagy, Z.; Werner-Stark, A.; Dulai, T.: Using Process Mining in Real-Time to Reduce the Number of Faulty Products. In (Welzer, T. et al. Hrsg.): Advances in Databases and Information Systems, Springer, Cham, S.89-104, 2019.

[Na17] Natschläger, C. et al.: A Practical Approach for Process Mining in Production Processes. In (Piazolo, F. et al. Hrsg.): Innovations in Enterprise Information Systems Management and Engineering, Springer, Cham, S. 87-95, 2017.

[NE19] Nuritha, I.; Er, M.: Behavioural Similarity Measurement of Business Process Model to Compare Process Discovery Algorithms Performance in Dealing with Noisy Event Log. In (Samsonovich, A. V.; Klimov, V. V. Hrsg.): Procedia Computer Science Pro. of the 5th Information Systems Int. Conf., S. 984–993, 2019.

[Pa15] Park, M. et al.: Workload and Delay Analysis in Manufacturing Process Using Process Mining. In: Bae, J., Suriadi, S., Wen, L. (Hg.): In (Bae, J.; Suriadi, S.; Wen, L. Hrsg.): Asia Pacific Business Process Management, Springer, Cham, S.138-151, 2015.

[Pa13] Paszkiewicz, Z.: Process Mining Techniques in Conformance Testing of Inventory Processes: An Industrial Application. In (Abramowicz, W. Hrsg.): Business Information Systems Workshops, Springer, Cham, S.302-313, 2013.

[Pi17] Pika, A. et al.: Analysing an Industrial Safety Process Through Process Mining: A Case Study. In (Mathew, L. et al. Hrsg.): Asset Intelligence through Integration, Springer, Cham, S. 491-500, 2017.

[RC17] Rbigui, H.; Cho, C.: Purchasing Process Analysis with Process Mining of a Heavy Manufacturing Industry. In (IEEE Hrsg.): Int. Conf. on Information and Communication Technology Convergence, IEEE, S. 495–498, 2017.

[RC09] R'bigui, H.; Cho, C.: Customer Oder Fulfillment Process Analysis with Process Mining. In (Association for Computing Machinery Hrsg.): Proc. of the 15th Int. Conf. on Mechatronics and Machine Vision in Practice, New York, S. 247–252, 2009.

[RAB18] Ribeiro, R.; Analide, C.; Belo, O.: Improving Productive Processes Using a Process Mining Approach. In (Rocha, A. et al. Hrag.): Trends and Advances in Information Systems and Technologies, Springer, Cham, S. 736-745, 2018.

[Ro19] Roldán, J. J. et al.: A Training System for Industry 4.0 Operators in Complex Assemblies Based on Virtual Reality and Process Mining. Robotics and Computer-Integrated Manufacturing, 59/19, S. 305-316, 2019.

[Ro09] Rozinat, A. et al.: Process Mining Applied to the Test Process of Wafer Scanners in ASML. IEEE Transactions on Systems, Man and Cybernetics, 39/09, S. 474-479, 2009.

[Ru18] Ruschel, E. et al.: Establishment of Maintenance Inspection Intervals: An Application of Process Mining Techniques in Manufacturing. Journal of Intelligent Manufacturing 467/18, S. 53-72 , 2018.

[TSS16] Thomé, A. M. T.; Scavarda, L. F.; Scavarda, A. J.: Conducting Systematic Literature Review in Operations Management. Production Planning & Control, 27/16, S. 408-420, 2016.

[TS16] Tu, T. B. H.; Song, M.: Analysis and Prediction Cost of Manufacturing Process Based on Process Mining. In (ICIMSA Hrsg): ICIMSA Int. Conf. on Industrial Engineering, Management Science and Applications, IEEE, S. 1-5, 2016.

[Va12a] van der Aalst, W. et al.: Process Mining Manifesto. In (Daniel, F.; Barkaoui, K.; Dustdar, S. Hrsg): Business Process Management Workshops, Springer, New York, S. 169-194, 2012.

[Va12b] van der Aalst, W.: Process Mining: Overview and Opportunities. ACM Transactions on Management Information Systems, 3/12, S. 1-17, 2012.

[Va16] van der Aalst, W.: Process Mining. Springer, Berlin, 2016.

[VRM18] Villanueva Zacarias, A.; Reimann, P.; Mitschang, B.: A Framework to Guide the Selection and Configuration of Machine-Learning-based Data Analytics Solutions in Manufacturing. In (Wang, L. Hrsg.): Proc. of the 51st CIRP Conference on Manufacturing Systems, Elsevier, S. 153-158, 2018.

[VW04] van der Aalst, W., Weijters, A.: Process mining: A Research Agenda. Computers in Industry, 53/04, S. 231-244, 2004.

[Wi20] Wilhelm, Y. et al.: Data Science Approaches to Quality Control in Manufacturing: A Review of Problems, Challenges and Architecture. In: Springer Proceedings Series Communications in Computer and Information Science (CCIS), 2020.

[Wu16] Wuest, T. et al.: Machine Learning in Manufacturing: Advantages, Challenges, and Applications. Production & Manufacturing Research, 4/16, S. 23-45, 2016.

[Zh11] Zhou, H. et al.: Supply Chain Integration and the SCOR Model. Journal of Business Logistics, 32/11, S. 332-344, 2011.

The Effect of Process Length on Process Acceptance

Lars Drewes[1] Volker Nissen[2]

Abstract: According to the process acceptance theory, the acceptance of processes can have an influence on their correct execution [Mü19]. If deviations and manipulations of processes of any kind are to be excluded, it is necessary to understand which factors lead to a change in acceptance. In the present paper, a large scale experimental study is carried out to investigate the influence of the number of process activities and the process throughput time on acceptance. A generic purchasing process is implemented online and executed in variants by the participants of the study. Process acceptance is measured using a questionnaire in three dimensions of attitude (cognitive, affective, conative). The analysis demonstrates that there is a significant difference in acceptance with regard to varying process throughput time as a measure of process length.

Keywords: BPM; Process; Acceptance; Experiment; Amazon's Mechanical Turk

1 Introduction

The functional orientation that prevailed in companies for decades has meanwhile often been replaced by process orientation to improve enterprise performance and competitiveness. However, good process design and active process management are essential prerequisites for achieving process innovation, cost reduction, and customer satisfaction [BK12], [HC06], [MN14], [Mü15]. Moreover, processes must actually be carried out as intended in the design. The decision to initiate and execute a process as required is based on the acceptance that the process executor (acceptance subject) attributes to the process (acceptance object) within a certain (acceptance) context. If this acceptance does not exist or is too low, modifications of that process occur or even a refusal of the execution takes place. It is also possible that unofficial processes, so-called shadow processes, arise as a result or that similar, already existing substitutional processes experience a much higher acceptance and for this reason are preferred [MN14], [Mü15].

Such process deviations can not only result in business disadvantages, such as loss of customer satisfaction and thus loss of sales but can also pose a potential risk to people and the environment. One example is the persistent outbreak of MRSA bacteria in hospitals, which is due to the lack of attention paid to hygiene processes. A deviation into illegality is also possible if, for example, embargo processes are circumvented, which may result in severe penalties [Mi08], [MN14], [Mü15].

[1] Technische Universität Ilmenau, Wirtschaftsinformatik für Dienstleistungen, 98693 Ilmenau, lars.drewes@tu-ilmenau.de
[2] Technische Universität Ilmenau, Wirtschaftsinformatik für Dienstleistungen, 98693 Ilmenau, volker.nissen@tu-ilmenau.de

doi:10.18420/inf2020_56

Based on the above, it is of great relevance to investigate and understand influencing factors of process acceptance. Müllerleile [Mü19] established a basic theoretical concept on the subject of process acceptance along with it's measurement. Furthermore, he identified influencing factors on acceptance by means of a qualitative study. However, in order to understand the mechanisms of process acceptance more precisely and to be able to derive optimization potentials and actions, further investigations are necessary to determine the effect strength of individual factors as well as potential interactions of the aforementioned factors. To this end, the present paper starts by investigating process length as a potential key factor of influence, taking into account the following research questions:

1. How can the variable process length be operationalized?
2. Which effects on process acceptance can be measured and which interactions occur?

2 Method

To investigate the effects and interactions of process length on process acceptance, an online experiment is developed to test a range of difference hypotheses. More specifically, an executable purchasing process was implemented which is accessible on a server. The participants actively and directly experience the process as if it would be delivered in a live business environment. The crowdsourcing platform Amazon's Mechanical Turk (MTurk) is used for the acquisition of participants. This platform allows conducting a cost-efficient study with a large sample size.

Before using MTurk, the basic comparability of the results generated there with those from a more conventional empirical field study should be checked. Such a comparability was previously shown by various researchers. In the field of political science, Berinsky et al. [BHL12] examined MTurk as a tool for the acquisition of study participants. Within this study, an experiment that was conducted by conventional means that was then replicated using MTurk. The authors concluded that the response of MTurk participants to variation of the independent variables in the experiment was consistent with previous research. Such comparative studies have also been conducted in the fields of decision making and experimental economics, demonstrating that study results are comparable and therefore MTurk is a suitable alternative to conventional empirical settings [HRZ11], [PCP10]. Apart from these comparisons, MTurk has already been successfully used in further behavioral studies (cf. [ES10], [MW09], [SW11]) [MS12]. Based on these results, and against the background of significant cost and scale advantages over classical field studies when using a crowdsourcing platform, MTurk is employed within this study for recruiting study participants. Nevertheless, future replications of the study in field experiments are planned to provide further validation of the results and ensure their generalizability.

In the context of online experiments, it is necessary to introduce control measures that prevent unwanted behaviour of participants, which could result in manipulations or distortions of the

results. The platform may be used by people to achieve monetary gains quickly and without any effort. For this purpose, bots are often used, which run through automated studies. From the researcher's point of view, these bots generate unusable data, because they are designed to maximize the benefit for the developer and therefore only try to solve the tasks somehow and trigger a payment. To counteract this, Mason and Suri [MS12] recommend adding an additional task that serves as a CAPTCHA[3], limiting the influence of bots and ensuring proper identification of unwanted records. They also advise that participants should be aware that payment will not be made until the CAPTCHA is successfully completed [MS12]. Moreover, there is always a tendency for participants to not conscientiously fulfil the required task. This results in inconsiderate answers, which in turn appear as outliers in the statistical evaluation. In addition, some participants may only skim the instructions or answer arbitrarily. Overall, this leads to a reduction in the test strength of the experimental study [OMD09]. As a countermeasure Oppenheimer et al. [OMD09] have developed a methodological tool, the IMC (instructional manipulation check). This tool identifies whether or not the participants read the instructions carefully. This is achieved through embedding a particular question in the experimental material, with it's length and format of the answers corresponding to the answer possibilities of previous/following questions. However, this particular question tells the participant to actually ignore the answer options, but instead extract the correct solution from the question text itself. If the correct solution is not chosen, the IMC is considered failed [OMD09].

3 Setup of the Online Experiment

3.1 Definition of the Theoretical Constructs

According to previous work, with regard to the factor 'process length', it can be expected that short processes experience a higher process acceptance [Mü15], [Mü19]. However, it remains unclear whether the length refers to elapsed time or the number of necessary activities in a process. Accordingly, this independent variable is to be divided into the following: Activity-related process length (APL) and Process throughput time (PTT).

The activity-related process length refers to the number of activities required to complete a defined process[4]. Therefore, it constitutes an observable variable [DB16]. Possible repetitions or built-in loops are to be taken into account in the measurement as far as they are intended by the process. The number of repetitions depends on the respective repetition parameters. Thus, this variable is defined as follows: *Activity-related process length is the number of activities that are necessary to execute a process in its entirety. Loops deliberately planned in the process are to be included with the intended repetition parameter.*

[3] A CAPTCHA is a test, regardless of its form, that can be generated automatically. Only one person should be able to solve this test and none of the currently existing computer programs [vBL04].

[4] Based on the Davenport process definition [Da93].

Throughput time in production systems is the sum of the time elapsed for activities plus the waiting times for certain required events. It too is an observable variable [DB16], which is defined as follows, based on the general throughput time definition by Voigt [Vo18]: *Process throughput time is the duration in time units required to fully execute all activities that must be performed within a process, aggregated with the sum of the waiting times for specific required events and decisions.* Consequently, the process throughput time varies with the individual duration of the summarized elements.

3.2 Experimental Design

Since two independent variables are to be investigated at the same time within this experiment, a multi-factorial experimental design is chosen [DB16]. However, before the experimental design can be established, the characteristics of the variables must be determined. Regarding APL, four different variants exist. The variation is achieved by eliminating one or two activities, as is described in more detail in section 3.3. The PTT constitutes two variants. Since the total throughput time of the experiment and thus of the process should be kept short, the variation of the variables must be in the range of a few seconds. The values 10 and 20 seconds are chosen. Thus, a $4x2$ design is used (cf. Tab. 1) and a total of $8(4 * 2)$ factor level combinations are formed. The sub-sample sizes shall be equal. Accordingly, $8 * n = N$ participants, which are randomly assigned to the experimental groups, are required.

		APL			
		APL_1	APL_2	APL_3	APL_4
PTT	PTT_1	S_{11}	S_{12}	S_{13}	S_{14}
	PTT_2	S_{21}	S_{22}	S_{23}	S_{24}

Tab. 1: Test Design

This field of research is still relatively unexplored. For this reason, it is not possible to determine the probable effect strength from the results of previous studies [HJ02]. Therefore, an small effect strength ($f = 0.1$ or a $\eta^2 = 0.0099$) according to Cohen [Co88] is postulated, so that further planning is adjusted to be able to prove at least this small effect. According to the tool G*Power[5] [Fa07] a total sample size (N) of 2191 participants ($n = 2191/8 = 274$ per group) would be necessary. With this information, the IT architecture necessary for the experiment could be planned. Nevertheless, it is more efficient to gradually increase the sample size. On the one hand for cost reasons, on the other hand, to reduce the runtime of the experiment. One of the advantages of MTurk is that the sample size can be increased dynamically without changing the experimental design, if you have basic web programming skills [PCP10]. For the initial phase, 400 participants could be found. 33 of the records were removed due to CAPTCHA and IMC violations, so that a total of 367 participants were taken into account regarding to the statistical analyses.

[5] Version 3.1.9.4.

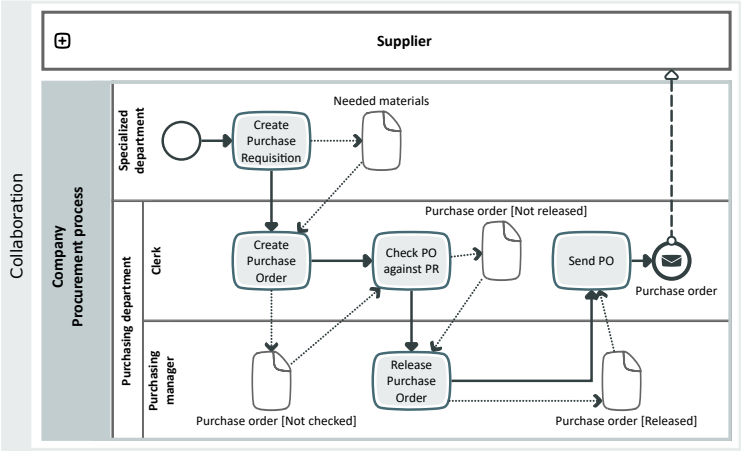


Fig. 1: Purchasing process used in the experiment

3.3 Process Design

A generic, fundamental purchasing process serves as the basis for the experiment (cf. Fig. 1). The goal is to achieve the closest possible proximity to the operational domain in practice to ensure external validity, which is often considered low for laboratory experiments, and thus allow a transfer of the results into practice. The process, including all variations, is implemented using the programming language JavaScript and a server architecture consisting of a JATOS[6] server and MTurk. Each experimental group is presented with one of the process variants outlined in Tab. 1.

Within the experiment, the variations of the independent variable APL is implemented as follows. Firstly, it is possible that the checking of a created purchase order (Check PO against PR) is omitted, which resembles an often occurring real-world problem. The omission is equivalent to the process optimization approach of 'omitting' according to Bleicher [Bl91]. Secondly, the manual sending of the purchase order could be omitted, as this is done automatically in the step following the release of the purchase order and can, therefore, be assigned to the activity 'Release purchase order'. This corresponds to the optimization approach 'Summarize' [Bl91]. The PTT is represented by a timer event after the release request, which automatically initiates the release after a previously defined time period (t). The values for t are set to 10 and 20 seconds.

The study proceeds as follows: The participant is assigned the role of a purchaser in a company. His/her task is to purchase certain materials. Before the start of each run, a description of the task is presented to the participant so that he/she receives all necessary information to be able to carry out the process (in its respective variant) correctly. Before

[6] Just another tool for online surveys (JATOS)

the participant starts the actual purchasing process, a CAPTCHA is presented. In this CAPTCHA, the participant must select the process that matches the previous description in the notation event driven process chain. If this fails, the participant is led back to the introduction of the study and can then try again. The failed attempt is recorded internally in the metadata of the study for further evaluation. The actual purchasing process begins in an order cockpit. In this cockpit, the participant receives new purchase requisitions. When a timer expires, four new purchase order items are displayed so that the participant is able to order them in his or her role as purchaser. Each position contains the name of the material, the quantity required, and details of the vendor and the recommended payment method. In this experiment, the vendor and payment method data are identical for all items. The required materials can now be ordered in the next step using an order form. Once the participant has selected the materials and chosen the necessary vendor and payment method details, the entries can be compared with the purchase requisitions. The purchase order must then be released by the purchasing manager, represented here by the server and a timer event. When the timer expires, the functionality of the send button is enabled. After sending, the participant receives a success message and is forwarded to the IMC, which contains the simple question 'What position do you have in the company?' The answer options should be ignored as indicated in the text and the title should be clicked. Failures in this test are also logged in the metadata and the participant can try again. If the participant features too many failures, the associated data set is eliminated, because this inattention calls into question the validity of the participant's answers. 33 records were removed as stated in chapter 3.2. If successful, the questionnaire will be loaded. Different variations of the process are created with regard to the aforementioned omissions of the check activity, the manual sending of the order and of both of these activities. In addition, changes were made to the timer event. The experiment runs until the desired sample size is reached. Potential confounding variables such as age, origin, education level, etc. are recorded and controlled using established mechanisms like randomization.

3.4 Research Hypotheses

In order to analyze the effects on the dependent variable with regard to the two defined factors (independent variables), research hypotheses must first be formulated. These were deductively derived from previous work on process acceptance according to Müllerleile et al. [MN14], [Mü15], [Mü19].[7] Results in this previous work indicate that a process is more accepted if fewer variables have a negative effect on it. However, this effect differs depending on the considered variables, and therefore must be defined individually. In the case of the activity-based process length, the negative effect rises the more activities the process contains. In the case of the process throughput time, it rises the longer the process takes. Consequently, the hypotheses to be tested are the following:

[7] cf. [HM94]

- EIH_{APL}: If the participants in the specified purchasing process are confronted with fewer activities (minus 1 and 2 activities), regardless of the content, the process acceptance increases, measured by the three attitude dimensions of the measurement construct described in [Mü19].
- EIH_{PTT}: With decreasing process throughput time varied by the waiting time for a release (10 and 20 seconds) in the specified purchasing process, process acceptance increases, measured by the three attitude dimensions of the measurement construct described in [Mü19].

The three dimensions reflecting process acceptance comprise cognitive, affective and conative attitude. The cognitive dimension describes the attitude towards the process constructed by the knowledge of the process executant. The affective dimension includes motivational-emotional sensations. The conative dimension describes the inner willingness to act according to the actual process [Mü19], [Gü03].

4 Data Analysis

The measurement model according to Müllerleile [Mü19], adapted to the process of this experiment, is used to record process acceptance in the three attitude dimensions (cf. Tab. 2, Tab. 3 and Tab. 4).

Question	Scale	Loading
I think this way of managing purchases is useful.	Yes/No	+
I think this way of managing purchases is important.	Yes/No	+
I think this way of managing purchases is complicated.	Yes/No	-
I think this way of managing purchases is laborious.	Yes/No	-
I think this way of managing purchases takes a long time.	Yes/No	-
I think I'm happy with the result.	Yes/No	+
I think this way of managing purchases is decent.	Yes/No	+

Tab. 2: Cognitive Dimension

Question	Scale	Loading
I would feel comfortable during the process.	Yes/No	+
I would feel included.	Yes/No	+
I would feel informed.	Yes/No	+
This purchasing process is troublesome.	Yes/No	-
This purchasing process is stressful.	Yes/No	-
I consider the purchasing process unpleasant.	Yes/No	-
I consider the purchasing process inconvenient.	Yes/No	-
I would feel insecure working the purchasing process.	Yes/No	-
I would feel annoyed working the purchasing process.	Yes/No	-

Tab. 3: Affective Dimension

Question	Scale	Loading
I would stand up for the maintaining of this purchasing process.	Yes/No	+
I would purchase materials exactly like this again.	Yes/No	+
I would have a proposal for modification.	Yes/No	-
I would like to delegate the process to another co-worker.	Yes/No	-
I would complain about this way of purchasing materials.	Yes/No	-
I would prefer another way of purchasing materials.	Yes/No	-
I would stand up for this purchasing process.	Yes/No	+

Tab. 4: Conative Dimension

Each dimension contains a set of items that have been operationalized according to the DLF IIST Binary approach by Rossiter [Ro11]. In this regard, the participant can choose between the answer options Yes and No. An answer option n/a (not available) is intentionally not offered here since an answer based on the personal inferenced threshold of satisfaction is to be enforced [Ro11]. Additionally, each dimension concludes with a bipolar single-item scale with 5 levels [Mü19]:

- Cognitive: [Overall I reject this way of purchasing material] -2 -1 0 1 2 [Overall I approve this way of purchasing material]
- Affective: [Overall I had a bad feeling] -2 -1 0 1 2 [Overall I had a good feeling]
- Conative: Overall I would like to purchase materials like suggested in this scenario [unlikely] -2 -1 0 1 2 [likely]

As already shown, the single-item scales are coded by means of a uniform interval $[-2; 2]$. The DLF IIST items are coded according to the effect coding. Which means, the values 1 and -1 are used. In the case of a regression, the regression coefficients can be interpreted as estimates of the treatment effects. This type of coding is commonly adopted within the General Linear Model for variance-analytical evaluations [BS16]. The loadings within the tables above indicate in which case the answer Yes is coded with 1 and No with -1 (Loading: +) and in which case Yes is coded with -1 and No with 1 (Loading: -). For each dimension, the response codes are aggregated according to Rossiter [Ro11] and thus provide the indicator of acceptance in this dimension and thus the input for the analyses. The statistical tests are carried out individually for each dimension.

The hypotheses presented are directed hypotheses. For this reason, a one-sided significance test is conducted. Prior, a descriptive-statistical test is carried out to assess the hypothesized direction of the effects. If the effects point in the wrong direction, a significance test is no longer necessary, since the alternative hypothesis cannot be accepted anyway [DB16]. The direction of the effects is analyzed by comparing the mean process acceptance of each dimension and variant. In the case of PTT, the mean acceptance must increase as the time decreases. As can be seen in Tab. 5, the average acceptance per acceptance dimension increases when time is shortened. Thus, the direction corresponds to the predicted one.

PTT	Cognitive	Affective	Conative
10 sec.	5.04	5.88	3.76
20 sec.	3.64	4.20	2.48

Tab. 5: Mean effects on the process acceptance per dimension (PTT)

For the variable APL, the hypothesized direction is confirmed if the average acceptance increases when the number of activities decreases. Two groups exist in which only one activity is omitted. The mean (cf. Tab. 6) between these two groups only differs slightly. The basic direction of the effects is noticeable and does not contradict the predicted one.

APL	Cognitive	Affective	Conative
Full variant	3.85	4.44	2.94
-1 Activity (w/o Check)	4.48	4.80	2.90
-1 Activity (w/o Sending)	4.30	4.85	3.18
-2 Activities	4.64	5.90	3.38

Tab. 6: Mean effects on the process acceptance per dimension (APL)

To test for significance, a two-way ANOVA is usually performed for a two-factorial design [DB16]. For an unbalanced experimental design as presented in this study, type III sums of squares must be used [BS16]. However, first of all, the requirements of ANOVA should be checked. The requirements are independence of the samples, the assumption of variance homogeneity and the assumption of normal distribution in each group [SR18]. The independence can be assumed as the participants have been acquired worldwide via MTurk and it has been technically ensured that each person can only participate once in the experiment. The variance homogeneity is to be demonstrated at this point by means of the Levene test [BS16]. The Levene test shows significant results for the affective dimension ($p_{Aff.} = 0,0020$). The results suggest that there is no variance homogeneity for this acceptance dimension. The cognitive and conative dimension show variance homogeneity ($p_{Cogn.} = 0,3970; p_{Con.} = 0,9540$). Since there is a violation of the prerequisite here, no two-factorial ANOVA should be applied, since it is not robust against this violation in an unbalanced design [MWT87]. Likewise, there is no normal distribution within the groups. The Shapiro-Wilk test for normal distribution yields a p-value < 0.05 for all dimensions. Normal distribution of the resulting measurement of acceptance for each group cannot be assumed. Since two requirements of two-factorial ANOVA have been violated, a non-parametric test, here an Aligned Rank Transform (ART) ANOVA, is performed [Wo11].

		Df	Df.res	F value	Pr(>F)	
1	APL	3	359	0.15726	0.9250	
2	PTT	1	359	12.89487	0.0004	***
3	APL:PTT	3	359	0.77340	0.5095	

Signif. Codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ‘ 1

Tab. 7: ART ANOVA for the cognitive dimension

In all three dimensions, the PTT has a p-value < 0.01. This means that significant differences

		Df	Df.res	F value	Pr(>F)	
1	APL	3	359	0.77047	0.5111	
2	PTT	1	359	7.55093	0.0063	**
3	APL:PTT	3	359	1.71581	0.1634	

Signif. Codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Tab. 8: ART ANOVA for the affective dimension

		Df	Df.res	F value	Pr(>F)	
1	APL	3	359	0.30814	0.8195	
2	PTT	1	359	7.31969	0.0071	**
3	APL:PTT	3	359	1.07096	0.3614	

Signif. Codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Tab. 9: ART ANOVA for the conative dimension

between the groups are demonstrated. Since there are only two variants and thus groups (10 sec. and 20 sec.), post-hoc tests are not necessary. To evaluate the given hypothesis, the effect strength of the variables is still necessary [HM94]. This is given by Cohen's f and the eta-square (cf. Tab. 10) [Co88].

Dimension	Cohens f	η_2
Cognitive	0.188	0.0341
Affective	0.144	0.0203
Conative	0.142	0.0196

Tab. 10: Effect strength per dimension for PTT

The differences between the APL groups are not statistically significant in any dimension. Consequently, the EIH_{APL} hypothesis cannot be accepted. This can be due to various reasons, such as insufficient sample size, and will be discussed further in chapter 5. The experiment performed has shown that process acceptance differs significantly within the PTT groups, i.e. between a waiting time of 10 and 20 seconds. However, before the given hypothesis can be accepted, an evaluation of the measured effect strength is of relevance, since even small effects can show significance in large samples [BS16], [HM94]. As the field of research is new and no comparable studies exist, no meta-analyses can be used to assess the effect strength. In such cases, the mean effect according to Cohen [Co88] ($f = 0.25$) can be used as the minimum effect strength achievable, which is common in studies in the field of psychology. The results should be supported by replication or similar studies [SG89], [SR18]. In line with Westermann and Hager [WH82], a critical effect strength with an f of 0.10 is assumed in this study, constituting a low effect when regarding Cohen's [Co88] definition. The critical effect strength has been chosen as a minimum threshold for the results of the analysis to be reasonable. With reference to these comparative values, the measured effect strengths ($f_{cognitive} = 0.188$, $f_{affective} = 0.144$, $f_{conative} = 0.142$) lead to a limitedly proven acceptance of the hypothesis EIH_{PTT} in all dimensions [Ha87], [WH82].

5 Conclusion and Discussion

5.1 Discussion of the Results

The experiment has shown that process acceptance in all three dimensions regarding a generic, fundamental purchasing process significantly depends on the process throughput time. The process throughput time was changed by varying a waiting time deterministically. According to the results, waiting times must be reduced in process optimization activities (optimization approach 'accelerate' [Bl91]) to increase process acceptance. This can be achieved, for example, through automating the release procedures by setting volume limits for releases that a system can check. The complete digitalization of the purchase process would be advisable if the decision for a release depends solely on algorithmically testable conditions. As a result, the process would have a much reduced throughput time, and thus, an even higher acceptance.

A significant impact of the number of activities in the process could not be proven. The impact of the reduction by one or two activities was too small for the sample size of this study. The power of the test ($(1-\beta) = 0.0788186$ for the lowest case (conative dimension)) leads to a high probability that the H_0-hypothesis is incorrectly assumed.

5.2 Limitations of the study

A number of limitations exist that have to taken into consideration when interpreting the results of this study. The experiment performed resembles a laboratory experiment, whose external validity (generalizability of the result) may be limited, which is mainly dependent on the artificially created environment it is conducted in. However, the traceability of the results to the variation of the independent variables is very high [DB16] in our case. The generalizability has been increased by the type of process serving as a basis of the conducted experiment. The purchasing process has been chosen to be as generic and fundamental as possible and largely reflects the standard ordering process of the widely used SAP ERP system. Nevertheless, it is important to verify the results by means of further studies, if possible with real field experiments, and thus make them more generalizable.

Furthermore, person-related confounding variables occur. These comprise for example the age, the level of qualification, the country of origin and language (the investigation was conducted in English). These have been controlled by means of randomization in group allocation. Nevertheless, it is useful to analyze the distributions of these variables among the participants. As far as age is concerned, the majority of participants are distributed in the interval of $24-37$ years. This is probably because MTurk is more likely to be used by people with an affinity for technology and that they are more likely to be found in this age interval [Mü19]. With regard to the degree of qualification, the most common degrees have been recorded. 56.7% of the participants hold a bachelor's degree, closely followed by college

students (11.2%) and persons with a Master's degree (10.9%). Therefore, it can be assumed that, in terms of educational level, the participants were able to fully understand the required task and to perform the process as required. Although MTurk makes it technologically possible to acquire participants worldwide, 63.5% of participants currently live in the USA and 27.2% in India. Thus, as the underlying process of the experiment is delivered in English, questions and answers are likely to have been understood by the participants.

5.3 Implications for Further Research

The significant difference in process throughput time within this study is based on the variation of waiting time. Although this is by definition an essential part of the throughput time, further investigations should be carried out with regard to it. For example, it should be investigated whether the measured effects are based on the fact that the participants had to wait for an event without further activity. For this purpose, the duration of a single activity, can be tested for significant differences. Furthermore, regression methods can be used to explain the relationship between process throughput time and process acceptance and thus enable predictions. Regarding APL, the EIH_{APL} hypothesis could not be accepted in the existing design. In further research, an experimental design could be chosen that contains a clearer reduction of the process activities so that the effect of the reduction is more noticeable. Moreover, the number of participants should be increased. Finally, our experiments focus on two independent variables, while other variables of potential influence on process acceptance in practice are excluded for reasons of complexity. Future studies will aim at measuring the importance of such other factors and also their possible interdependences.

References

[BHL12] Berinsky, A. J.; Huber, G. A.; Lenz, G. S.: Evaluating Online Labor Markets for Experimental Research: Amazon.com's Mechanical Turk. Political Analysis 20/03, pp. 351–368, 2012.

[BK12] Becker, J.; Kahn, D.: Der Prozess im Fokus. In (Becker, J.; Kugeler, M.; Rosemann, M., eds.): Prozessmanagement. Springer Gabler, Berlin and Heidelberg, pp. 3–16, 2012.

[Bl91] Bleicher, K.: Organisation: Strategien - Strukturen - Kulturen. Gabler, Wiesbaden, 1991.

[BS16] Bortz, J.; Schuster, C.: Statistik für Human- und Sozialwissenschaftler: Extras online. Springer, Berlin and Heidelberg, 2016.

[Co88] Cohen, J.: Statistical power analysis for the behavioral sciences. Erlbaum, Hillsdale, NJ, 1988.

[Da93] Davenport, T. H.: Process innovation: Reengineering work through information technology. Harvard Business School Press, Boston, Mass., 1993.

[DB16] Döring, N.; Bortz, J.: Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften. Springer, Berlin and Heidelberg, 2016.

[ES10] Eriksson, K.; Simpson, B.: Emotional reactions to losing explain gender differences in enteringa risky lottery. Judgment and Decision Making 2010/5, pp. 159–163, 2010.

[Fa07] Faul, F.; Erdfelder, E.; Lang, A.-G.; Buchner, A.: G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behavior Research Methods 39/2, pp. 175–191, 2007.

[Gü03] Güttler, P. O.: Sozialpsychologie: Soziale Einstellungen, Vorurteile, Einstellungsänderungen. Oldenbourg, München, 2003.

[Ha87] Hager, W.: Grundlagen einer Versuchsplanung zur Prüfung empirischer Hypothesen der Psychologie. In (Lüer, G.; Becker, D., eds.): Allgemeine experimentelle Psychologie. UTB für Wissenschaft Grosse Reihe, Fischer, Stuttgart, pp. 43–264, 1987.

[HC06] Hammer, M.; Champy, J.: Reengineering the corporation: A manifesto for business revolution. Collins Business Essentials, New York, 2006.

[HJ02] Hussy, W.; Jain, A.: Experimentelle Hypothesenprüfung in der Psychologie. Hogrefe Verlag für Psychologie, Göttingen, 2002.

[HM94] Hussy, W.; Möller, H.: Hypothesen. In (Herrmann, T.; Tack, W. H., eds.): Methodologische Grundlagen der Psychologie. Vol. 1, Enzyklopädie der Psychologie, Hogrefe–Verlag für Psychologie, Göttingen and Seattle, pp. 475–507, 1994.

[HRZ11] Horton, J. J.; Rand, D. G.; Zeckhauser, R. J.: The online laboratory: conducting experiments in a real labor market. Experimental Economics 14/3, pp. 399–425, 2011.

[Mi08] Militz, M.: Strategien zur Eingrenzung multiresistenter Erreger im Krankenhaus. Trauma und Berufskrankheit 10/S1, pp. 140–145, 2008.

[MN14] Müllerleile, T.; Nissen, V.: When Processes Alienate Customers: Towards a Theory of Process Acceptance. In (van der Aalst, W.; Mylopoulos, J.; Rosemann, M.; Shaw, M. J.; Szyperski, C.; Nanopoulos, A.; Schmidt, W., eds.): S-BPM ONE - Scientific Research. Vol. 170, Lecture Notes in Business Information Processing, Springer International Publishing, Cham, pp. 171–180, 2014.

[MS12] Mason, W.; Suri, S.: Conducting behavioral research on Amazon's Mechanical Turk. Behavior Research Methods 44/1, pp. 1–23, 2012.

[Mü15] Müllerleile, T.; Ritter, S.; Englisch, L.; Nissen, V.; Joenssen, D. W.: The Influence of Process Acceptance on BPM: An Empirical Investigation. In: 2015 IEEE 17th Conference on Business Informatics. IEEE, pp. 125–132, 2015.

[Mü19] Müllerleile, T.: Prozessakzeptanz: Theoretische und empirische Untersuchung der Akzeptanz und Ablehnung betrieblicher Prozesse. Springer Gabler, Wiesbaden, 2019.

[MW09] Mason, W.; Watts, D. J.: Financial incentives and the "performance of crowds". In (Bennett, P., ed.): Proceedings of the ACM SIGKDD Workshop on Human Computation. ACM, New York, p. 77, 2009.

[MWT87] Milligan, G. W.; Wong, D. S.; Thompson, P. A.: Robustness properties of nonorthogonal analysis of variance. Psychological Bulletin 101/3, pp. 464–470, 1987.

[OMD09] Oppenheimer, D. M.; Meyvis, T.; Davidenko, N.: Instructional manipulation checks: Detecting satisficing to increase statistical power. Journal of Experimental Social Psychology 45/4, pp. 867–872, 2009.

[PCP10] Paolacci, G.; Chandler, J.; Panagiotis, I. G.: Running experiments on Amazon Mechanical Turk. Judgment and Decision Making 2010/5, pp. 411–419, 2010.

[Ro11] Rossiter, J. R.: Measurement for the Social Sciences: The C-OAR-SE Method and Why It Must Replace Psychometrics. Springer Science+Business Media LLC, New York, 2011.

[SG89] Sedlmeier, P.; Gigerenzer, G.: Do studies of statistical power have an effect on the power of studies? Psychological Bulletin 105/2, pp. 309–316, 1989.

[SR18] Sedlmeier, P.; Renkewitz, F.: Forschungsmethoden und Statistik: Für Psychologen und Sozialwissenschaftler. Pearson, Hallbergmoos, 2018.

[SW11] Suri, S.; Watts, D. J.: Cooperation and contagion in web-based, networked public goods experiments. PloS one 6/3, e16836, 2011.

[vBL04] von Ahn, L.; Blum, M.; Langford, J.: Telling humans and computers apart automatically. Communications of the ACM 47/2, pp. 57–60, 2004.

[Vo18] Voigt, K.-I.: Definition: Durchlaufzeit, ed. by Springer Gabler, 2018, URL: https://wirtschaftslexikon.gabler.de/definition/durchlaufzeit-32490/version-256033.

[WH82] Westermann, R.; Hager, W.: Entscheidung über statistische und wissenschaftliche Hypothesen: Zur Differenzierung und Systematisierung der Beziehungen. Zeitschrift für Sozialpsychologie/13, pp. 13–21, 1982.

[Wo11] Wobbrock, J. O.; Findlater, L.; Gergle, D.; Higgins, J. J.: The aligned rank transform for nonparametric factorial analyses using only anova procedures. In (Tan, D.; Fitzpatrick, G.; Gutwin, C.; Begole, B.; Kellogg, W. A., eds.): Conference proceedings and extended abstracts / the 29th Annual CHI Conference on Human Factors in Computing Systems. ACM, New York, pp. 143–146, 2011.

Identifikation von Gestaltungsfaktoren der Prozessmodellierung

Erste Ergebnisse einer qualitativen Studie

Aleksandra Dzepina[1] Franz Lehner[2]

Abstract: Das Geschäftsprozessmanagement leistet seit mehreren Jahrzehnten einen wichtigen Beitrag für den Unternehmenserfolg. Im Rahmen der Aufgaben des Geschäftsprozessmanagements nimmt die Prozessmodellierung eine zentrale Stellung ein. Dennoch haben insbesondere kleine und mittelständische Unternehmen Schwierigkeiten bei der erfolgreichen Umsetzung. Dies liegt unter anderem an einer fehlenden Prozessmanagementabteilung, aber auch an vorwiegend dezentral organisierten Modellierungseinheiten. Dieser Beitrag gewährt Einblick in die aktuelle Umsetzung der Prozessmodellierung in deutschsprachigen Unternehmen und bezieht sich dabei auf 26 durchgeführte Interviews, mit Experten aus Forschung und Praxis. Dabei wird deutlich, dass sich nach wie vor keine Notationssprache als Modellierungsstandard etablieren konnte und auch die Beurteilung der Qualität der Prozessmodelle weiterhin zweitrangig bleibt. Ferner wird ersichtlich, dass aus Sicht der Praxis Handlungsbedarf beim Aufbau von Prozesshäusern besteht, da es bis heute an einer standardisierten Lösung für eine personen- und anforderungsspezifische Navigation fehlt. Im Rahmen der Studie werden durch eine induktive Vorgehensweise fünf Gestaltungsfaktoren abgeleitet, die für eine erfolgreiche Umsetzung der Prozessmodellierung von Bedeutung sein können.

Keywords: Geschäftsprozessmanagement; Prozessmodellierung; Gestaltungsfaktoren; Reifegrad

1 Einleitung

Das Geschäftsprozessmanagement wurde erstmals von *Hammer und Champy* in den frühen 90iger Jahren vorgestellt, damals noch unter der Prämisse der Neuorganisation und –restrukturierung, auch besser bekannt als Business Process Reengineering [HC93]. Darauf folgte die Entwicklung zahlreicher Prozessmodellierungssprachen, die sich heutzutage einer großen Beliebtheit für die graphische Darstellung von Prozessen erfreuen. Es gibt kaum ein Unternehmen, das mit der Verwendung einer Prozessmodellierungssprache nicht vertraut ist und sei es nur die Abbildung des Prozesses in einem Flow-Chart oder mittels Microsoft Visio. Dennoch beschreibt *Vanderhaagen* die Heterogenität verwendeter Modellierungssprachen und die häufig monolithische Architektur von der dahinter stehenden Systeme als große Herausforderung, der Unternehmen heutzutage begegnen müssen [VFL10]. Das weitere Aufkommen von neuen Technologien, wie zum Beispiel Data Mining und Process Mining,

[1] Universität Passau, Wirtschaftsinformatik, Innstr. 43, 94032 Passau, aleksandra.dzepina@uni-passau.de
[2] Universität Passau, Wirtschaftsinformatik, Innstr. 43, 94032 Passau, franz.lehner@uni-passau.de

doi:10.18420/inf2020_57

hat zusätzlich Vorteile mit sich gebracht. Dennoch bilden diese Technologien lediglich Schnittstellen zum Geschäftsprozessmanagement, sodass eine fehlerhafte oder wenig erfolgreiche

Umsetzung dessen, sich auf die Nutzung der neuen Technologien auswirken kann. Da die Verwendung von Process-Mining-Methoden ein fundiertes Prozesswissen voraussetzt, fällt es Unternehmen schwer die neuen Methoden zu adaptieren, wenn die Prozessmodellierung nicht einen gewissen Reifegrad aufweist. Es hat sich gezeigt, dass insbesondere KMUs, denen es nicht möglich war eine separate Prozess-managementabteilung aufzubauen, zu Nachzüglern im Wettbewerb wurden. Die fehlenden finanziellen Ressourcen für die Lizenzen der Modellierungssoftware führten zu einer Verwendung von frei erhältlichen Modellierungstools, die zum Teil auf andere Vorhaben ausgelegt waren und zu Prozessmodellen geringerer Qualität führten. Obwohl sich die Prozessmodellierung innerhalb der letzten Jahrzehnte etablieren konnte, bleibt weiterhin die Frage bestehen, inwiefern Forschung und Praxis mit der Umsetzung zufrieden sind. Es ist anzunehmen, dass sich die Beachtung von Gestaltungsvorgaben, wie beispielsweise die einheitliche Verwendung einer Modellierungssprache, positiv auf die Umsetzung der Prozessmodellierung auswirken kann. Somit gilt es zu klären, welche Gestaltungsfaktoren für die Prozessmodellierung von Bedeutung sein können. Mit dieser Studie soll an diese Frage angeknüpft werden und die Gestaltungsfaktoren, die innerhalb der letzten Jahrzehnte für die Prozessmodellierung identifiziert wurden, benannt und erläutert werden. Zudem soll untersucht werden, inwiefern sich die Unternehmensgröße und –branche auf die erfolgreiche Umsetzung der Gestaltungsfaktoren auswirkt. Das Ergebnis ist eine Liste an Gestaltungsfaktoren, die die Experten in den Interviews beschreiben und deren Umsetzung positive Einflüsse auf die Prozessmodellierung verspricht.

2 Einordnung in die bisherige Forschung

Es haben sich bisher zahlreiche Autoren mit der Entwicklung des Geschäftsprozessmanagements befasst. Reifegradmodellen kommt dabei eine besondere Bedeutung zu. Es wird zwischen Prozess-Reifegradmodellen und Prozessmanagement-Reifegradmodellen unterschieden, wobei letztere das gesamtheitliche Geschäfts-prozessmanagement, insbesondere die Ausgestaltung und Steuerung der Prozesse unterstützen [RK12]. Unter den Prozessmanagement-Reifegradmodellen finden sich einige wieder, die auch Aspekte der Prozessmodellierung abdecken [RB05; Ro09; Va10]. Roeglinger et al. kommen zu der Erkenntnis, dass es aber bisher an einem einheitlichen Verständnis für den Reifegradbegriff der Prozessmodellierung fehlt. Ferner erschweren unveröffentlichte Studien zu Reifegradmodellen, die Teilaspekte der Prozessmodellierung abdecken, die Anwendung durch die Praxis.

Neben den Reifegradmodellen spielen Referenzmodelle und Referenzmodellierungssprachen eine besondere Rolle, die es Modellierern ermöglichen soll, flexible Änderungen an Prozessmodellen vorzunehmen [BDK04]. Der Prozess der

Referenzmodellierung beinhaltet dabei drei Aspekte: Charakterisierung, Konstruktion und Anwendung. Diese decken wiederum Fragestellungen der Prozessmodellierung ab, weil man sich beispielsweise mit dem optimalen Zugang zu Prozessmodellen, der Tooleignung oder der Auswahl einer geeigneten Prozessmodellierungssprache beschäftigt [FLZ05].

Es existieren darüber hinaus weitere Forschungsansätze, die sich zwar mit ähnlichen Fragestellungen beschäftigen, aber keine Einordung in einen Reifegrad oder die Anpassung eines Referenzmodells vorsehen. Bucher und Winter untersuchen in einer empirischen Studie sogenannte Gestaltungsfaktoren der Prozessmodellierung, deren Beachtung sich positiv auf die wahrgenommene Reife im Unternehmen auswirken kann. Zu diesen Gestaltungsfaktoren zählen beispielsweise die Performance-Messung oder die Nutzung etablierter Modellierungsstandards und Methoden [BW14]. Kurz wiederum, spricht von einer Selbstorganisation des Geschäftsprozessmanagements und strebt dabei das selbstständige Erkennen von Verbesserungspotenzialen durch Mitarbeiter an. Dabei sollen beispielsweise alle Prozessmitarbeiter für die Methoden des Geschäfts-prozessmanagements geschult werden, um die Qualität der Prozessmodelle sicherzustellen [Ku11]. Zur Muehlen kommt in einem mit Phil Gilbert, IBM, durchgeführten Interview zur Erkenntnis, dass die Entwicklung sogenannter ‘Centers of Excellence‘ der Prozessmodellierung zuträglich wäre [Zu12]. Dabei handelt es sich unter anderem um Plattformen mit optimaler Technologieunterstützung, Kernaspekte der IT, die bereits auch in Reifegradmodellen Verwendung fanden.

Den Forschungsstand zusammenfassend, haben sich bereits einige Autoren mit für die Prozessmodellierung relevanten Kriterien beschäftigt. Diese waren jedoch häufig Bestandteil von Reifegrad- oder Referenzmodellen. Davon abgeleitet, sind weitere Studien hervorgegangen, die sich ebenfalls mit einer erfolgreichen Umsetzung der Prozessmodellierung befasst haben, dies aber ohne Einordung in dafür vorgesehene Reifegrade oder Anpassung von Referenzmodellen. Dabei handelt es sich um Studien zu sogenannten Gestaltungsfaktoren, die nach jetzigem Wissensstand noch nicht aus Sicht der Praxis und Forschung untersucht worden sind. Diese Studie ergänzt somit die bisherigen Erkenntnisse in Hinblick auf Gestaltungsfaktoren und stellt erstmals die Unterschiede zwischen Forschung und Praxis dar. Zudem können die aus der Studie induktiv hervorgegangenen Gestaltungsfaktoren zukünftig in Kombination mit bereits bekannten Reifegrad- und Referenzmodellen Verwendung finden und diese ergänzen.

3 Forschungsdesign

Für die Beantwortung der Fragestellung hinsichtlich der Gestaltungsfaktoren der Prozessmodellierung wurde die qualitative Erhebungsmethode einer quantitativen vorgezogen. Die strukturierte, quantitative Datenerhebung in Verbindung mit theoretischen Modellen wäre aufgrund nicht eindeutig definierter Begriffe zu den Gestaltungsfaktoren nur erschwert möglich gewesen. Zudem stand die Erhebung individueller Aussagen im Vordergrund, wodurch sich das explorative Experteninterview als geeigneter herausstellte. Die Interviews wurden auf Basis eines Leitfadens mit 18 offen formulierten Fragen durchgeführt. Dabei wurden

die Interviewpartner zur Gestaltung der Prozessmodellierung befragt, insbesondere zur Verwendung von Modellierungssprachen, zur Ausgestaltung von Geschäftsprozessmodellen und zu Mängeln dieser. Der Interviewleitfaden diente zwar der Eingrenzung des Themengebiets, das Interview selbst wurde aber bewusst in einer halbstrukturierten Form gestaltet, um die Offenheit der qualitativen Forschung zu sichern. Für die Auswahl der Experten aus der Praxis wurde Xing, ein soziales Netzwerk für berufliche Kontakte, verwendet. Es wurden Experten ausgewählt, deren Tätigkeitsbeschreibung ein entsprechendes Prozessverständnis erkennen ließ. Dementsprechend wurden die folgenden Tätigkeitsbezeichnungen als Suchparameter verwendet: Prozessmanager, Prozesseigner, Qualitätsmanager, Auditor und Prozessanalyst. Dabei wurde zusätzlich darauf geachtet, dass die Experten mindestens ein Jahr Berufserfahrung im Prozessbereich aufweisen konnten. Zusätzlich zu den Personen aus der Praxis wurde gezielt nach Experten aus der Forschung gesucht. Hier diente überwiegend die Internetpräsenz von Lehrstühlen der Wirtschaftsinformatik als Anlaufstelle. Es wurden Mitarbeiter jener Lehrstühle angeschrieben, die einen Schwerpunkt im Bereich Geschäftsprozessmanagement und Prozessmodellierung aufweisen konnten. Schließlich konnten 26 Interviewpartner für die Studie gewonnen werden.

4 Darstellung des Untersuchungsfeldes

Es wurde ein möglichst heterogenes Untersuchungsfeld gewählt, das unabhängig von der Unternehmensgröße, -branche und der Tätigkeitsbeschreibung der befragten Person ist. Die Experten aus der Forschung wurden überwiegend durch Wirtschaftsinformatik- Absolventen repräsentiert. Die nachfolgenden Abbildungen stellen die absoluten Häufigkeiten der Interviewpartner aus Praxis und Forschung dar.

Abb. 1: Verteilung der Interviewpartner Praxis

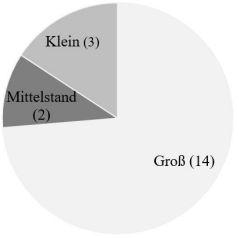


Abb. 2: Verteilung der Interviewpartner Forschung

Die nachfolgende Grafik gibt die absoluten Häufigkeiten der befragten Unternehmensbranchen wieder.

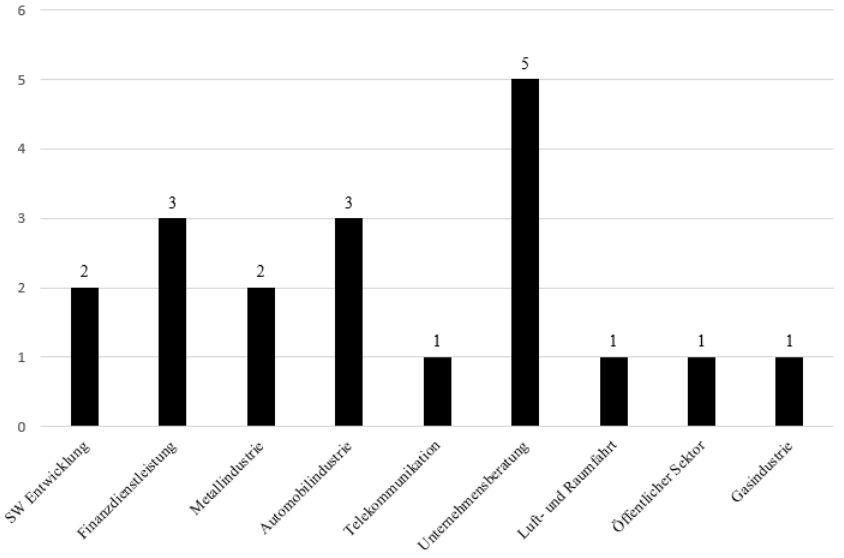


Abb. 3: Anzahl der Interviewpartner nach Unternehmensbranche

5 Durchführung der Interviews und Darstellung der Untersuchungsergebnisse

Die aufgezeichneten Interviews wurden zunächst unter Verwendung von MaxQDA 2020 transkribiert, codiert und ausgewertet. Da zwei Interviews mit Praxispartnern eine sehr schlechte Tonqualität aufwiesen, wurden diese zusätzlich in Anschluss an die Interviews in Gedächtnisprotokollen festgehalten. Nachdem die Transkripte fertiggestellt waren, wurden diese auf Rechtschreibung überprüft. Die fertiggestellten Transkripte dienten dann als Basis für die Codierung. Aus den Interviews resultierten somit 24 Transkriptionen und zwei Gedächtnisprotokolle. Für die Codierung der Interviews wurde die typisierende Strukturierung der qualitativen Inhaltsanalyse ausgewählt, da diese die Beschreibung von Extremen, Häufigkeiten und Prototypen zulässt [Ma15]. Eine rein zusammenfassende Inhaltsanalyse wäre für die Gegenüberstellung von Praxis und Forschung nicht sinnvoll gewesen, da für die Bewertung der Umsetzung der Gestaltungsfaktoren eine Zählung ausschlaggebend war.

Es wurden dabei Sinneinheiten codiert, die mindestens einen Satz enthielten. Der Text wurde in Hinblick auf die Fragestellung und unter Berücksichtigung der definierten Codiereinheiten analysiert. Die Codiereinheiten gingen aus den geschilderten Erfolgen und Misserfolgen bei der Prozessmodellierung sowie Problemen und Best-Practice-Ansätzen hervor. Die relevanten Textstellen wurden markiert, gesammelt und paraphrasiert. Im

Anschluss wurden die Paraphrasierungen generalisiert und jene Aussagen mit inhaltlich gleicher Aussage reduziert. Während einer zweiten Reduktion wurden Paraphrasen mit gleicher Aussage gebündelt und Paraphrasen mit mehreren Aussagen zusammengeführt. Die Interviewpartner beschrieben zudem weitere Faktoren, die sich nicht auf den Prozess der Modellierung, sondern das Produkt, die Prozessmodellqualität, bezogen. Diese wurden im Anschluss separat analysiert. Als Ergebnis der qualitativen Inhaltsanalyse gingen fünf Gestaltungsfaktoren hervor, die induktiv abgeleitet wurden.

5.1 Gestaltungsfaktoren und ihre Ausprägungen

Die Interviews wurden in Hinblick auf Gestaltungsfaktoren untersucht und bewertet. Hierbei wurde ersichtlich, dass insbesondere Praxis und Forschung unterschiedliche Vorstellungen von einer erfolgreichen Umsetzung der Prozessmodellierung aufweisen. Zudem wurde deutlich, dass vor allem kleinere Unternehmen Verbesserungspotenzial in puncto Prozessmodellierung sehen, aber Schwierigkeiten damit haben, Erfolgspotenziale zu erkennen und zu nutzen. Im Anschluss sollen die aus der typisierenden Strukturierung hervorgegangenen 'Prototypen', die sogenannten Gestaltungsfaktoren, aufgezählt und näher beschrieben werden: Zentrale Modellierungseinheit, Verwendung der BPMN, Aufbau eines Prozesshauses, separate Qualitätssicherung der Prozessmodelle und Messung von Prozessmodellqualität.

(1) Zentrale Modellierungseinheit

Aus den Interviews geht hervor, dass im öffentlichen Sektor das Bewusstsein für eine zentrale Modellierungseinheit erst erworben werden musste. So wird in einem Fall beschrieben, dass 177 Prozessmodelle einst dezentral modelliert wurden. Dies führte dazu, dass nicht alle modellierenden Personen mit den Modellierungsrichtlinien vertraut waren und somit einheitliche Prozessmodelle nur schwer sichergestellt werden konnten. Zudem führten unterschiedliche Granularitätsstufen zu unterschiedlich informationsintensiven Prozessmodellen. Man konnte das Problem beheben, indem eine zentrale Modellierungseinheit von vier Personen die Modellierung der Prozesse übernahm. Zudem wird die Auffassung geteilt, dass nur 25% der Arbeitszeit eines Modellierers für das Zeichnen verwendet werden sollte. Die restlichen 75% werden für das Erfassen, Nachfragen und Layoutüberlegungen benötigt. Es stellte sich heraus, dass insbesondere Personen, die sich mit der Modellierung nur nebenbei befasst hatten, Schwierigkeiten bei der Umsetzung hatten. Es kamen Fragen der Prozesswürdigkeit, insbesondere zur Erhebung, Gestaltung und Steuerung von Prozessen sowie deren Umsetzbarkeit auf, die bei Personen mit mehr Modellierungserfahrung nicht mehr auftraten. Daher stellte sich die zentrale Modellierungseinheit als beste Lösung dar. Bei einer großen internen Unternehmensberatung von insgesamt 400.000 Mitarbeitern wird beispielsweise zwischen der Erstaufnahme und Wartung von Prozessmodellen unterschieden. Hier wird es nicht für sinnvoll erachtet, Mitarbeiter zur Verfügung zu stellen, die sich ausschließlich mit der Modellierung von Prozessmodellen befassen. Nur für die Erstaufnahme eines Prozessmodells stehen sogenannte Modellierungsexperten zur Verfügung. Dies stellt unter anderem sicher, dass die Prozessmodelle keine syntaktischen Fehler aufweisen und leicht verständlich sind. Muss das Prozessmodell im Anschluss gewartet werden, so

sollte dies idealerweise von dem verantwortlichen Prozessmanager selbst durchgeführt werden, da dieser mit den Prozessabläufen vertraut ist. Dies hat jedoch zur Folge, dass insbesondere die Prozessmanager, aber auch in Prozessen involvierte Mitarbeiter, mit der Prozessmodellierung vertraut sein sollten und regelmäßig ihr Wissen aktualisieren sollten. Eine mittelständige Unternehmensberatung sieht ebenfalls Vorteile in einer zentralen Modellierungseinheit, insbesondere, weil ein Vergleich zwischen zentral und dezentral möglich ist. Die Vorteile, die sich durch die zentrale Modellierungseinheit ergeben, sind insbesondere die Integration in das Gesamtmodell, Widerspruchsfreiheit und verbesserte Wartbarkeit. Es kann bei großen heterogenen Gruppen jedoch bis zu einem Jahr dauern, bis die gewünschte Modellierungskompetenz erreicht wird.

(2) Verwendung der BPMN

Für die BPMN wird empfohlen, nicht sofort mit der Notation zu beginnen, sondern die Erstaufnahme des Prozesses mit einem Flipchart und mehreren Post-its durchzuführen. Dies bietet den Vorteil, dass man sich in den Workshops auf den Sachverhalt konzentrieren kann. Später kann die Abbildung des Prozesses dann von einer geschulten, zentralen Modellierungseinheit erfolgen. Ein kleines Softwareentwicklungsunternehmen für Prozessautomatisierungslösungen beschreibt beispielsweise, dass sie die Erfahrung gemacht haben, dass jeder Kunde auf unterschiedliche Prozessmodellierungssprachen und –software setzt. ARIS sei immer noch verbreitet, Prozessmodelle werden aber ebenso mit nicht ausschließlich dafür vorgesehener Software wie z.B. Visio oder Powerpoint modelliert. Das Fehlen von formalen Vorgaben ist insofern verwunderlich, da insbesondere für Prozessautomatisierungen reine Text- oder Arbeitsanweisungen nicht ausreichend sind. Zudem wird beschrieben, dass für viele Prozessmodelle rein fachliche und nicht formale Vorgaben von Bedeutung sind. Ein Großkonzern der Metallindustrie stellt die Prozessmodelle zwar hauptsächlich mittels ereignisgesteuerter Prozessketten dar, zu Teilen aber auch mittels einer symboleingeschränkten BPMN. Die Anzahl der BPMN Symbole wurde dabei bewusst beschränkt, führte im Umkehrschluss aber zu reduzierten Prozessmodellen, insbesondere den Detailgrad betreffend. Ein Finanz-dienstleistungsunternehmen mit 3.500 Mitarbeitern beschreibt zum Beispiel, dass für die Durchführung eines IT-Projektes ca. 25 IT-Prozesse beteiligt sind, darunter insbesondere Softwareerstellung und Projektmanagement. Prozessbeschreibungen dieser 25 Prozesse sind nicht in Prozessmodellen aufzufinden, jedoch in zwanzig-, bis dreißigseitigen Dokumenten, textuell beschrieben. Dies verhält sich jedoch anders bei Geschäftsprozessen, diese werden in der Regel in BPMN modelliert. Aus Sicht der Forschung ist bekannt, dass viele Prozesse mittels Visio oder Powerpoint modelliert werden. Dennoch wird aus Sicht der Forschung eine einheitliche Lösung, beziehungsweise die Verwendung von Frameworks, wie zum Beispiel Camunda empfohlen sowie der Gebrauch gängiger Modellierungssprachen wie BPMN. Zudem wird beschrieben, dass in der Lehre häufig mit EPKs begonnen wird, da diese intuitiv und semantisch leicht verständlich sind, um dann die Grundlagen für die BPMN zu schaffen. In der Forschung wird die BPMN als Standardnotation angesehen, was in einem Widerspruch zu Einsatzhäufigkeit in der Praxis steht.

(3) Aufbau eines Prozesshauses

Ein Prozesshaus dient in der Regel der Strukturierung und Visualisierung von Prozessmodellen unterschiedlicher Abstraktionsebenen. In der Forschung für Fertigungstechnik werden für die Erstellung eines Prozesshauses Referenzmodelle berücksichtigt, insbesondere um den erstmaligen Modellierungsaufwand zu reduzieren. Das Baukastenprinzip der Referenzmodelle ermöglicht die schnelle Erweiterung um detailliertere Ebenen. Es muss jedoch beachtet werden, dass die Ergebnisse der Referenzprozesse an die Realwelt angepasst werden müssen, um exaktere Prozessmodelle hervorbringen zu können. Die Vorteile eines Prozesshauses liegen in dem strukturierten Aufbau der Prozesslandschaft und dem Sichtenkonzept, mit dem Prozessmodelle für unterschiedliche Benutzergruppen mit unterschiedlichen Informationen angereichert werden. Allerdings ist noch kein Tool bekannt, in welchem unterschiedliche Sichten oder Varianten eines Prozesses möglichst benutzerfreundlich verwaltet werden könnten. Die Praxis versucht sich zu helfen, indem sie eigene Lösungen entwickelt. Für einen Großkonzern der Gasindustrie beispielsweise, konnten nicht alle Anforderungen an ein Prozesshaus mit einer herkömmlichen Softwarelösung abgedeckt werden. Insbesondere im Bereich Navigation hatte man sich mehr Unterstützung erhofft und somit auf eine eigens entwickelte Prozesshauslösung gesetzt. Ein Großkonzern der Automobilindustrie beschreibt hauptsächlich die Herausforderung der Ländergrenzen und damit verbundener, offener Schnittstellen. Ziel einer Prozesshauslösung sollte es sein, diese offenen Schnittstellen zu beseitigen. Auch der Großkonzern hat sich für eine eigens entwickelte Prozesshauslösung entschieden, die die Überprüfung der Prozessmodelle auf Korrektheit unterstützt und somit Prozessmodelle mit noch offenen Schnittstellen gar nicht erst zur Verfügung stellt, sondern vorerst eine Fehlermeldung ausgibt.

(4) Separate Qualitätssicherung der Prozessmodelle

Der Mittelstand der Automobilindustrie verfügt über cirka 20 Prozesseigner, die wiederum Mitarbeiter zur Verfügung gestellt bekommen. Diese sogenannten Prozessersteller sind für die Modellierung der Prozesse zuständig. Dennoch wird die Freigabe der Prozessmodelle separat vom IT-Prozessmanager durchgeführt, um sicherzustellen, dass alle Prozesse einheitlich und auf einem qualitativ guten Niveau sind. Sind die Prozessmodelle nicht in Ordnung, werden diese von dem IT-Prozessmanager mit einem Vermerk versehen, von dem Prozessersteller ausgebessert und erneut geprüft. Handelt es sich dabei um Kleinigkeiten, wie beispielsweise Tipp- oder Benennungsfehler, werden diese vom IT-Prozessmanager selbst ausgebessert. Erst wenn die Prozessmodelle von dem IT-Prozessmanager für gut befunden wurden, werden diese für einen gesonderten Freigabeworkflow genehmigt. Eine große, interne Unternehmensberatung arbeitet im Rahmen der Qualitätssicherung und Bereitstellung der Prozessmodelle mit Staging- und Release-Bereichen. Sobald ein Prozessmodell bearbeitet wird, kann es nicht mehr im Produktiv-, sondern im Stagingbereich aufgefunden werden. Erst bei Freigabe der neuen Änderung, wird das Prozessmodell wieder im Produktivbereich sichtbar. Auf diese Weise kann sichergestellt werden, dass prozessinvolvierte Personen stets mit einem geprüften und freigegebenen Modell arbeiten. Hierbei wird großen Wert auf die syntaktische Korrektheit gelegt.

(5) Messung von Prozessmodellqualität

Die Stimmen bezüglich der Messung der Prozessmodellqualität sind vor allem bei den Experten aus der Praxis kritisch, was überwiegend mit einem großen Zeitaufwand begründet wird. Obwohl die Messung der Prozessqualität mit Kennzahlen, wie beispielsweise Durchlaufzeiten, Gesamtkosten etc. verbreitet ist, wird die Messung der dahinterliegenden Prozessmodellqualität nicht vollzogen. Lediglich einer von 19 befragten Praxispartnern, Mitarbeiter einer internen Unternehmensberatung in einem Großkonzern, kann berichten, dass die Prozessmodellqualität gemessen wird. Allerdings werden hier andere Prozessmodellqualitätskennzahlen erhoben, als in der Forschung. So soll beispielsweise die Anzahl der Change Requests nach Freigabe eines Prozesses Aussage über die Reife eines Prozesses treffen. Man leitet davon ab, dass Prozesse mit einer großen Anzahl an Change Requests für eine fehlende Reife und somit geringe Prozessmodellqualität sprechen. Der Nachteil der sich aus dieser Kennzahl ergibt ist, dass die Reife erst nach Freigabe des Prozessmodells bewertet werden kann. Die Experten aus der Praxis erachten jedoch insbesondere bei einer dezentralen Modellierung die Erhebung von zusätzlichen Prozessmodellqualitätskennzahlen als sinnvoll. Auf diese Weise könnte leichter sichergestellt werden, dass die modellierten Prozesse einheitlich gestaltet werden, um auch den Nacharbeitungsaufwand oder die separate Qualitätssicherung zu reduzieren.

5.2 Zusammenfassender Überblick

Die nachfolgenden Tabellen geben Auskunft darüber, wie viele Experten sich für einen der fünf Gestaltungsfaktoren ausgesprochen haben. Dabei steht ein ‘ ‘ für die Umsetzung des Gestaltungsfaktors und eine ‘0‘ für keine Umsetzung. Die Gestaltungsfaktoren der Tabelle sind in der in Kapitel 5.1 verwendeten Reihenfolge angeordnet: Zentrale Modellierungseinheit, Verwendung der BPMN, Aufbau eines Prozesshauses, separate Qualitätssicherung der Prozessmodelle und Messung von Prozessmodellqualität. Das Buchstabenkürzel hinter dem anonymisierten Experten stellt dabei die Unternehmensgröße dar, wobei k für klein, g für groß und m für mittelständig steht.

6 Fazit

Die Interviews geben einen Einblick in den aktuellen Stand der Prozessmodellierung in deutschsprachigen Unternehmen. Sie zeigen auf, dass die Erwartungen hinsichtlich der Reife der Prozessmodellierung bisher nicht erfüllt werden konnten. Dies wird insbesondere bei der unterschiedlichen Verwendung der Modellierungssprachen deutlich. Obwohl die internationale Forschung BPMN 2.0 als Standard ansieht, konnte sich diese bisher im deutschsprachigen Raum nicht wirklich durchsetzen. Verwunderlich erscheint zudem, dass Unternehmen mit Automatisierungsanliegen und –potenzialen die Abbildung ihrer Prozesse in BPMN 2.0 bisher nicht umgesetzt haben. Die Verwendung nicht maschinenlesbarer Modellierungssprachen und der Verzicht der Abbildung von Mensch-System-Kommunikation

Tab. 1: Zusammenfassende Erkenntnisse der Praxisinterviews

Experte	Zentralisierung	BPMN	Prozesshaus	Qualitätssicherung	Qualitätsmessung
1 - (k)	**0**	**0**	**0**	✓	**0**
2 - (g)	✓	✓	**0**	✓	**0**
3 - (g)	**0**	**0**	✓	✓	**0**
4 - (g)	**0**	✓	**0**	**0**	**0**
5 - (g)	**0**	✓	✓	**0**	**0**
6 - (g)	**0**	✓	**0**	**0**	**0**
7 - (m)	✓	✓	**0**	✓	**0**
8 - (g)	✓	✓	✓	✓	✓
9 - (k)	**0**	**0**	**0**	**0**	**0**
10 - (g)	**0**	**0**	**0**	**0**	**0**
11 - (m)	✓	✓	✓	✓	**0**
12 - (g)	**0**	**0**	✓	✓	**0**
13 - (g)	✓	**0**	✓	✓	**0**
14 - (g)	**0**	**0**	**0**	**0**	**0**
15 - (g)	**0**	**0**	**0**	**0**	**0**
16 - (g)	✓	**0**	**0**	✓	**0**
17 - (g)	✓	**0**	✓	**0**	**0**
18 - (k)	✓	✓	**0**	✓	**0**
19 - (g)	✓	✓	✓	✓	**0**
Σ	**9**	**9**	**8**	**11**	**1**

in Prozessmodellen, erschwert die Automatisierungs-bestrebungen zusätzlich. Ferner ist die dezentrale Modellierung und der Mangel an Prozessmodellqualität ein Problem, das in den Interviews deutlich wird. Eine Vielzahl an Modellierern erschwert die einheitliche Gestaltung von Prozessmodellen und sorgt unter anderem für mehrere Überarbeitungszyklen. Es ist aus den Interviews keine klare Präferenz für eine zentrale, oder dezentrale Modellierungseinheit ablesbar. Dennoch berichteten die Unternehmen mit zentraler, separater Modellierungseinheit von weniger Schwierigkeiten. Zudem wurde ersichtlich, dass sowohl Praxis, als auch Forschung, sich mehr Unterstützung bei der Umsetzung von Prozesshäusern wünschen. Frameworks, wie ARIS, Camunda und ADONIS dienen der Orientierung für den Aufbau einer Prozesslandschaft. Sie ermöglichen aber keine Navigation durch unterschied-

Tab. 2: Zusammenfassende Erkenntnisse der Forschungsinterviews

Experte	Zentralisierung	BPMN	Prozesshaus	Qualitätssicherung	Qualitätsmessung
20	✓	✓	✓	**0**	✓
21	**0**	✓	**0**	**0**	**0**
22	**0**	✓	**0**	**0**	**0**
23	✓	✓	✓	**0**	**0**
24	**0**	✓	**0**	**0**	**0**
25	✓	✓	**0**	✓	✓
26	**0**	**0**	✓	**0**	**0**
∑	**3**	**6**	**3**	**1**	**2**

liche Prozessanwendersichten. Wegen des fehlenden Modellierungsstandards und einer Vielzahl an Modellierern wird die separate Qualitätssicherung häufig zur Pflicht. Dabei wäre ein durchgängiges und schnittstellenübergreifendes Prozess- und Modellierungswissen wünschenswert, um separate Qualitätssicherungseinheiten zu vermeiden. Es bleibt zudem fraglich, auf welcher Basis die Entscheidung für die Modellierung von Prozessen getroffen wird. Das Beispiel eines großen Finanzdienstleisters zeigte beispielsweise, dass IT-Prozesse überhaupt nicht in Prozessmodellen erfasst werden. Da davon auszugehen ist, dass diese als Unterstützungsprozesse dienen und die Prozesslandschaft vervollständigen, wäre ihre Modellierung naheliegend. Es kann somit festgehalten werden, dass die Umsetzung der Prozessmodellierung Verbesserungspotenziale aufweist.

Weiters wird ersichtlich, dass sich die Unternehmensgröße durchaus auf die erfolgreiche Umsetzung der Prozessmodellierung bezüglich der fünf Gestaltungsfaktoren auswirkt. Größere Unternehmen verfügen im Vergleich zu kleinen und mittelständigen häufiger über eine zentrale Modellierungseinheit und sind vertrauter mit der BPMN Darstellung. Sie können ein Prozesshaus im Zweifel selbst implementieren und die ersten Überlegungen hinsichtlich der Messung der Prozessmodellqualität vornehmen. Dies führt dazu, dass bei großen Unternehmen, eher von einer erfolgsversprechenderen Umsetzung der Prozessmodellierung ausgegangen werden kann. Vergleicht man die großen Unternehmen jedoch miteinander, wird deutlich, dass auch hier der Reifegrad voneinander abweicht. Die unterschiedlichen Unternehmensbranchen betreffend, konnte jedoch kein signifikanter Unterschied in Hinblick auf die Gestaltungsfaktoren festgestellt werden.

Vergleicht man das Geschäftsprozessmanagement und insbesondere die Prozessmodellierung mit verwandten Aufgaben wie dem Datenmanagement, so kann vermerkt werden, dass sich letzteres eines höheren Reifegrads erfreut. Die Messung von Daten- und Datenmodellqualität ist in der Praxis verbreitet [Fe09]. Die Situation der Prozessmodellierung lässt sich jedoch mit einem vergleichsweise heterogenem Anwendungsfeld erklären. Die Prozessmodellierung verfügt über zahlreiche Schnittstellen, zum Beispiel zum Projekt- und Qualitätsmanagement und unterliegt somit, sofern keine zentrale Prozessmanagementabtei-

lung vorhanden ist, den Anweisungen der anderen Abteilungen. Dies kann im Zweifel dazu führen, dass unterschiedliche Richtlinien verfolgt werden müssen, was sich wiederum in der Ausgestaltung der Prozesse widerspiegeln kann.

Da die Interviews mit Experten innerhalb des DACH-Gebiets durchgeführt wurden, lassen die Ergebnisse nur Erkenntnisse für den deutschsprachigen Raum zu. Zusammenfassend kann festgehalten werden, dass die Empfehlungen aus der Forschung in puncto Prozessmodellierung nicht immer mit den Umsetzungen in der Praxis übereinstimmen. So wird beispielsweise deutlich, dass der Forschung das Problem des integrativen Prozesshauses zwar bekannt ist, es aber bisher keine durchgängige, leicht implementier- und integrierbare Lösung gibt. Ziel der nächsten Jahre wird es sein, auch in Hinblick auf zunehmende Automatisierungs- und Digitalisierungsbestrebungen, workflow-unterstützende Prozessmodellierungssprachen weiter zu fördern. Auf diese Weise kann auch die Prozessmodellqualität verbessert werden und eine separate Qualitätssicherung obsolet werden. Großer Handlungsbedarf besteht in der Messung der Prozessmodellqualität, eine benutzerfreundliche Lösung aus der Forschung wäre hier wünschenswert. Das wichtigste Ergebnis stellt die Identifikation von fünf Gestaltungsfaktoren dar: Zentrale Modellierungseinheit, Verwendung der BPMN, Aufbau eines Prozesshauses, Separate Qualitätssicherung der Prozessmodelle und Messung von Prozessmodellqualität. Diese fünf Gestaltungsfaktoren bieten eine erste Orientierungshilfe für die Praxis und versprechen mit ihrer Umsetzung positive Auswirkungen auf die Prozessmodellierung.

Literaturverzeichnis

[BDK04] Becker, J., Delfmann, P., Knackstedt, R.: Konstruktion von Referenzmodellierungssprachen. Ein Ordnungsrahmen zur Spezifikation von Adaptionsmechanismen für Informationsmodelle. In: Wirtschaftsinformatik 46(4), 2004, S. 251-264.

[BW14] Bucher, T., Winter, R.: Geschäftsprozessmanagement. Einsatz, Weiterentwicklung und Anpassungsmöglichkeiten aus Methodiksicht. In: HMD Praxis der Wirtschaftsinformatik. 46, 2014, S. 5-16.

[FLZ05] Fettke, P., Loos, P., Zwicker, J.: Business Process Reference Models: Survey and Classification. In: Business Process Management Workshops 2005, S. 469-483.

[Fe09] Fettke, P.: Ansätze der Informationsmodellierung und ihre betriebswirtschaftliche Bedeutung: Eine Untersuchung der Modellierungspraxis in Deutschland. In: Schmalenbachs Zeitschrift für betriebswirtschaftliche Forschung 61, 2009, S. 550-580.

[HC93] Hammer, M., Champy, J.: Reengineering the Corporation. A Manifesto for Business Revolution, New York, 1993.

[Ku11] Kurz, M.: BPM 2.0: Selbstorganisation im Geschäftsprozessmanagement. In: Sinz EJ, Bartmann D, Bodendorf F, Ferstl OK (Hrsg) Dienstorientierte IT-Systeme für hochflexible Geschäftsprozesse. Schriften aus der Fakultät Wirtschaftsinformatik und Angewandte Informatik der Otto-FriedrichUniversität Bamberg, Bd 9. University of Bamberg Press, Bamberg, 2011, S. 193–216.

[Ma15] Mayring, P.: Qualitative Inhaltsanalyse. Grundlagen und Techniken. 12. Aufl., Beltz, Weinheim und Basel, 2015.

[RB05] Rosemann, M., de Bruin, T.: Towards a Business Process Management Maturity Model. In: ECIS 2005 Proceedings of the Thirteenth European Conference on Information Systems, 2005, S. 1-12.

[Ro09] Rohloff, M.: Case Study and Maturity Model for Business Process Management Implementation. In: International Conference on Business Process Management. BPM 2009, S. 128-142.

[RK12] Röglinger, M., Kamprath, N.: Prozessverbesserung mit Reifegradmodellen. Eine Analyse ökonomischer Zusammenhänge. In: Journal of Business Economics 82(5), 2012, S. 509-538.

[VFL10] Vanderhaeghen D., Fettke P., Loos P.: Organisations- und Technologieoptionen des Geschäftsprozessmanagements aus der Perspektive des Web 2.0. Ergebnisse eines gestaltungsorientierten Forschungsansatzes unter besonderer Berücksichtigung von Selbstorganisation und kollektiver Intelligenz. Wirtschaftsinformatik 52(1), 2010, S. 17–32.

[Va10] Van Looy, A.: Does IT matter for business process maturity? A comparative study on business process maturity models. In: Meersman R, Herrero P (Hrsg.) OTM 2010 workshops. Lecture notes in computer science, Bd 6428. Springer, Berlin, 2010, S. 687–697.

[Zu12] zur Muehlen, M.: Interview mit Phil Gilbert über Geschäftsprozessmanagement und Datenintegration. In: Wirtschaftsinformatik 54, 2012, S. 287-289.

Spezifikation, Ausführung und Monitoring von Workflows in verteilten Wissensgraphen (Abstract)

Tobias Käfer[1], Andreas Harth[2]

1 Zusammenfassung

Während BPM-Systeme typischerweise Web-Service-Schnittstellen annehmen, auf welchen Daten im Datenmodell XML übertragen werden, haben sich im Web REST-APIs etabliert, auf welchen Daten in anderen Datenmodellen übertragen werden, beispielsweise Wissensgraphen in RDF. Diese Kombination von RDF mit REST wird in der Literatur häufig mit Read-Write-Linked-Data [Be09] bezeichnet, wobei man Linked Data auch als verteilten Wissensgraphen ansehen kann. Die Systemumgebung Read-Write-Linked-Data unterscheidet sich fundamental von den Annahmen über die Systemumgebung traditioneller BPM-Systeme, da Read-Write-Linked-Data (1) nicht auf der Annahme der Weltabgeschlossenheit (Closed-World Assumption) basiert und (2) ohne Ereignisdaten arbeitet, da Read-Write-Linked-Data auf dem Architekturstil REST basiert. Komponenten mit Read-Write-Linked-Data-Schnittstellen finden sich beispielsweise im kürzlich standardisierten Web der Dinge [Kä20]; durch entsprechendes Lifting der Daten nach RDF kann man auch REST-basierte Microservices [Ne15] als Read-Write-Linked-Data ansprechen.

In [KH18b] präsentieren wir eine Ontologie zur Beschreibung von Workflows, die auf Komponenten ausgeführt werden können, welche mit Read-Write-Linked-Data-Schnittstellen ausgestattet sind. Hierzu geben wir für die Ontologie eine operationale Semantik in ASM4LD an, einem regelbasierten Rechnermodell für Read-Write-Linked-Data [KH18a].

2 Übersicht über Methoden und Ergebnisse

Auf der theoretischen Seite basiert [KH18b] auf formalen Überlegungen zu Workflows und zu den grundlegenden Technologien der Systemumgebung Read-Write-Linked-Data, d. h. REST und RDF. Aus diesen Überlegungen leiten wir für eine Workflow-Ontologie, welche auf dem Refined Process Structure Tree [VVK08] basiert, eine operationale Semantik in Regeln ab und zeigen, dass ihre Ausdrucksmächtigkeit die Basic Workflow Patterns umfasst.

[1] Institut AIFB, Karlsruher Institut für Technologie (KIT), Karlsruhe, tobias.kaefer@kit.edu
[2] Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Nürnberg, andreas.harth@fau.de

doi:10.18420/inf2020_58

Auf der praktischen Seite führen wir die in [KH18b] beschriebenen Regeln auf einer entsprechenden Regelengine [St13] aus. Die praktische Anwendbarkeit des Ansatzes demonstrieren wir anhand eines prototypischen Systems zur Beobachtung von Piloten beim Ausführen von Workflows in Flugzeugcockpits in der Virtuellen Realität. Die Skalierbarkeit unserer Lösung zeigen wir anhand eines synthetischen Benchmarks aus der Gebäudeautomatisierung, in welchem wir ein lineares Verhalten in der Zahl der Geräte beobachten.

3 Diskussion der Ergebnisse

Mit [KH18b] wollen wir zwei Technologien zusammenbringen: Workflows auf der einen Seite, welche eine verständliche Spezifikation von Systemverhalten erlauben, und Linked Data auf der anderen Seite, welches dezentrales Erstellen von Systemkomponenten auf eine Art und Weise erlaubt, die später die Integration mittels formaler Methoden erleichtert.

Diese Kombination ermöglichte uns in öffentlich geförderten Forschungsprojekten zusammen mit der Industrie sowie in Demonstratoren auf internationalen Konferenzen erfolgreich Systeme zu bauen, welche auf flexible und leichtgewichtige Art und Weise orchestriert und integriert wurden.

Literatur

[Be09] Berners-Lee, T.: Read-Write Linked Data, Design Issues, 2009, URL: http://www.w3.org/DesignIssues/ReadWriteLinkedData.html.

[Kä20] Käbisch, S.; Kamiya, T.; McCool, M.; Charpenay, V.; Kovatsch, M., Hrsg.: Web of Things (WoT) Thing Description, Recommendation, W3C, 2020, URL: https://www.w3.org/TR/wot-thing-description/.

[KH18a] Käfer, T.; Harth, A.: Rule-based Programming of User Agents for Linked Data. In: Proceedings of the 11th International Workshop on Linked Data on the Web (LDOW) at the 27th Web Conference (WWW). 2018.

[KH18b] Käfer, T.; Harth, A.: Specifying, Monitoring, and Executing Workflows in Linked Data Environments. In: Proceedings of the 17th International Semantic Web Conference (ISWC). 2018.

[Ne15] Newman, S.: Building microservices. O'Reilly, Sebastopol (CA), USA, 2015.

[St13] Stadtmüller, S.; Speiser, S.; Harth, A.; Studer, R.: Data-Fu: A Language and an Interpreter for Interaction with Read/Write Linked Data. In: Proceedings of the 22nd International Conference on World Wide Web (WWW). 2013.

[VVK08] Vanhatalo, J.; Völzer, H.; Koehler, J.: The Refined Process Structure Tree. In: Proceedings of the 6th International Conference on Business Process Management (BPM). 2008.

R. Reussner, A. Koziolek, R. Heinrich (Hrsg.): INFORMATIK 2020,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2021 665

Towards an Automated Assessment of Graphical (Business Process) Modelling Competences: A Research Agenda

Michael Striewe[1], Constantin Houy[2], Jana-Rebecca Rehse[3], Meike Ullrich[4], Peter Fettke[2], Niclas Schaper[5], Andreas Oberweis[4]

Abstract: In Business Process Management (BPM), graphical process modelling plays an important role because process models can significantly support BPM endeavors in many different ways. Hence, it is also crucial in the context of learning and teaching BPM. Graphical modelling in general is also a curricular core component of higher education in related disciplines such as information systems engineering or software engineering. There are numerous concepts and tools which support learning and teaching of graphical modelling, but most of them have so far been isolated from each other and are used only locally. In order to increase the quality of learning and teaching in modelling courses, it is desirable to better integrate these approaches and identify commonalities. This paper discusses several challenges of integrating learning and teaching approaches for graphical modelling and outlines an approach to a solution that is currently pursued in the ongoing research project KEA-Mod.

Keywords: Business Process Modelling; Modelling education; Educational technology

1 Introduction

Graphical process modelling plays an important role in Business Process Management (BPM) as many different functionalities supporting successful BPM are based on process models. Hence, process modelling is a crucial aspect for the qualification of BPM experts. In that context, teaching graphical modelling approaches, such as data modelling, is also relevant because they provide insights into the manner of thinking and designing artifacts of computer science-related disciplines. In theses disciplines, graphical modelling is an essential core component of the curriculum, e. g., [As18]. Commonly used languages are Petri nets, Event-driven process chains (EPC), Business Process Model and Notation (BPMN), Entity-Relationship diagrams (ERD), or Unified Modeling Language (UML).

Due to the complexity of the topic, courses with a high number of participants cannot provide individual support in solving modelling tasks, as there are often multiple correct

[1] Universität Duisburg-Essen, michael.striewe@s3.uni-due.de
[2] Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI) und Universität des Saarlandes, constantin.houy@dfki.de, peter.fettke@dfki.de
[3] Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI) und Universität Mannheim, rehse@uni-mannheim.de
[4] Karlsruher Institut für Technologie, meike.ullrich@kit.edu, andreas.oberweis@kit.edu
[5] Universität Paderborn, nschaper@mail.uni-paderborn.de

doi:10.18420/inf2020_59

solutions. This makes individual coaching with individual feedback on the modelling solution especially important. Digital exercises and exams as well as automated feedback generation for models have been developed and used at various universities over the past decade (e.g., [SG14; Th16]). However, up to now, most of them have been isolated solutions tailored to the individual situation of each university and course. Moreover, many solutions focus on plain model creation tasks, while learning objectives at higher levels, which emphasize competence orientation, are rarely considered appropriately.

We see this as a major obstacle in improving the quality of teaching and learning. Hence, we started a collaborative research project to set up a methodically grounded, integrated e-assessment platform that can be applied to multiple assessment scenarios. This includes the support of different graphical modelling languages, task types, and levels of modelling expertise. The platform will merge existing technical applications into a uniform concept that can be adapted for different courses in specific contexts. The platform shall be designed to enable (i) the extension towards additional modelling languages and (ii) an individual adaptation to the requirements of the respective institution and course by means of open interfaces, parameterisation and modularisation. The purpose of the paper at hand is to shed light onto the current state of research in the field of e-assessment for graphical modelling and to outline the agenda underlying the ongoing research project KEA-Mod.[6]

2 Conceptual Foundations

Designing and understanding models are core competences of graphical modelling. According to learning objectives for modelling in university education, students should learn "both to read and understand existing models and to build models" and "learn suitable modelling languages to the extent that they are able to apply the conceptual knowledge about modelling"[Gl08, p. 427]. In the taxonomy of Anderson and Krathwohl (revised according to Bloom) [An00], those objectives can be classified into the categories "applying" and "analysing" that form the intermediate level of competences. We also consider verifiable learning objectives for modelling at the two highest levels, "Evaluate" and "Create", of the taxonomy as most relevant, since they appropriately reflect the actual practice of modelling.

Tasks on *model understanding* usually present complete models to students and ask different questions about them. Examples are questions about the correctness of a model regarding the syntax or about the meaning, i. e., whether a certain statement is actually covered by the model. For automated assessments, such task types can be implemented with multiple choice questions. The advantage is that student answers can be checked automatically and that an automatic generation of possible answers is feasible. This allows for the automation of the exercise creation, so that individualized, adaptive examination formats can be enabled.

[6] Kompetenzorientiertes E-Assessment für die grafische Modellierung (KEA-Mod), gefördert vom BMBF (FKZ: 16DHB3022-16DHB3026), `https://keamod.gi.de/`

Tasks on *model creation* typically provide a textual description of a situation. Students have to translate that into a graphical model using a modelling language. In addition to mastering the modelling language, cognitive skills such as text comprehension and abstraction are essential for solving the task. Grading such tasks manually is at least time-consuming if not also error-prone due to inconsistencies or accidental mistakes. A (partially) automated check can reduce the error rate, ensure transparent and consistent feedback, and save a lot of time. However, these benefits are based on formal digital models and cannot be achieved with hand-drawn diagrams. In current teaching and assessment methods, solutions are still mostly drawn on paper. Hence, they are often less readable and the manual correction effort for the grader is additionally increased.

A wide range of tools is already available for the digital creation of models. Most of those tools include support functionalities that check essential syntactic rules of the modelling languages, making it difficult or even impossible to violate these rules. Since teachers might be interested in assessing their students' knowledge of those syntactical modeling rules, they cannot use any existing tools for this learning objective (this would be analogous to testing the students' knowledge of spelling rules using a tool with automatic spell checking). Hence, any input editors used for teaching purposes must be limited in this respect.

The *competence-orientation* of the curricula reflects a central content-related and didactic claim of the reformed bachelor's and master's programmes. The course design should not only be based on a clearly defined competence profile and corresponding goals, which particularly represent the professional requirements of graduates, but should also include teaching and learning methods designed towards the acquisition of competence and corresponding examination tasks and formats. However, this requirement is often met only rudimentarily in the conception and implementation of the curriculum. So far, few studies have been conducted that explicitly deal with a subject-related didactic competence model for modelling in higher education teaching. In a superordinate competence framework model, the competence to model information systems is of central importance, although requirements for the use of modelling languages and tools have not been specified in detail so far in this model [Li13]. Regarding this goal, a discussion of curriculum development and learning objectives for modeling in university teaching [Gl08] has been published.

3 Existing technical solutions

In general, e-assessment systems for complex examination formats are already used in many places in the higher education sector, e. g., for checking source code in programming tasks [KJH18]. Although different approaches for computer-aided analysis of student solutions for graphical modelling are described in literature, their use in higher education is not yet widespread. While implementations of e-assessments for graphical modelling have been known for almost 20 years [Ts02], they are primarily concerned with the basic feasibility of e-assessments for this type of artefact and neither with the subject-specific challenges of certain model types nor with the direct connection to a competence-oriented teaching

and examination concept. These early approaches demonstrate the inherent structural commonality of the different model and diagram types used for graphical modelling (e. g., ER, UML, BPMN, EPC, Petri nets): each of them is a specific graph in which different node elements are connected to each other via edges. These common properties can be used to develop a cross-language technical solution. Previous work demonstrated that the transferability of existing approaches to other model types of the same nature (graph structure) can be implemented with little effort [Fe16].

Since the basic technical feasibility has been established, current research focuses on three remaining technical challenges. First, there is an inherent ambiguity of modelling problems that allow for more than one correct solution. In order to be able to accept different correct solutions, current systems do not only rely on the fixed comparison of a student solution with sample solutions, but use rule-based approaches [SG11]. Solution specifications are divided into smaller sections, whose correct implementation can be checked individually and provided with feedback. Second, also a linguistic ambiguity of model element labels can cause uncertainties in the semantic interpretation. Current approaches take synonyms, typing errors, abbreviations, and alike into account or apply distance measures, text transformations, and other methods of natural language processing to deal with such ambiguities. Finally, execution semantics have to be considered for model or diagram types for which a purely static examination of semantics is not appropriate. To test the execution semantics of models or diagrams, similar techniques as in the checking of programming tasks can be used. Depending on the model or diagram type, all conceivable or only selected sequences can be simulated and compared with a specification [SG14].

Among the different diagram types used in teaching, the automated assessment of ER and UML class diagrams in particular has attracted the attention of researchers (e. g. [WE15]). A web-based software prototype based on the tool *RefModMiner* enables the automated assessment of EPC diagrams [Th16]. The automated assessment of student business process models in the form of Petri nets can be coupled with user feedback and a gamification approach [Pf16].

4 Remaining challenges and questions to answer

In order to achieve progress beyond the current state, several questions must be answered before an improved e-assessment platform can be developed and applied in regular lectures. The *first question* addresses the relation between required modelling competences in professional tasks and the learning objectives in university curricula. Looking only at one of the two sides would risk missing relevant aspects. We assume that reviewing existing research on professional modelling activities and curricula can contribute to finding answers. We plan to use questionnaires to gather additional information if necessary.

The *second question* aims at finding out which competence-oriented formats and task types are actually used at which point in the teaching/learning cycle to promote certain

competences. It has to be captured which subject-related cognitive performance and competences are addressed by identified task types. Here, several task types might be suitable to measure a specific competence. At the same time, the complexity of modelling tasks makes it very unlikely that one task type corresponds to exactly one single competence. Hence, inferences and dependencies must be carefully analysed. To this effect, examination and exercise material currently used in higher education has to be reviewed.

The *third question* aims at finding out which task types address higher cognitive levels (applying, analysing, evaluating, creating) in an appropriate way. Only a certain subset of all potential modeling task types will be integrated into the e-assessment platform. However, not every relevant task type might be technically suitable for automation, so a further challenge is to find out which competences can be promoted by task types that can be (i) individually and adaptively generated and (ii) completely automatically checked. Notably, it is not the goal of the research project to change exercises so that they fit the current capabilities of existing e-assessment tools. Instead, the goal is to extent e-assessment features to cover as much of all relevant exercise types as possible. Potentially, the answer to the third question reveals that existing competence-oriented assessments are not suitable or sufficient regarding their relevance to learning objectives and competences. Then an additional question has to be answered on how to change or modify assessment procedures/criteria in order to establish the desired correspondence between learning objectives and competences. With the developed list of competences and related task types, the final question is how to design automated individual and elaborated feedback for various learning scenarios.

System usability is a further important topic. Since students and lecturers are two equally important user groups with different requirements, a careful analysis based on media pedagogy and instructional design should be carried out. From a media didactic point of view, it can also be important to enable the platform functionality to be embedded in the existing technical infrastructure of the respective university. In order to achieve this, the platform will be implemented on the basis of a micro-service system architecture, enabling the integration of existing technical solutions. In addition, the basic requirements for a modular adaptation and maintenance are given, such that small and clearly arranged system components can be extended or exchanged without affecting the functionality of the overall system. Cloud-based implementation also allows access to the platform independent of location and time. The digital processing of student solutions on the platform also opens up the possibility of learning process analyses in the sense of *learning analytics*.

5 Conclusion

We have identified several open questions that need to be answered on the way towards an automated assessment of graphical modelling competences which can significantly contribute to the education of future BPM experts. While there is a lot of promising research on the technical aspects of model analysis, feedback generation and exercise generation, a strong connection to competence measurement is still missing. Against this background,

we are currently working with teachers, practitioners and fellow researchers towards the creation and validation of a suiting competence model and related assessment tools which shall be integrated in an overarching platform for automated assessment of graphical models.

References

[An00] Anderson, L. W.; Krathwohl, D. R.; Airasian, P. W.; Cruikshank, K. A.; Mayer, R. E.: A Taxonomy for Learning, Teaching, and Assessing. Pearson Education (US), 2000.

[As18] Association for Computing Machinery (ACM): Curricula Recommendations, Online [https://www.acm.org/education/curricula-recommendations], 2018.

[Fe16] Fellmann, M.; Fettke, P.; Houy, C.; Loos, P.; Oberweis, A.; Schoknecht, A.; Striewe, M.; Thaler, T.; Ullrich, M.: Evaluation automatisierter Ansätze für die Bewertung von Modellierungsaufgaben. In: DeLFI 2016 - Die 14. E-Learning Fachtagung Informatik. Pp. 203–214, 2016.

[Gl08] Glinz, M.: Modellierung in der Lehre an Hochschulen: Thesen und Erfahrungen. Informatik-Spektrum 31/5, pp. 425–434, 2008.

[KJH18] Keuning, H.; Jeuring, J.; Heeren, B.: A systematic literature review of automated feedback generation for programming exercises. ACM Transactions on Computing Education (TOCE) 19/1, pp. 1–43, 2018.

[Li13] Linck, B.; Ohrndorf, L.; Schubert, S.; Stechert, P.; Magenheim, J.; Nelles, W.; Neugebauer, J.; Schaper, N.: Competence model for informatics modelling and system comprehension. In: 2013 IEEE Global Engineering Education Conference (EDUCON). Pp. 85–93, 2013.

[Pf16] Pflanzl, N.: Gameful Business Process Modeling. In: Workshop on Enterprise Modeling and Information Systems Architecures (EMISA). Pp. 17–20, 2016.

[SG11] Striewe, M.; Goedicke, M.: Automated checks on UML diagrams. In: 16th Annual SIGCSE Conference on Innovation and Technology in Computer Science Education (ITiCSE). Pp. 38–42, 2011.

[SG14] Striewe, M.; Goedicke, M.: Automated Assessment of UML Activity Diagrams. In: 19th Annual SIGCSE Conference on Innovation and Technology in Computer Science Education (ITiCSE). P. 336, 2014.

[Th16] Thaler, T.; Houy, C.; Fettke, P.; Loos, P.: Automated Assessment of Process Modeling Exams: Basic Ideas and Prototypical Implementation. In: Modellierung 2016 Workshopband. LNI, pp. 63–70, 2016.

[Ts02] Tsintsifas, A.: A Framework for the Computer Based Assessment of Diagram Based Coursework, PhD thesis, University of Nottingham, 2002.

[WE15] Weerasinghe, A.; Evans, B.: UML-IT: An ITS to Teach Multiple Modelling Tasks. In: Artificial Intelligence in Education (AIED). Pp. 816–819, 2015.

R. Reussner, A. Koziolek, R. Heinrich (Hrsg.): INFORMATIK 2020,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2021 671

Agilität im Geschäftsprozessmanagement

Eine systematische Literaturanalyse

Janek Ziehmann[1], Birger Lantow[2]

Abstract: Wollen sich Unternehmen im Zuge der Digitalisierung der damit verbundenen Dynamik stellen, reicht es nicht, agile Ansätze nur in der IT-Entwicklung anzuwenden. Auch das Management der Geschäftsprozesse sollte der Dynamik Rechnung tagen. Diese Arbeit untersucht auf Basis einer systematischen Literaturanalyse, welche Ansätze zur Umsetzung von Agilität im Geschäftsprozessmanagement vorgeschlagen werden. Der aktuelle Stand der Wissenschaft auf diesem Gebiet wird systematisiert und zukünftig notwendige Entwicklungen werden abgeleitet.

Keywords: Geschäftsprozessmanagement; Business Process Management; Agilität; Social BPM

1 Einleitung

Im Zuge der Digitalisierung basieren immer mehr Geschäftsprozesse ganz oder teilweise auf IT-Komponenten. Die unternehmensinterne und -externe Bereitstellung der IT folgt dabei häufig dem Dienstleistungsparadigma. IT-Dienstleistung bzw. IT-Services stellen IT-Funktionalität zur Unterstützung eigener Geschäftsprozesse oder von Kundenprozessen bereit. Dabei ist die Bereitstellung von IT-Services selber an Geschäftsprozesse gebunden, welche diese ermöglichen und unterstützen. In der IT-Entwicklung haben sich in den letzten Jahren agile Methoden etabliert. Zusammen mit aktuellen Ansätzen wie DevOps und Continuous Delivery ergibt sich eine hohe Flexibilität in der Bereitstellung von IT-Funktionalität. Außerdem ist ein organisatorisches Umfeld zu schaffen, welches der Softwareentwicklung als wissensintensives Tätigkeitsfeld Rechnung trägt. Hier ist ein geeigneter Umgang des Geschäftsprozessmanagements mit wissensintensiven Prozessen notwendig.

Um die bestehenden Anforderungen im Kontext von IT-Services umsetzen zu können, stellt sich die Aufgabe der Integration von Agilität in das Geschäftsprozessmanagement bzw. Business Process Management (BPM) der verbundenen Geschäftsprozesse. [Ba19] definieren das BPM als agil, „[...] wenn es das Bemühen um Wertschöpfung beinhaltet, indem es dazu beiträgt, Veränderungen zu schaffen und/oder proaktiv im Vorfeld auf

[1] Universität Rostock, Lehrstuhl für Wirtschaftsinformatik, Albert-Einstein-Str. 22, D-18051 Rostock, janek.ziehmann@uni-rostock.de
[2] Universität Rostock, Lehrstuhl für Wirtschaftsinformatik, Albert-Einstein-Str. 22, D-18051 Rostock, birger.lantow@uni-rostock.de

doi:10.18420/inf2020_60

Veränderungen zu reagieren und/oder aus Veränderungen zu lernen, während es gleichzeitig zur Wahrnehmung der Wirtschaftlichkeit, Qualität und Einfachheit eines Prozesses durch die Verbraucher beiträgt und diese nicht beeinträchtigt. Darüber hinaus sollten die ergriffenen Maßnahmen kontinuierlich sein und ein Minimum an Zeit und Kosten erfordern“. Im Vergleich zur systematischen Literaturanalyse von Badakhshan et al. [Ba19] zeichnet sich die vorliegende Arbeit durch den Einbezug anderer Literaturdatenbanken und eine andere inhaltliche Ausrichtung aus. In der Konsequenz decken sich die gefundenen, relevanten Quellen der beiden Arbeiten nur teilweise. Bei der Auswertung der Quellen liegt in dieser Arbeit der Fokus auf der Umsetzung von Prinzipien des agilen BPMs und Kontexte in denen ein agiles BPM erfolgreich sein kann. Die Studie von Badakhshan et al. fokussiert eher auf ein theoretisches Rahmenwerk zur Einordnung von Forschungsarbeiten auf dem Gebiet des agilem BPM.

Ziel dieser Arbeit ist es, auf Basis einer systematischen Literaturanalyse zu untersuchen, welche Ansätze existieren, um agile Prinzipien oder Agilität auf das BPM zu übertragen. Dabei sollen folgende Forschungsfragen beantwortet werden:

FF1: Welche Prinzipien für agiles BPM werden in der Forschung vorgeschlagen?

FF2: In welchem Kontext wird die Anwendung von agilem BPM vorgeschlagen?

FF3: Was sollten künftige Forschungsschwerpunkte bezüglich agilem BPM sein?

Im weiteren Verlauf beschreibt Abschnitt 2 die angewendete Vorgehensweise der Literaturrecherche und stellt sowohl den Suchterm, als auch die gefundenen Ergebnisse dar. Darauf aufbauend werden in Abschnitt 3 Prinzipien aus der Literatur abgeleitet, die den in der Literatur betrachteten Ansätzen zugrunde liegen und so die Forschungsfrage FF1 beantwortet. Die Forschungsfrage FF2 wird darauffolgend in Abschnitt 4 betrachtet und analysiert den Anwendungskontext von agilem BPM. Da es sich bei dem agilen BPM um ein relativ junges Forschungsthema handelt, werden im Abschnitt 5 anhand der Forschungsfrage FF3 zukünftige Forschungsschwerpunkte betrachtet und ein Überblick über den Veröffentlichungszeitraum gegeben. Zusammenfassung und Ausblick finden sich dann im letzten, sechsten Abschnitt.

2 Systematische Literaturanalyse

Die Literaturrecherche soll einen Überblick über die in der Wissenschaft existierenden Ansätze zu agilem BPM generieren und den Ansätzen agilen BPM zugrunde liegende Prinzipien betrachten.

Da das BPM ein sehr weitreichendes Themenfeld ist und somit viele Facetten und Forschungsrichtungen bietet ist die systematische Literaturrecherche auf Basis der Vorgehensweisen von [Ki04] und [Le06] durchgeführt worden. Zu Beginn wurde relevante Literatur mit Hilfe eines aufgestellten Suchterms in verschiedenen Datenbanken identifiziert. Aufgrund der

Annahme eines noch recht jungen Themafelds wurde die Identifikation relevanter Literatur durch eine Volltextsuche in folgenden Datenbanken durchgeführt: Scopus, Sciencedirect, SpringerLink und IEEE.

Aus diesen Suchergebnissen wurde anhand der Titel der Ergebnisse die als relevant empfundene Literatur herausgefiltert. Darauffolgend ist eine Selektion der Literatur mit Hilfe der jeweiligen Abstracts durchgeführt worden. Handelte es sich um ein Buch oder ein Kapitel eines Buches, zu welchem kein Abstract existiert, so wurde das Inhaltsverzeichnis systematisch durchgegangen und per Volltextsuche im Dokument die Begriffe des Suchterms ausfindig gemacht. Die so selektierte Literatur diente als Grundlage zur Datenextraktion mit Bezug auf die oben genannten Forschungsfragen. Zur Identifikation relevanter Literatur ist folgender Suchterm verwendet worden:

„agiles Geschäftsprozessmanagement" OR „agiles BPM" OR „agiles Business Process Management" OR „agile Geschäftsprozesse" OR „agile BPM" OR „agile Business Process Management"

Der Suchterm enthält sowohl die deutschen, als auch die englischen Formulierungen, um ein möglichst breites Spektrum an Literatur zu erhalten. Agilität stellt ein internationales Thema dar und wird so auch viel im nicht deutschsprachigen Raum behandelt.

Die initiale Suche in den aufgeführten Datenbanken ergab kombiniert eine Anzahl von 371 Treffern. Durch die Eliminierung von Duplikaten und die Selektion der Ergebnisse anhand ihrer Titel ergab sich eine Summe von 90 Treffern. Bei der Selektion anhand der Titel musste entweder der Begriff „Agilität" o.ä. oder der Begriff „Geschäftsprozessmanagement" o.ä. bzw. eine Kombination beider Begriffe oder deren englischsprachigen Synonyme im Titel vorkommen. Diese Selektionskriterien sind ebenfalls bei der Selektion der 90 Ergebnisse anhand ihrer Abstracts angewandt worden. Bei Buchkapiteln oder Büchern ohne Abstract wurden die Selektionskriterien im Inhaltsverzeichnis gesucht und durch eine Volltextsuche im Dokument ausfindig gemacht, um ihren Kontext bei der Selektion berücksichtigen zu können. Da es in Einzelfällen vorgekommen ist, dass der Abstract auf Englisch, aber der weitere Text in anderen Sprachen formuliert war, mussten diese Ergebnisse zusätzlich aussortiert werden. Durch diesen Selektionsschritt ist eine Eingrenzung auf 42 Treffer erzielt worden. Die Literatur aus dieser Ergebnismenge wurde daraufhin inhaltlich betrachtet und auf ihre Relevanz bezüglich der formulierten Forschungsfragen untersucht und bewertet. Hieraus resultierte eine endgültige Ergebnismenge, welche zur Beantwortung der formulierten Forschungsfragen genutzt wird. Die endgültige Ergebnismenge besteht nach den Selektionsschritten aus insgesamt 17 Treffern. Die genaue Zusammensetzung der Ergebnismenge sowie die Verweise der Elemente der Ergebnismenge auf das Literaturverzeichnis sind in der Tabelle 1 übersichtlich dargestellt.

In der Literatur ließen sich grundlegende Ansätze finden, auf welche sich jeweils bezogen wurde. Eine Zuordnung von Ansätzen und Quellen ist in Tabelle 2 zu finden. Auffällig ist dabei die große Zahl an Publikationen zum Thema Social BPM (SBPM).

Tab. 1: Suchverlauf und Selektion relevanter Literatur

Datenbank	**#Initial**	**#Selektion**	**#Final**	**Quellen**
Scopus	31	11	6	[Tr18, Ro18, Ba19, Me14a, Me11, Me14b, Br11]
ScienceDirect	19	7	6	[Ra19, Ro15, Pa16, Za17, Ma17]
SpringerLink	322	22	4	[Ra15, Ko13, Pa11, Th13]
IEEE	7	2	1	[Sc07]

Tab. 2: Zuordnung von BPM-Ansätzen und Literaturquellen

Ansatz	Quellen
Social BPM	[Ra19, Tr18, Ra15, Ro15, Br11, Me14b, Pa11, Me11, Ro18]
BPM&ACM	[Ko13]
Sense-and-Respond	[Sc07]
SSCRUM und BPM	[Th13, Pa16]
BPPAM	[Ma17, Za17]
$BPMN^{easy}$	[Me14a]
Technische Unterstützung	[Ba19]

In den nachfolgenden Abschnitten werden die Ergebnisse der Literaturanalyse dargestellt und so die formulierten Forschungsfragen FF1 bis FF3 beantwortet. Abschnitt 3 fokussiert dabei zunächst auf die Identifikation von generalisierbaren Prinzipien für agiles BPM in den einzelnen Ansätzen.

3 Prinzipien agilen Geschäftsprozessmanagements

Anhand der Literatur sowie der verschiedenen Ansätzen zu agilem BPM lassen sich unterschiedliche Prinzipien ableiten, an denen die agilen BPM-Ansätze beim klassischen BPM ansetzen und dieses verändern. Ein Prinzip stellt dabei ein grundlegendes Element der Ansätze zum agilen BPM dar, welches zumindest theoretisch Bestandteil mehrerer Ansätze sein kann. Jeder Ansatz kann mehrere Prinzipien einbeziehen und ist nicht zwingend auf ein Prinzip beschränkt. Anhand der Prinzipien verfolgen die jeweiligen Ansätze unterschiedliche Ziele. Bezieht ein Ansatz mehrere Prinzipien mit ein, so sollten sich die zugrundeliegenden Prinzipien in ihren Zielen nicht widersprechen. Sie können sich jedoch gegenseitig verstärken, sofern sie dasselbe Ziel verfolgen. Eine Übersicht über die in den jeweiligen Quellen behandelten Prinzipien des agilen BPMs bietet die Tabelle 3. Folgende Prinzipien werden in der Tabelle 3 betrachtet, aus den Quellen abgeleitet und nachfolgend näher erläutert:

1. Veränderungen am BPM-Lebenszyklus
2. Vereinfachte Notationen

3. Einbeziehung von agilen Rollen
4. Verstärkung der Kollaboration
5. Einbindung von Wissensmanagement (zur Wissenssicherung und -vervielfältigung)
6. Prozessanalyse mit Hilfe moderner Technologien

Die Prinzipien sind dabei aus getätigten Vergleichen der vorhandenen Literatur abgeleitet und adaptiert worden oder auf Gemeinsamkeiten der betrachteten Ansätze zurückzuführen. Dabei stellen die Prinzipien Gesetzmäßigkeiten dar, mit Hilfe derer die Agilität im BPM-Kontext vorangetrieben wird.

Die Prinzipien (1) Veränderungen am BPM-Lebenszyklus, (2) Vereinfachte Notationen und (5) Einbindung von Wissensmanagement lassen sich auf [Ma17] zurückführen. Für den Vergleich von Geschäftsprozess-Methoden, wie unter anderem AGILIPO (agile business process) oder BPPAM (Business Process and Practice Alignment Methodology), werden durch [Ma17] die Vergleichspunkte Lebenszyklus, Modellierung und Wissen herangezogen. Das Prinzip der (1) Veränderung am BPM-Lebenszyklus bezieht sich grundsätzlich auf den gesamten Lebenszyklus und wird durch die Literatur auf unterschiedlich granularen Ebenen betrachtet. Durch die Veränderungen am BPM-Lebenszyklus werden beispielsweise neu entwickelte iterative Vorgehensweisen oder ad-hoc Prozesse in den BPM-Lebenszyklus integriert. Dieses Prinzip lässt sich in verschiedenen Ansätzen des agilen BPMs, wie beispielsweise dem BPM(N)$^{\text{Easy1.2}}$ wiederfinden und verfolgt differenzierte Ziele. Zu den verfolgten Zielen kann die stärkere Einbindung von Stakeholdern oder auch die Sicherung der Qualität der entstandenen Modelle zählen. Das Prinzip (2) Vereinfachte Notationen wird im Gegensatz zum BPM-Vorgehen nur sehr vereinzelt in der Literatur aufgegriffen und zudem lediglich stark vereinfacht dargestellt. Im BPM-Lebenszyklus lässt sich das Prinzip der vereinfachten Notation hauptsächlich dem Schritt der Definition/Modellierung von Geschäftsprozessen zuordnen und kommt dabei hauptsächlich im Ansatz des BPM(N)$^{\text{Easy1.2}}$ zum Tragen, wobei auch hier keine spezifischen Änderungen der Notation aufgezeigt werden. Durch das Prinzip der vereinfachten Notationen sollen die Kommunikation und das Prozessverständnis der am Prozess beteiligten Stakeholder gefördert werden, was wie bereits beim BPM-Vorgehen in einer stärkeren Einbindung der Stakeholder münden kann. Das Prinzip (5) Einbindung von Wissensmanagement lässt sich im BPM-Lebenszyklus verstärkt den Schritten der Definition/Modellierung von Geschäftsprozessen sowie der Prozessimplementierung zuordnen, da sich hier die größten Effekte für ein Wissensmanagement ergeben. In der Literatur wird das Wissensmanagement hauptsächlich mit den SBPM-Ansätzen behandelt, aber auch im Zusammenhang mit den Ansätzen der BPPAM oder der Kombination von BPM und modernen Technologien berücksichtigt. Als Ziel der Einbindung von Wissensmanagement können die Verbesserung der Kommunikation eingebundener Stakeholder oder auch die Förderung des Prozessverständnisses betrachtet werden.

Das Prinzip (3) Einbeziehung von agilen Rollen kann im weitesten Sinne auf [Za17] zurückgeführt werden, welcher in dem durchgeführten Vergleich von Geschäftsprozess-Konzepten unter dem Konzept der Prozessressource die Rollen ausmacht. Rollen können daher für das BPM als relevant angesehen werden und sollten in Bezug auf agiles BPM in Form eines Prinzips berücksichtigt werden. Dieses Prinzip lässt sich im BPM-Lebenszyklus nicht klar zuordnen, dient dabei aber meist der Zuordnung von Verantwortlichkeiten der beteiligten Stakeholder und kann so die Kommunikation fördern. Das Prinzip der Einbindung von Rollen wird beispielsweise durch den agilen BPM-Ansatz zur Kombination von BPM und SCRUM berücksichtigt.

Das Prinzip (4) Verstärkung der Kollaboration kann nur bedingt auf Quellen zurückgeführt werden. [Im19] führen Kollaboration als eine zu betrachtende Thematik für BPM an, beschreiben dies aber nicht detaillierter. Das Prinzip ist hier aufgeführt, da besonders durch den Ansatz des SBPM, aber auch durch andere Ansätze, die Kollaboration im BPM-Kontext von Interesse ist und verstärkt betrachtet wird. Zudem ist eine stärkere Kollaboration besonders im Kontext agiler Methoden von Bedeutung. Aus besagten Gründen ist eine verstärkte Kollaboration trotz mangelnder Quellen als Prinzip von agilem BPM anzusehen. Als Ziele können die Verbesserung der Kommunikation sowie eine stärkere Einbindung beteiligter Stakeholder angesehen werden. Dabei lässt sich auch dieses Prinzip nicht klar einzelnen Schritten des BPM-Lebenszyklus zuordnen.

Neben Prinzipien der agilen BPM-Ansätze, die stark auf Stakeholder sowie deren Kommunikation oder das generelle Prozessverständnis ausgerichtet sind, gibt es in der Literatur auch ein Prinzip, das grundsätzlich keine methodischen Änderungen des BPM-Lebenszyklus erfordert und sich so auch nicht auf diesen auswirkt. Das Prinzip (6) Prozessanalyse mit Hilfe moderner Technologien lässt sich auf den Ansatz nach [Ba19] zurückführen. Dies ist ein noch recht junger Ansatz und im BPM-Kontext der Einzige, welcher sich mit dem Einsatz von Technologien zur agilen Gestaltung von BPM auseinandersetzt. Da das derzeitige Geschäftsumfeld einer starken Dynamik unterliegt und auf Optimierung sowie Automatisierung ausgerichtet ist und vermehrt Technologien zur Geschäftsprozess-Unterstützung Einfluss erhalten ist das Prinzip der Prozessanalyse mit Hilfe moderner Technologien als ein für die Zukunft relevantes Prinzip zu betrachten. Moderne Technologien erhalten derzeit in vielen Bereichen vermehrt Einfluss und werden sich auch auf das traditionelle BPM auswirken. Daher ist die Prozessanalyse mit Hilfe moderner Technologien als Prinzip agilen BPMs anzusehen und im Zuge künftiger Forschungen auf andere Teilbereiche als die Prozessanalyse im BPM-Kontext auszuweiten. Das beschriebene Prinzip lässt sich derzeitig im BPM-Lebenszyklus klar dem Schritt der Prozessanalyse zuordnen. Für die Einbindung moderner Technologien lässt sich als Ziel die Digitalisierung, Automatisierung und Optimierung der Prozessabläufe ausgeben.

Durch die Beschreibung der Prinzipien wird deutlich, dass die Kommunikation oder auch die Einbindung der Stakeholder elementare Ziele der Veränderungen von traditionellem BPM zu agilem BPM sind. Dies lässt sich auch darauf zurückführen, dass Agilität im Allgemeinen einen hohen Grad an Kommunikation und Kollaboration erfordert. Besonders

zum Tragen kommt der Aspekt der Kollaboration und Kommunikation in den agilen Methoden wie beispielsweise SCRUM, in denen zum Teil selbstorganisierte Teams für die Projektumsetzung zuständig sind.

Die behandelten Prinzipien und deren Berücksichtigung in den verschiedenen Quellen sind in der Tabelle 3 übersichtlich dargestellt. Prinzipien, welche in der jeweiligen Quelle berücksichtigt wurden sind gekennzeichnet (•). Eine detailliertere Unterscheidung in weiteren Abstufungen der Berücksichtigung wird hier nicht vorgenommen, da es sich um Ansätze des agilen BPMs handelt, die für eine Überführung in ein methodisches Vorgehen weiterer Betrachtung bedürfen und die Prinzipien teils nur sehr allgemein oder stark vereinfacht betrachten.

Tab. 3: Prinzipien des agilen BPM

Quellen	Veränderungen am BPM-Lebenszyklus	Vereinfachte Notation	Einbeziehung von agilen Rollen	Verstärkung der Kollaboration	Einbindung von Wissensmanagement	Prozessanalyse mit Hilfe moderner Technologien
[Ra19]				•	•	
[Tr18]				•	•	
[Ra15]		•	•	•		
[Ro15]	•			•		
[Br11]	•			•	•	
[Me14b]	•					
[Pa11]				•		
[Me11]	•			•		
[Ko13]				•	•	
[Ro18]	•			•		
[Sc07]	•					
[Th13]	•		•	•		
[Pa16]	•		•			
[Ma17]	•			•	•	
[Za17]				•		
[Me14a]	•	•		•		
[Ba19]	•			•		•

Für eine erfolgreiche Etablierung sowie ein erfolgreiches Betreiben von agilem BPM ist das Ausschöpfen der Potenziale der verschiedenen Ansätze notwendig. Dabei sollte sich neben den reinen Potenzialen und Zielen der Ansätze auch verstärkt mit den Zusammenhängen und Gemeinsamkeiten beschäftigt werden. Verschiedene Ansätze kombiniert und so die Kombination von noch mehr Prinzipien im Rahmen eines Ansatzes für agiles BPM können eine noch größere Wirkung entfalten und sollten daher verstärkt im Mittelpunkt der Forschung liegen.

4 Kontext agilen Geschäftsprozessmanagements

Auch wenn Agilität, und besonders Agilität im BPM-Kontext, derzeit einen Trend darstellt, ist es nicht immer sinnvoll, ein von Agilität beeinflusstes BPM zu betreiben. Nachfolgend wird das Thema beleuchtet, wann ein agiles BPM sinnvoll ist und so auf die Forschungsfrage FF2 eingegangen.

Angesichts der vermehrt zunehmenden Dynamik in der heutigen Geschäftsumgebung steht der traditionelle BPM-Ansatz regelmäßig vor der Herausforderung, schnell auf Veränderungen zu reagieren und bestehende Prozesse anzupassen, um so dem Kunden kontinuierlich einen Mehrwert bieten zu können. Aufgrund der Starrheit traditionellen BPMs ist dies aber lediglich eingeschränkt möglich. Diese Herausforderung verdeutlicht den Mangel der traditionellen BPM-Ansätze an Agilität [Ba19].

Der genaue Anwendungskontext der agilen BPM-Ansätze wird in der Literatur meist vernachlässigt oder nur sehr kurz angerissen. Verschiedenen Quellen gemein ist, dass agiles BPM dort angewendet werden sollte, wo traditionelles BPM sowie die traditionellen Prozessmodelle mit ihren starr vorgegebenen Strukturen zu Einschränkungen der geschäftlichen Beweglichkeit führen. Dies ist zumeist dort der Fall, wo locker strukturierte, wissensintensive Prozesse ablaufen, bei denen es sinnvoll ist, die Reihenfolge der Aktivitäten erst zur Laufzeit zu bestimmen und von den Entscheidungen der Ausführenden oder den Ereignissen der übergeordneten Prozesse abhängig zu machen. Solche Prozesse werden als ad-hoc Prozesse oder auch als Wissensarbeit bezeichnet. Bei der Modellierung von ad-hoc Prozessen müssten traditionelle Prozessmodelle alle möglichen Szenarien aufzeigen, was zu einem sehr chaotischen Prozessmodell führen kann. Dies hätte den Verlust des wichtigsten Nutzens der Prozessmodellierung zur Folge [Ko13, Tr18]. Daher kann durch Agilität im BPM ermöglicht werden, dass zur Laufzeit der Prozesse die Reihenfolge der Aktivitäten bestimmt und so von einer vorab Modellierung aller Prozessmodelle abgesehen werden kann.

Ein weiterer Anwendungskontext des agilen BPM sind Situationen, die aufgrund der Umgebungsdynamik häufige Prozessänderungen erfordern oder in denen es wichtig ist, kontinuierlich Prozessverbesserungen herbeizuführen. Dies kann beispielsweise bei entscheidungsorientierten Prozessen der Fall sein, welche ein stark ereignisorientiertes und situationsbewusstes Verhalten oder auch nicht-standardisierte, wissensintensive Fälle aufweisen und so mehr Agilität erfordern. Das agile BPM ist auch für solche Situationen kein „[...] Allheilmittel, dass sich auf alle Szenarien anwenden lässt [...]“ [Ro15]. Es bietet aber Arbeitsweisen, die häufig ändernden Geschäfts- und Kundenanforderungen oder anderen Bedingungen der Unsicherheit gerecht werden und so „[...] schnelle Erfolge bei der Entwicklung von Fähigkeiten, Dienstleistungen oder Systemen [...]“ erzielen können [Ro15, Pa11].

Autoren, die explizit darstellen, wann ein agiles BPM und wann ein traditionelles BPM angewendet werden sollte, sind [Th13]. Die Autoren betrachten dabei die Anwendung

von agilem BPM oder traditionellem BPM im Projektkontext von BPM, was darauf zurückzuführen ist, dass [Th13] als agilen BPM-Ansatz die Verbindung von traditionellem BPM mit der agilen Methode für Softwareentwicklungsprojekte SCRUM betrachten (siehe Tabelle 2). Die nach [Th13] entwickelten Parameter für die Anwendung des jeweiligen Ansatzes sind in Abbildung 1 dargestellt und bieten eine Orientierung für die Anwendung von agilem oder traditionellem BPM.

Zusammenfassend zeigt sich, dass die Anwendung von agilen BPM-Ansätzen meist in Geschäftsumgebungen vorgezogen wird, die einer starken Dynamik unterliegen und somit häufige Änderungen und Anpassungen an den Prozessen erfordern. Agiles BPM sollte also angewendet werden, wenn traditionelles BPM im betrachteten Kontext an seine Grenzen kommt und viele wissensintensive Arbeitsabläufe stattfinden oder ad-hoc Prozesse durchgeführt werden. Zudem muss die Unternehmensorganisation und -kultur den Einsatz von agilem BPM zulassen. Im Zuge von BPM muss sich nicht für das gesamte BPM eines Unternehmens für oder gegen agiles BPM entschieden werden. Für jeden Anwendungskontext kann zwischen agilem und traditionellem BPM entschieden werden.

Abb. 1: Anwendungskontext agilen BPMs nach [Th13]

Die Literatur zeigt jedoch auch, dass der Anwendungskontext von agilen BPM-Ansätzen noch wenig Berücksichtigung findet und sich stattdessen vorrangig auf die Entwicklung von agilen BPM-Ansätzen konzentriert wird.

5 Offene Forschungsfragen laut Literatur

Dieser Abschnitt soll kurz mögliche zukünftige Forschungsschwerpunkte des agilen BPM betrachten und setzt sich somit mit der Forschungsfrage FF3 auseinander.

Das Thema agiles BPM ist in der Forschung noch ein verhältnismäßig junges Thema. Erste Veröffentlichungen gab es zwar bereits in den frühen 2000er Jahren (beispielsweise im Jahr 2007 von [Sc07]). Der Großteil der Veröffentlichungen zu dem Thema wurde jedoch innerhalb der letzten Dekade vollzogen. Der Verlauf der Veröffentlichungen ist in Abbildung

2 anhand eines Zeitstrahls dargestellt. Aus der Abbildung wird ebenfalls ersichtlich, dass besonders der SBPM-Ansatz für agiles BPM bereits von Beginn an regelmäßiges Interesse in der Forschung hervorgerufen hat.

Abb. 2: Zeitstrahl der Veröffentlichungen

Bei dem noch verhältnismäßig jungen Thema des agilen BPM ist jedoch nicht nur von Interesse, was bereits erforscht worden ist, sondern besonders diskutierte zukünftige Forschungsschwerpunkte und -gebiete. Mögliche zukünftige Forschungsschwerpunkte sind in der im Zuge dieser Arbeit behandelten Literatur wenig berücksichtigt worden. Viele Autoren behandeln das Thema zukünftiger Forschungsschwerpunkte nicht oder sehen in diesen die Weiterentwicklung und Evaluierung der in der jeweiligen Literatur entwickelten Ansätze sowie die Ausprägung von Methoden zu den von den jeweiligen Autoren vorgeschlagenen Ansätzen.

Bei der Betrachtung zukünftiger Forschungsschwerpunkte durch die Literatur muss die Zeitachse der Veröffentlichungen berücksichtigt werden. So sehen [Th13] als Forschungsschwerpunkt die Zusammenhänge von BPM und Agilität. Hierbei ist zu beachten, dass [Th13] ihre Veröffentlichung im Jahr 2013 relativ zu Beginn der Forschung zu agilem BPM getätigt haben.

[Ra19] und [Ba19] sind wesentlich aktuellere Veröffentlichungen, die zu dem Thema möglicher Forschungsschwerpunkte weiter ins Detail gehen. [Ra19] haben sich mit dem Ansatz des SBPM auseinandergesetzt und sehen als einen Forschungsschwerpunkt die Verbindung zwischen Wissensmanagement und BPM. Aufgrund des betrachteten SBPM-Ansatzes ist für [Ra19] ebenfalls von Interesse, welche Arten von social Software aktuell in Unternehmen zur Unterstützung von Wissensmanagement eingesetzt werden und welche Teile des Wissensmanagements durch social Software unterstützt werden können. In diesem Zusammenhang ist auch die Unterstützung von social Software in den einzelnen BPM-Lebenszyklusphasen als möglicher Forschungsschwerpunkt zu betrachten [Ra19].

Da [Ba19] als Schwerpunkt ihrer Veröffentlichung einen anderen Ansatz für agiles BPM als [Ra19] betrachtet haben, kann man für ihre Forschungsschwerpunkte auch eine andere Ausrichtung erkennen. Nach [Ba19] sind folgende Forschungsgebiete zu betrachten:

- Unterstützung neuer Technologien beim BPM
- Bewertung der Technologien im Hinblick auf die Anwendung von Agilität
- Soziale Auswirkungen von agilem BPM
- Lernen, Verlernen, Neulernen beteiligter Stakeholder
- Anreicherung traditionellen BPM hin zu agilem BPM

Neben den beschriebenen Forschungsschwerpunkten nach [Ba19] stellen die Autoren ebenfalls in Zusammenhang mit den sechs Kernelementen des BPM (Strategische Ausrichtung, Governance, IT, Methoden, Kultur, Menschen) mögliche Forschungsgebiete detailliert dar. Auf die Kernelemente des BPMs bezogene zukünftige Forschungsschwerpunkte sind im Folgenden beispielhaft anhand der Elemente Strategische Ausrichtung und Menschen aufgeführt:

Strategische Ausrichtung:

- Bewertung der Auswirkungen von Agilität auf die Verknüpfung von Organisationsstruktur und Prozessfähigkeiten
- Bewertung der Auswirkungen von Agilität auf die schnelle Einbettung von BPM in die Organisationsstruktur
- Bewertung der Auswirkungen der Agilität auf Prozessänderungspläne sowie die Förderung von intelligentem BPM und die organisatorische Agilität

Menschen:

- Bewertung der Auswirkungen von Agilität auf die Verbesserung der Prozesszusammenarbeit und -kommunikation
- Bewertung der Auswirkungen von Agilität auf die aktive Beteiligung von Interessensgruppen an BPM-Initiativen sowie dem Ermöglichen präziser Verantwortlichkeiten durch die Bereitstellung angemessener Autonomie

Zusammenfassend wird aufgezeigt, dass es sich bei dem Forschungsthema des agilen BPM um ein noch relativ junges Forschungsgebiet handelt, welches besonders in den letzten zehn Jahren an Bedeutung gewonnen hat. An der geringen Historie ist ebenfalls festzumachen, dass sich viele Autoren auf die Entwicklung und Evaluierung der jeweiligen Ansätze konzentrieren und sich weniger mit möglichen Forschungsschwerpunkten auseinandersetzen. [Ra19] und [Ba19] zeigen mit ihren aktuellen Veröffentlichungen jedoch auf, welche Gebiete bei bestehenden Ansätzen von besonderem Interesse für die künftige Forschung sind. Dabei spielt für die Ableitung möglicher Forschungsschwerpunkte der betrachtete Ansatz eine zentrale Rolle.

6 Zusammenfassung und Ausblick

Diese Arbeit hat sich mit der Bedeutung sowie der Umsetzung von agilem BPM auseinandergesetzt und in dem Zuge die Forschungsfragen FF1 bis FF3 beantwortet.

Über die letzten Jahre hinweg sind viele Veröffentlichungen auf dem Gebiet des agilen BPMs vorgenommen worden. Dabei wurde eine Vielzahl an Ansätzen und Herangehensweisen vorgestellt. Die verschiedenen Ansätze weisen trotz ihrer teilweise sehr unterschiedlichen Herangehensweisen auch Gemeinsamkeiten auf. So setzten sich die Ansätze des Sense-and-Respond, des BPM(N)$^{Easy1.2}$, des BPPAM sowie der Ansatz zur Verbindung von BPM und SCRUM mit der Einführung und Integration von Zyklen in den BPM-Kontext auseinander. Ebenfalls einigen verschiedenen Ansätzen gemein sind die zu integrierenden Aspekte der Kollaboration und verbesserten Kommunikation. Die Ansätze des BPM(N)$^{Easy1.2}$, des BPPAM sowie des SBPM betrachten verschiedene Aspekte zur Förderung der Zusammenarbeit/Kollaboration und der Kommunikation der beteiligten Stakeholder.

Neben den Ansätzen, die in unterschiedlichen Ausprägungen Gemeinsamkeiten aufweisen, gibt es auch Ansätze für das agile BPM, die sich von anderen Ansätzen vollkommen unterscheiden. So weist der Ansatz [Ba19] für die Kombination von BPM mit modernen Technologien nur bedingt Gemeinsamkeiten mit anderen Ansätzen auf. [Ba19] betrachten als Einsatzpunkt moderner Technologien hauptsächlich die Unterstützung der Prozessanalyse. Moderne Technologien können im BPM-Kontext jedoch vielfältig eingesetzt werden und so neben der Analyse und Auswertung von Prozessen auch andere Teilbereiche des BPM unterstützen. Denkbar wäre der Einsatz moderner Technologien zur Unterstützung der Kollaboration und Kommunikation beteiligter Stakeholder, wodurch sich Überschneidungen mit dem SBPM Ansatz ergeben. Weitere Einsatzmöglichkeiten moderner Technologien im BPM-Kontext können sich bei detaillierterer Betrachtung ergeben.

Die entwickelten Ansätze dienen der Bewältigung von Herausforderungen, die sich derzeit auf dem Gebiet des BPM ergeben. Mit Hilfe der Ansätze für agiles BPM können die Probleme der traditionellen BPM-Methoden und Vorgehensweisen zum Teil umgangen oder abgemildert werden. Die meisten Autoren sehen die Anwendung der agilen BPM-Ansätze im Umfeld von locker strukturierten wissensintensiven Prozessen (ad-hoc Prozesse) sowie in Umgebungen, die aufgrund ihrer Dynamik häufige Prozessänderungen erfordern. Dieser Anwendungskontext stellt zudem den Bereich dar, in welchem die traditionellen BPM-Methoden an ihre Grenzen stoßen.

Den betrachteten Ansätzen für agiles BPM liegen verschiedene Prinzipien zugrunde, die im Abschnitt 3 abgeleitet und beschrieben werden. Auffällig ist zudem, dass über den gesamten betrachteten Veröffentlichungszeitraum der Ansätze zu agilem BPM das SBPM bzw. die sozialen Komponenten im Zusammenhang mit BPM als einzige kontinuierlich betrachtet worden sind und auch in aktuellen Veröffentlichungen weiter betrachtet werden. Die sozialen Komponenten spiegeln somit kritische Komponenten des agilen BPMs wider und sind bei dessen Betrachtung von besonderer Bedeutung.

Eine Erweiterung der Wissensbasis durch die Einbeziehung und Analyse der von Badakhshan et al. in [BA19] identifizierten Arbeiten zum Thema agiles BPM kann potentiell zum einen die Validität der gewonnen Erkenntnisse erhöhen und zum anderen neue Aspekte aufzeigen.

Literaturverzeichnis

[Ba19] Badakhshan, P.; Conboy, K.; Grisold, T.; vom Brocke, J.: Agile business process management. A systematic literature review and an integrated framework, *Business Process Management Journal,* 2019.

[Br11] Bruno, G.; Dengler, F.; Jennings, B.; Khalaf, K.; Nurcan, S.; Prilla, M.; Sarini, M.; Schmidt, R.; Silva, R.: Key challenges for enabling agile bpm with social software, *Journal of Software Maintenance and Evolution*, 23(4):297–326, 2019.

[Im19] Imgrund, F.; Janiesch, C.: Understanding the need for new perspectives on bpm in the digital age. An empirical analysis, *International Conference on Business Process Management*, pages 288–300. Springer, 2019.

[Ki04] Kitchenham, B.: Procedures for performing systematic reviews, *Keele, UK, Keele University*, 33(2004):1–26, 2004.

[Ko13] Kolar, J.; Dockal, L.; Pitner, T.: A dynamic approach to process design. A pattern for extending the flexibility of process models, *The Practice of Enterprise Modeling*, pages 176–190, Berlin, Heidelberg. Springer, 2013.

[Le06] Levy, Y.; Ellis, T. J.: A systems approach to conduct an effective literature review in support of information systems research, *Informing Science*, 9, 2006.

[Ma17] Martins, P. V.; Zacarias, M.: An agile business process improvement methodology, *Procedia Computer Science*, 121:129 – 136, 2017.

[Me14a] Mevius, M.; Ortner, E.; Wiedmann, P.: Modeling using ordinary language as basis for agile business process management, *Lecture Notes in Informatics (LNI), Proceedings - Series of the Gesellschaft fur Informatik (GI)*, volume P225, pages 433–448, 2014.

[Me14b] Meyer, N.;Schiffner, S.: Democratizing business process management. empowering process participants to contribute to the enactment of business processes, *2014 IEEE 16th Conference on Business Informatics*, volume 2, pages 93–100. IEEE, 2014.

[Me11] Meziani, R.; Saleh, I.: Towards a collaborative business process management methodology, *International Conference on Multimedia Computing and Systems - Proceedings*, 2011.

[Pa16] Paschek, D.; Rennung, F.; Trusculescu, A.; Draghici, A.: Corporate development with agile business process modeling as a key success factor, *Procedia Computer Science*, 100:1168 – 1175, 2016.

[Pa11] Paschke, A.: A semantic rule and event driven approach for agile decision-centric business process management, *European Conference on a Service- Based Internet*, pages 254–267, Springer, 2011.

[Ra19] Ramadhani, F.; ER, M.: A conceptual model for the use of social software in business process management and knowledge management, *Procedia Computer Science*, 161:1131 – 1138, 2019.

[Ra15] Rangiha, M. E.; Comuzzi, M.; Karakostas, B.: Role and task recommendation and social tagging to enable social business process management, *Enterprise, Business-Process and Information Systems Modeling*, pages 68–82, Springer, Cham, 2015.

[Ro18] Rodríguez, D.; Molina, E. S.: The experience of implementation with agile business process management, *Advances in Science, Technology and Engineering Systems*, 3(4):284–294, 2018.

[Sc07] Schatten, A.; Schiefer, J.: Agile business process management with sense and respond, *IEEE International Conference on e-Business Engineering (ICEBE'07)*, pages 319–322, 2007.

[Th13] Thiemich, C.; Puhlmann, F.: An agile bpm project methodology, *Business Process Management*, pages 291–306, Springer, Berlin, Heidelberg, 2013.

[Tr18] Triaa, W.; Gzara, L.; Verjus, H.: A new approach for sbpm based on competencies management, *ICEIS 2018 - Proceedings of the 20th International Conference on Enterprise Information Systems*, volume 2, pages 673–681, 2018.

[Ro15] von Rosing, M.; von Scheel, J.; Gill, A. Q.: Applying agile principles to bpm, *The Complete Business Process Handbook*, pages 557 – 581, Morgan Kaufmann, Boston, 2015.

[Za17] Zacarias, M.; Martins, P. V.; Gonçalves, A.: An agile business process and practice meta-model, *Procedia Computer Science*, 121:170 – 177, 2017.

Prozessdigitalisierung: Just do it!

Prozessmanagement im Zeitalter der Digitalisierung – Vorgehen und Workshopkonzept – eingeladener Beitrag zum Workshop ZuGPM

Arno Müller[1], Hinrich Schröder[2], Lars von Thienen

Abstract: In der Workshop-Session zur Prozessdigitalisierung wird der Weg vom Kundenproblem zu konkreten Anforderungen an die technische Implementierung am praktischen Beispiel erlebbar gemacht. Es wird ein exemplarischer Serviceprozess aus Kunden- und Unternehmenssicht analysiert und optimiert. Ziel ist es, nicht nur die IT-Implementierung agil durchzuführen, sondern alle Phasen eines BPM-Projektes mit den Prinzipien agiler Ansätze zu durchlaufen. Im Beitrag wird zunächst an einem Beispiel das Vorgehen bei der Prozessdigitalisierung vorgestellt, das im Arbeitskreis „Digineering Re-Engineering im digitalen Zeitalter“ entwickelt wurde. An dem von den Autoren durchgeführten praxisorientieren Forschungsprojekt nahmen ca. 50 IT- und Prozessexperten von 13 Unternehmen teil. Zum Abschluss des Beitrags wird das Vorgehen skizziert, das auf dem ZuGPM-Workshop als ein Beitrag der "alternativen Form" interaktiv mit den Teilnehmenden durchgespielt wird.

Keywords: Agiles Prozessmanagement; Praktische Anwendung; Digineering; Workshop Prozessdigitalisierung; BPM im Zeitalter der Digitalisierung

1 Warum muss das BPM-Methodenset erweitert werden?

Business Process Management (BPM) strebt die Verbesserung der funktionsübergreifenden Arbeitsweise von Unternehmen an und stellt sicher, dass unternehmensweite Fähigkeiten zur Verfügung stehen, die ein effektives Management des gesamten Geschäftsprozesslebenszyklus ermöglichen [SS20]. Dies impliziert die Verbindung der unternehmensinternen Prozesse mit den Erwartungen des Kunden.

Im Konzept des BPM ist die Unternehmensstrategie die Vorgabe für die Prozessgestaltung, und die Soll-Prozesse liefern die Anforderungen an die IT. Diese wird als „Enabler“ der definierten Soll-Prozesse angesehen. Zentrale Ziele des BPM sind die Standardisierung von Prozessen und der eingesetzten IT, um die Prozesskosten zu reduzieren. In der betrieblichen Praxis wurde IT oft nicht als Wettbewerbsfaktor [Ca04] angesehen, sondern als notwendige

[1] NORDAKADEMIE – Hochschule der Wirtschaft, Prozessmanagement und Logistik, arno.mueller@nordakademie.de https://www.nordakademie.de
[2] NORDAKADEMIE – Hochschule der Wirtschaft, IT-Management, Betriebswirtschaftliche Anwendungen, hinrich.schroeder@nordakademie.de

doi:10.18420/inf2020_61

Ressource, um Prozesse umzusetzen. Vor diesem Hintergrund kommt der Kostenreduktion für diese Ressource durch Standardisierung und Skaleneffekte mit Outsourcing eine besondere Bedeutung zu.

In der heutigen Situation, die durch neue Potentiale zur Prozessautomatisierung und neue Möglichkeiten der digitalen Prozessgestaltung gekennzeichnet ist, reicht der kostenfokussierte, reaktive Ansatz für BPM nicht mehr aus. IT darf nicht nur als Enabler definierter Soll Prozesse angesehen werden, sondern muss als eigenständige Option zur Verbesserung des Kundenerlebnisses, der Definition neuer Geschäftsmodelle und der Prozessgestaltung eingesetzt werden. Im Mittelpunkt steht der Beitrag der IT Steigerung der Wettbewerbsfähigkeit [Po15] und des Unternehmenswertes [Kes13].

Die klassische BPM-Methode [Sc02; SS20] umfasst nicht die komplette Projektkette von der IST-Analyse eines Kundenproblems bis zum Roll-Out der IT-Anwendung. In der Regel wird in einem kundenorientierten Projekt der Kundenbedarf definiert, und falls es nötig ist beginnt auf dieser Basis das BPM-Team mit der Gestaltung der Soll-Prozesse. Wenn für die neuen Prozesse neue IT benötigt wird, werden die Anforderungen im Fachkonzept definiert und anschließend in der Software-Entwicklung umgesetzt. Dies führt zu erheblichem Know-how Verlust im Projekt und langer Laufzeit.

Digineering als integriertes Methodenset zur Prozessdigitalisierung setzt auf ein Big Picture und die iterative Definition von Prozessen und IT in dezentralen Teams, um diese Komplexität zu bewältigen.

2 Vom Big Picture zum digitalisierten Prozess

Unternehmen, die sich mit der Prozessdigitalisierung befassen, fragen oft als erstes nach der Technologie, die zum Einsatz kommen soll. Der Weg zum digitalisierten Geschäftsprozess darf aber nicht mit dem Blick auf die Technologie beginnen. Steve Jobs hat bereits 1997 klargestellt: “You have to start with the customer experience and work backwards to the technology.” [Jo97]

Digineering beginnt deshalb mit der Analyse der Kunden und deren Anforderungen. Während der agile Ansatz in der Unternehmenspraxis fast ausschließlich im Software-Engineering und oft nur in der IT-Abteilung zur Anwendung kommt [ABC17], setzt das Digineering darauf, die gesamte Projektkette unter Einsatz eines agil iterativen Vorgehens zu gestalten.

Das „Big Picture“ für einen E2E-Prozess wird in kurzer Zeit erarbeitet. Es liefert eine Prozess-Vision und somit die Orientierung für die Optimierung einzelner Teilprozesse bei der iterativen Umsetzung. Im Big Picture wird zunächst das Kundenproblem analysiert und daraus werden Ideen zur Digitalisierung abgeleitet. Im Anschluss wird untersucht, wie die internen Prozesse digitalisiert werden können. Im Ergebnis entstehen Epics / User Stories für die iterative Umsetzung. Die Verknüpfung der Kundensicht mit der internen Prozesssicht erfolgt über das Customer Journey Mapping – siehe Abb. 1.

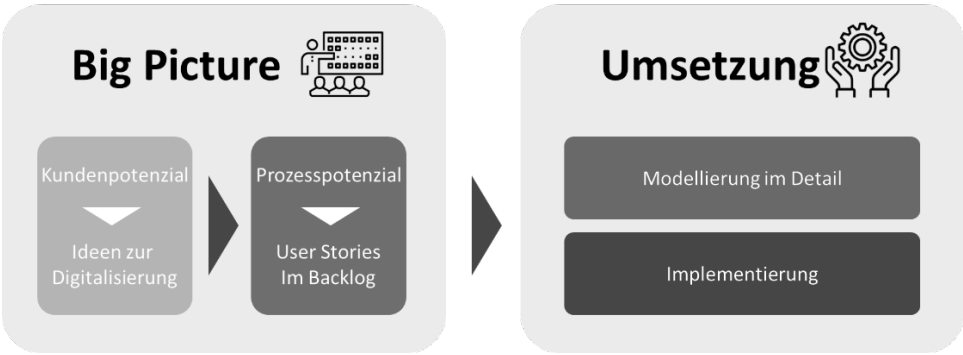


Abb. 1: Ablauf des Digineering

2.1 Analyse des Kundenpotenzials

Das Ziel der Prozessdigitalisierung ist es, für den Prozesskunden eine optimale Leistung zu erbringen. Es geht um die Begeisterung des Kunden und schnelle, schlanke Abläufe unter Einsatz der IT. Eine intensive Betrachtung des Kunden und seiner Erwartungen ist somit der Einstieg.

Hierzu werden im Folgenden drei aufeinander aufbauende Methoden vorgestellt, um die zentralen Aufgaben im Rahmen der Analyse des Kundenpotenzials durchzuführen.

1. Den Kunden verstehen – Persona zur Beschreibung des Prozesskunden nutzen
2. Komplexität der Kundensituation reduzieren – Kundenbedarfsfälle abgrenzen
3. Kundenproblem und -wunsch ableiten – Customer Journey untersuchen und gestalten

Im Customer-Journey-Mapping wird deutlich, an welchen Stellen im Ablauf die Verärgerung des Kunden entsteht und es können sofort Ideen zur Optimierung entwickelt werden.

Abschließend werden die gesammelten Ideen zur Digitalisierung als Ambition beschrieben und grob bewertet. Als Ambition wird hier eine konkretisierte Idee bezeichnet, für die die erforderlichen Maßnahmen grob definiert sind und eine Abschätzung von Kosten und Nutzen möglich ist.

2.2 Analyse des Prozesspotenzials

Mit den ersten drei Methoden *Persona*, K*undenbedarfsfall* und *Customer-Journey-Mapping* werden Ideen zur Digitalisierung aus Kundensicht identifiziert. Die nächsten drei Schritte

der „Analyse des Prozesspotenzials" beziehen sich auf die Digitalisierung interner Prozesse unter Beachtung der Touchpoints zum Kunden.

Wenn die Brücke von der Kundensicht zu den internen Prozessen durch das Customer-Journey-Mapping erfolgt ist, können diese im Detail analysiert und schließlich automatisiert werden. In diesem Baustein des Vorgehens werden Methoden des BPM mit Methoden im Kontext der Digitalisierung und des Software-Engineering verbunden:

1. Prozessteckbrief und -modellierung als BPM-Methoden
2. Bewertung des Digitalisierungspotenzials als Methode zur Digitalisierung
3. Formulierung von Epics / User Stories als Methode des Software-Engineering

Somit wird es möglich, ein durchgängiges Konzept von der Kundensicht bis zur konkreten Anforderung an die Digitalisierung abzuleiten.

Die Umsetzung der erarbeiteten Epics erfolgt bevorzugt in dezentralen Teams je Fachabteilung oder verwendeter Technologie. Ziel muss es sein, die Modellierung der Soll-Prozesse im Detail und die technische Umsetzung iterativ und dezentral zu realisieren, statt ein zu großes und komplexes Projekt zu starten. Die dezentrale Umsetzung birgt das Risiko, dass der E2E-Prozess aus dem Blick verloren wird. Die Methode stellt somit hohe Anforderungen an das Team, das Big Picture mit der konkreten Umsetzung zu verknüpfen. Die Lösung des Zielkonflikts liegt in einer Governance, die eine lose Kopplung im E2E-Prozess bei gleichzeitig großen Freiräumen in der technischen Umsetzung für dezentrale Teams gewährleistet. Die Templates für diese Art der Governance liefern die Ansätze von DevOps und agilen Frameworks [Pf19].

3 Vorstellung des Workshop-Beitrags

Im geplanten Workshop-Beitrag der sog. "alternativen Form" wird das oben beschriebene im Rahmen des Digineering genutzte Methodenset an einem konkreten Prozessbeispiel praktisch eingesetzt und zur Diskussion gestellt. Es handelt sich um den Service Prozess eines Unternehmens, das Heizungsanlagen herstellt. Das Projektteam hat die Aufgabe, diesen Prozess zu digitalisieren. Im Fokus steht der Kundenbedarf, eine defekte Heiztherme wieder in Betrieb zu nehmen. Die Fallstudie [Mü19] ist im Internet abrufbar [https://medium.com/digineering/kleine-fallstudie-zur-prozessdigitalisierung-die-defekte-heizung-136223fcb527].

Der Workshop wird als interaktive Online Veranstaltung abgehalten. Für die Fallstudie werden Ergebnisse der Persona-Analyse, die Definition des Bedarfsfalls und die Customer-Journey vorbereitet und präsentiert. Mit einem Umfragetool werden Feedback eingeholt und Anregungen gesammelt, und die Teilnehmer diskutieren die Vor- und Nachteile des

Einsatzes der Methoden. Digitalisierungsideen zu dem Fall werden an einem digitalen Whiteboard gesammelt und priorisiert. Eine Idee wird als Ambition beschrieben. Dieser letzte Teil soll zeigen, dass das Vorgehen die Kreativität fördert und viele gute Vorschläge zu Tage bringen kann.

Im zweiten Workshopteil werden die Methoden der Prozesspotenzialanalyse genutzt: Der Prozessteckbrief und das Prozessmodell werden präsentiert und mit den Teilnehmern diskutiert. Für das Feedback wird erneut das Umfragetool eingesetzt. Danach werden die Prozessschritte hinsichtlich der Digitalisierungspotenziale bewertet. Für die Prozessschritte mit dem höchsten Potenzial werden abschließend Epics / User Stories formuliert. Zum Abschluss erfolgt die Auswahl geeigneter Technologien zur Umsetzung. Hier wird das digitale Whiteboard mit vorbereiteten Templates eingesetzt.

So wird im Workshop-Beitrag nach dem Motto „Just do it“ das komplette Vorgehen von der Analyse des Kundenproblems bis zur Auswahl der Technologie zusammen mit den Teilnehmenden anhand eines Beispiels nachvollzogen und erlebbar gemacht. Für die Teilnehmenden entsteht so einerseits ein Einblick in das Methodenset des Digineering zur Prozessdigitalisierung. Andererseits können durch eine praktische Anwendung auch Erkenntnisse in Bezug auf die damit verbundenen Potenziale und Herausforderungen gesammelt werden, die nach dem Workshop ausgewertet werden können.

Literaturverzeichnis

[ABC17] Agile Business Consortium (Hrsg.): Agile Project Management Handbook V2, Ashford, 2017.

[Ca04] Carr, N. G.: Does It Matter? Information Technology and the Corrosion of Competitive Advantage, Boston, 2004.

[CC16] Capgemini Consulting: Robotic process automation The next revolution of Corporate Functions, 2016.

[Da15] Davenport, T. C.: Process Automation and the Rebirth of Reengineering, in Wall Street Journal 08.07.2015.

[Da16] Davenport, T. C.; Kirby J.: Only Humans Need Apply: Winners and Losers in the Age of Smart Machine; New York, 2016.

[Da20] Dettmers, S. u.a.: Agile Unternehmen Zukunftstrend oder Mythos der digitalen Arbeitswelt?, Studie der StepStone GmbH und der Kienbaum Institut für Leadership & Transformation GmbH, 2020.

[Ke13] Kembel, J.: CX Journey Mapping Workshop (2013) https://de.slideshare.net/jkembel/cx-journey-mapping-workshop-intro-activity-20130624-26372628?next_slideshow=1

[Kes13] Kesten, R.; Müller A.; Schröder H.: IT-Controlling - IT-Strategie, Multiprojektmanagement, Projektcontrolling und Performancekontrolle 2. Auflage München 2013,

[Jo97] Jobs, S.: Vortrag auf Apple Developers Conference (1997) https://www.imore.com/steve-jobs-you-have-start-customer-experience-and-work-backwards-technology

[Ku18] Kusay-Merkle, U.: Agiles Projektmanagement im Berufsalltag; Berlin, 2018.

[Mü19] Müller, A.: Fallstudie zur Prozessdigitalisierung https://medium.com/digineering/kleine-fallstudie-zur-prozessdigitalisierung-die-defekte-heizung-136223fcb527

[Ni19] Nielsen L.: Personas User Focused Design, London, 2019.

[Pf19] Pfister A.; Müller P.: Psychologische Grundlagen agilen Arbeitens, in: Negri, C. Hrsg. Führen in der Arbeitswelt 4.0, Berlin, 2019.

[Po15] Porter M., Heppelmann J.E.. How Smart, Connected Products Are Transforming Companies, Harvard Business Review, October, 2015.

[Sc02] Scheer, A.-W.: ARIS – Vom Geschäftsprozess zum Anwendungssystem 4. Aufl., Berlin, 2002.

[SS20] Schmelzer, H.J.; Sesselmann, W.: Geschäftsprozessmanagement in der Praxis: Kunden zufrieden stellen Produktivität steigern Wert erhöhen; 9. vollständig überarbeitete Auflage; München, 2020.

[Pr06] Pruitt, J.; Adlin T.: The Persona Lifecycle: Keeping People in Mind Throughout Product Design; San Francisco, 2006.

[St18] Stansbury P.: Agile Digital Service Handbook, Agile Business Consortium, Ashford, 2018.

[We14] Westermann, G.; Bonnet, D. McAffee A.: Leading Digital Turning Technology into Business Transformation, Boston, 2014.

Konzeptionelle Herausforderungen für die KI

Konzeptionelle Herausforderungen für die KI

Workshop auf der INFORMATIK2020: Back to the Future – 50. Jahrestagung der Gesellschaft für Informatik

Reinhard Kahle[1], Klaus Mainzer[2]

Die neue, auf stastischer Analyse großer Datenmengen aufbauende KI hat eine ganze Reihe von zukunftsträchtigen Erfolgen vorzuwiesen (z.B. zu Bilderkennung; Medizin; selbstfahrende Autos; etc.). Allerdings ist die KI auch kein „Allheilmittel“. Man wird von ihr nicht erwarten, z.B. das Haltproblem zu lösen, und wären effiziente Faktorisierungsalgorithmen in ihrer Reichweite, müßte wohl auch das Internetbanking gestoppt werden.

Auf diesem Workshop werden computationelle Fragestellungen diskutiert, für die durch die moderne KI (noch?) keine Lösungsmöglichkeiten zu sehen sind. Im besonderen soll dabei auch der Frage nachgegangen werden, wie u.U. formal gezeigt werden könnte, daß diese Fragestellungen mit KI-Methoden nicht lösbar sind.

Eingeladene Sprecher sind Prof. Dr. Wolfgang Bibel und Prof. Dr. Kristian Kersting, deren Vorträge um zwei Beiträge der Organisatoren ergänzt werden.

Im Sinne des Mottos der 50. Jahrestagung der Gesellschaft für Informatik *Back to the Future* wird mit Prof. Bibel ein Pionieer der Künstlichen Intelligenz in Deutschland in seinem Beitrag *Laßt hundert Blumen blühen* den Bogen von den Anfängen der KI zu aktuellen Fragestellungen spannen.

Prof. Kersting stellt die Herausforderungen einer *Systemischen KI* vor:

> Wir Menschen treffen Schlussfolgerungen, die weit über das aktueller Verfahren des Maschinellen Lernens hinauszugehen scheinen. Während Menschen reichhaltigere Darstellungen lernen und sie für ein breiteres Spektrum an Lernaufgaben verwenden, wurden Algorithmen des Maschinellen Lernens bisher hauptsächlich für Inselbegabungen eingesetzt, indem sie aus einer Tabelle mit Lernbeispielen eine einzelne Funktion konstruierten. In diesem Vortrag werde ich auf die Systeme KI eingehen. Ähnlich wie in der Systembiologie

[1] Carl Friedrich von Weizsäcker-Zentrum, Universität Tübingen, Keplerstr. 2, 72074 Tübingen, Deutschland, und CMA, FCT, Universidade Nova de Lisboa, 2829-516 Caparica, Portugal, kahle@mat.uc.pt
[2] TU München und Carl Friedrich von Weizsäcker-Zentrum, Universität Tübingen, Deutschland, k.mainzer@outlook.com

doi:10.18420/inf2020_62

> nimmt sie eine systemische Sicht auf die KI ein, die die Interaktion einzelner KI-Bausteine erfasst, versteht und nutzt, um ein einziges, komplexes KI-System zu beschreiben. Die System-KI kann dabei helfen, menschliche Lernaspekte zu erfassen, indem verschiedene KI- und ML-Modelle mit Hilfe von High-Level-Programmierung kombiniert werden. Da Schlussfolgern „intractable“ bleibt, nutzen aktuelle Ansätze tiefes Lernen für die Inferenz. Anstatt „nur den neuronalen Weg zu gehen“, werde ich argumentieren, auch probabilistische Schaltkreise zu verwenden, eine tiefe, aber „tractable“ Architektur für Wahrscheinlichkeitsverteilungen. Dieser hybride Ansatz kann die Inferenz beschleunigen, wie ich an einigen Beispielen verdeutlichen werde.

Der Beitrag von Prof. Mainzer behandelt in einem weiten Rahmen die Aufgabe, Beweisassistenten zur Zertifizierung kausaler Zusammenhänge in der KI einzubringen. Klaus Mainzer ist Mitglied der Steuerungsgruppe im Auftrag der Bundesregierung für eine KI-Normungsroadmap und schlägt daher beim Thema Verifikation und Zertifizierung den Bogen von der Theorie zur Praxis.

Prof. Kahle verweist auf zahlentheoretischen Fragestellungen, von denen die moderne Kryptographie abhängig ist und plädiert in diesem Zusammenhang dafür, daß die KI Methoden entwickeln sollte, um ihre eigenen Grenzen formal erfassen zu können.

Auf dem Workshop wird ausreichend Zeit zur Diskussion gegeben, und die Veranstalter erhoffen sich einen fruchtbaren Austausch unter reger Beteiligung der Teilnehmer, so daß sich neue Herausforderungen und Perspektiven für die Künstliche Intelligenz in den kommenden Jahren aufzeigen lassen.

Literaturverzeichnis

[Ma19] Mainzer, K. (2019), Künstliche Intelligenz. Wann übernehmen die Maschinen? Springer: Berlin 2. erweiterte Auflage (engl. Übersetzung).

Wie sicher ist KI?

Grundlagen und Verifikation von Künstlicher Intelligenz

Klaus Mainzer[1]

Abstract: In Zeitalter der Digitalisierung nimmt die Künstliche Intelligenz eine Schlüsselstellung ein. Was ist aber Künstliche Intelligenz? Was kann sie heute und was kann sie nicht? Im Unterschied zu den logischen Formalismen der klassischen symbolischen KI wird das aktuelle Machine Learning durch statistisches Lernen dominiert, das die Bewältigung großer Datenmengen mit leistungsstarken Rechnern in Technik und Wirtschaft verspricht. Wie sicher sind aber statistische Korrelationen? Was sind und was können demgegenüber Kausalmodelle leisten? Kausalanalysen sind erkenntnistheoretisch und ethisch mit Verantwortungsfragen eng verbunden. Nur wenn diese Grundlagen der KI-Technologie klar verstanden sind, lassen sich auch ihre Anwendungen beurteilen und ethisch-rechtlich bewerten. Daher plädiert dieser Beitrag für eine Kombination von Machine Learning mit kausalem Lernen und zertifizierten KI-Programmen durch Beweisassistenten.

1 Formale Verifikationsprogramme der symbolischen KI

In ihren Anfängen wurde Künstliche Intelligenz als Automatisierung logischen Denkens und Beweisens verstanden. Man spricht in diesem Fall auch von symbolischer KI [KM20].

1.1 Von Logik und Beweistheorie zur SAT-Verifikation

David Hilbert hatte in den 1920er Jahren einen Formalismus entwickelt, um mathematische Konzepte in der formalen Sprache der Prädikatenlogik zu formulieren. Damit stand ein formales Verfahren zur Verfügung, um die Allgemeingültigkeit (bzw. Widersprüchlichkeit) einer Formel mechanisch nachzuweisen. Eine Formel kann nur dann widersprüchlich sein, wenn es ein endliches Anwendungsbeispiel gibt, das den Widerspruch aufzeigt. Außerdem gibt es zu jeder Formel ein „allgemeinstes“ (i.A. unendlich großes) Anwendungsbeispiel.

Dieses nach dem Logiker Jacques Herbrand benannte Beispiel besteht aus aussagenlogischen Formeln, die man systematisch und sukzessiv aus den in der Formel vorkommenden Symbolen konstruieren kann. Diese Formeln sind generische Namen für Individuen, die in einem beliebigen Anwendungsbeispiel nach Maßgabe der prädikatenlogischen Formel

[1] TU München und Carl Friedrich von Weizsäcker-Zentrum, Universität Tübingen, Deutschland, k.mainzer@outlook.com

doi:10.18420/inf2020_63

notwendigerweise vorkommen müssen. Jeder endlich große Anfang des Beispiels stellt nun eine Konjunktion von aussagenlogischen Formeln dar, deren Erfüllbarkeit (oder Widersprüchlichkeit) man mit rein aussagenlogischen Mitteln in endlicher Zeit entscheiden kann (z.B. mit Wahrheitstabellen).

Mit der Verfügbarkeit von Computern in den 1950er Jahren begannen mehrere Forscher in den USA das genannte Beweisverfahren konkret auszuformen und auf verschiedene Arten zu implementieren. Das Kernproblem war dabei, zunächst ein effizientes Verfahren für die aussagenlogische Erfüllbarkeit (SAT-Solving) zu finden, die wiederholt für endliche Anfänge des kombinatorisch stark anwachsenden Herbrand Beispiels nachzuweisen war. Der Philosoph Willard V. O. Quine hatte dazu Wahrheitstabellen vorgeschlagen, die aber exponentielles Wachstum aufweisen [Qu55]. Paul Gilmore hatte stattdessen eine disjunktive Form (der Art $A \wedge B \wedge C \cdots \vee E \wedge F \wedge G \cdots \vee \ldots$) benutzt und auf einer IBM 704 implementiert, aber auch dieser Ansatz war schon auf kleinen Beispielen ineffizient [Gi60]. IBM produzierte von 1954 bis 1960 das Modell 704 mit Schaltkreisen aus Röhren, etwa 18KB bis 144KB Hauptspeicher aus Magnetkernen sowie Magnetbandspeicher von je 5MB. Martin Davis und Hilary Putnam fassten 1960 die Gesamtsituation zusammen und präsentierten zudem ein völlig neues Verfahren für die Erfüllbarkeitsprüfung, das im Kern auf der Idee der Resolution aufbaut [DP60]. Sie konnten damit von Hand ein Beispiel von Gilmore rechnen, bei dem dessen Programm abbrechen musste.

Danach implementierten Davis, Logemann und Loveland das neue Verfahren ebenfalls auf einer IBM 704 und entdeckten sofort, dass es an einer entscheidenden Stelle verändert werden musste, um nicht ebenfalls ein explosives Wachstum der Formeln auszulösen. In ihrer 1962 publizierten Arbeit wird der heute als „DPLL“ bekannte rekursive Algorithmus beschrieben, der erstmalig ein praktikables Verfahren zum SAT-Solving darstellt und bis in die 1990er Jahre hinein Bestand hatte [Da+62]. Das Beispiel von Gilmore wurde damit in nur 2 Minuten automatisch gelöst.

Im Jahr 1965 publizierte J. A. Robinson eine neue Methode zum Beweisen in Prädikatenlogik. Dieses „Resolution“ genannte Verfahren erlaubt es, direkt aus der Menge der prädikatenlogischen Klauseln neue logisch implizierte Klauseln abzuleiten, ohne wie zuvor einen Umweg über die Aussagenlogik zu nehmen. Falls die ursprüngliche Formel (bzw. deren Klauselmenge) widersprüchlich ist, kann man immer in endlicher Zeit die „leere“ Klausel ableiten, die nicht erfüllt werden kann und so den Widerspruch zu Tage treten lässt. Die aussagenlogische Variante wurde bereits erwähnt. Prädikatenlogisch ist die Sache komplizierter, denn die beiden Klauseln $\{P(x)\}$ und $\{\neg P(f(y))\}$ widersprechen sich auf den ersten Blick noch nicht. Allerdings sind x und y implizit als allquantifiziert zu verstehen, sodass aus $\forall x.P(x)$ auch $\forall x.P(f(x))$ folgt und damit ein Widerspruch zu $\neg P(f(y))$.

Zum Vergleich: in der „alten“ Methode aus den 1950er Jahren hätte man den Widerspruch durch sukzessives und im Prinzip zielloses Aufbauen des Herbrand-Beispiels wie folgt ermitteln müssen: Soll es ein nicht-leeres Beispiel geben, so muss mindestens ein Individuum „a“ vorhanden sein und es muss eine Funktion „f“ geben. Damit muss es auch ein „$f(a)$“

geben, und weiter „$f(f(a))$“, „$f(f(f(a)))$“, usf. Wir prüfen nun, ob durch Einsetzen von a (in x und y) alleine ein Widerspruch zu Tage tritt; das ist nicht der Fall. Auch durch Einsetzen von a für x und $f(a)$ für y ergibt sich kein Widerspruch. Erst durch Einsetzen von $f(a)$ für x und a für y tritt der Widerspruch zu Tage. Durch geschicktes Einsetzen kommt man also (viel) schneller zum Ziel als durch ungeschicktes Einsetzen, und nach jedem Einsetzen braucht man eine erneute Erfüllbarkeitsprüfung. Demgegenüber vergleicht Robinson die prädikatenlogischen Formeln systematisch und stellt die Gleichheit durch gezieltes Einsetzen („Unifikation“) her. Im Beispiel sehen wir, dass wir im alten Verfahren für x ein Individuum „$f(a)$“ einsetzen müssen, wenn wir für y „a“ einsetzen, oder eben für x ein „$f(f(a))$“ und für y dann ein „$f(a)$“ etc.

Prädikatenlogische Resolution ist deutlich effizienter als das vorhergehende Verfahren; die Gilmore Formel wird nach nur 5 einfachen Resolutionsschritten (Ableitung neuer Klauseln) als widersprüchlich erkannt, was auch von Hand problemlos zu bewältigen ist. Im Gegenzug waren die ersten 24 Einsetzungen von Davis & Putnam jeweils erfüllbar, und erst die 25. Einsetzung erzeugte einen Widerspruch.

Aufgrund der deutlich besseren Effizienz wendete sich die damalige KI sofort der Resolutionsmethode als neuem Werkzeug für das zentrale Problem des „Reasoning“ (Herstellen von Schlussfolgerungen) zu. Ein unabhängig agierender intelligenter Akteur oder Agent muss seine Umwelt wahrnehmen können, daraus Schlussfolgerungen für sein intendiertes Handeln ableiten und diese sodann in Aktionen übersetzen. Hieraus ergaben sich Teilgebiete der KI wie Bilderkennung („Computer Vision“) und Sprachverstehen einerseits und Robotik andererseits sowie als zentrale Komponente Wissensrepräsentation mit intelligentem Schlussfolgern und Planen. Nach der herkömmlichen wissenschaftlichen Methode modelliert man die Umwelt mathematisch um durch Berechnungen Vorhersagen treffen können. Da die Mathematik nach Hilbert prädikatenlogisch gefasst werden kann, liegt es nahe, eine prädikatenlogische Repräsentation und einen automatischen Beweiser als zentrales Werkzeug vorzusehen.

Ein wichtiger Schritt war das Lehrbuch „Problem solving methods in AI“ von Nils Nilsson [Ni71, Ni80]. Neben verschiedenen Algorithmen um Gewinnstellungen in Lösungsräumen z.B. von Brettspielen zu suchen, widmen sich 3 von 8 Kapiteln dem Automatischen Beweisen mit der Resolutionsmethode und ihren Anwendungsmöglichkeiten. Hier geht es nun nicht mehr nur um den effizienten Beweis eines logischen Widerspruchs, sondern vor allem auch um die Extraktion von konkreten Handlungsanweisungen und Antworten aus dem Widerspruchsbeweis. Ein Beispiel ist die Extraktion eines Plans zum Erfüllen eines Ziels bis hin zum automatischen Schreiben eines Programms.

Eine eingeschränkte Form von Klauseln (sog. Horn-Klauseln) wurden zur Grundlage der neuen Programmiersprache PROLOG. Grundidee ist hier, sowohl Daten als auch Berechnungsregeln als Menge von Klauseln zu schreiben und einen universellen Beweiser für die konkreten Berechnungen zu nutzen. Dies sollte transparente, leicht wartbare Programme ermöglichen ohne die Notwendigkeit komplexer Schleifenkonstrukte und

Verzweigungen. Einer der ersten Prolog Compiler wurde um 1978 am „Department for Artificial Intelligence“ der Universität Edinburgh fertiggestellt [CM81].

In Deutschland begann eine kleine Gruppe um Jörg Siekmann mit der Programmierung eines Resolutionsbeweisers. Ein Ziel sollte es sein, die Beweise in einem Lehrbuch der Berechnungstheorie automatisch nachzuvollziehen. Diese Gruppe organisierte 1983 die führende Tagung „International Joint Conference of AI (IJCAI-83)“ in Karlsruhe. Themen waren System Support, Theorem Proving, Cognitive Modelling, Automatic Programming, Planning and Search, Knowledge Representation, Learning and Knowledge Acquisition, Logic Programming, Natural Language, Expert Systems, Vision, Robotics.

Aber in Deutschland war die KI nicht unumstritten. Einen eigenen Fachbereich gab es dafür anders als in Edinburgh nicht. Erst 1988 wurde das Deutsche Forschungszentrum KI (DFKI) gegründet. Ein führendes Lehrbuch der Informatik sprach noch von der „sogenannten künstlichen Intelligenz“. Allerdings erwiesen sich viele kühne Ziele der KI auch als schwieriger als zunächst vermutet. Nicht nur erwies sich Intelligenz als ein unerwartet komplexes Phänomen. Bereits mathematische Logik und ihre Beweise sind nicht einfach zu bewältigen, besonders wenn es effizient und in realisierbarer Zeit geschehen soll. Bereits 1962 hatten Davis, Logemann und Loveland resümiert: „We hoped that some mathematically meaningful and, perhaps nontrivial, theorems could be solved. The actual achievements in this direction were somewhat disappointing“.

Bis in die 1990er Jahre trat das SAT-Solving der Aussagenlogik in den Hintergrund, da es für Beweise in der Prädikatenlogik nicht mehr gebraucht wurde. Dann aber wurden mikroelektronische Schaltungen immer komplexer, deren Schaltpläne im Wesentlichen in Boole'scher Algebra (Schaltalgebra) vorliegen. Ein weit beachtetes Ereignis war 1994 ein Fehler in der Gleitkommadivision des Pentium 5 Prozessors von Intel. Im Zuge des neu erwachten Interesses an der Verifikation kombinatorischer Schaltkreise gelang Joao Markes-Silva [Ma95] und Karem Sakallah 1996 ein richtungsweisender Durchbruch beim SAT-Solving: die intelligente Kombination der DPLL Suchprozedur mit dem Deduktionsverfahren Resolution [MS96, MS99].

Deduktionsverfahren häufen ständig neues Wissen an, indem sie dynamisch neue Klauseln erzeugen und leiden in der Folge unter diesem Speicherverbrauch. Reines DPLL-Suchen hat keinen dynamischen Speicherverbrauch, kann aber unter langen Laufzeiten leiden, wenn es durch backtracking wiederholt auf dieselbe Weise bei der Lösung derselben Teilmenge von Formeln scheitert. Das conflict driven clause learning (CDCL) besteht nun darin, dass man aus den Konflikten lernt, auf die man bei der DPLL Suche stößt. Ein Konflikt besteht darin, dass eine der Klauseln unter der momentan untersuchten Variablenbelegung falsch (false) wird. Dieser Fall tritt niemals direkt durch eine *true/false* Entscheidung auf, sondern immer durch weitere Wertepropagationen aufgrund einer Entscheidung.

Solche Propagationen werden immer durch Klauseln erzwungen: z.B. erzwingt eine Klausel $\{x, y\}$ die Propagation von $y = true$ sobald $x = false$ gesetzt wird. Falls eine Klausel false

wird, so liegt das daran, dass sie die umgekehrte Propagation erzwingen würde, im Beispiel etwa $\{x, \neg y\}$. In dieser Situation gibt es aber immer eine Resolvente, im Beispiel wäre das $\{x\}$. Als Resolvente ist die Klausel eine logische Konsequenz der zu lösenden Formel und kann als zusätzliche Klausel „gelernt" werden. In unserem Beispiel würde sie verhindern, dass nach backtracking jemals wieder eine Lösung mit x=false versucht wird, denn $\{x\}$ erzwingt zuvor sofort $x = \mathit{true}$.

Eine in der Praxis wichtige Konsequenz des CDCL Solving besteht darin, dass der Solver im Fall einer unlösbaren Formel nun die leere Klausel „lernt" und dazu einen Resolutionsbeweis konstruiert. Nun kann das bislang lapidare Ergebnis „UNSAT" durch einen nachvollziehbaren (und unabhängig nachprüfbaren) Beweis begründet werden. In Fällen wo „UNSAT" auf einen Fehler in der Anwendung (z.B. einem Schaltkreis) zeigt, kann aus dem Beweis oft eine Anleitung zur konkreten Lokalisation des Fehlers und zu seiner Korrektur entnommen werden.

Neben dem Lernen wurden für moderne CDCL Solver weitere wichtige Komponenten entwickelt, wie z.B. hoch effiziente Implementierungen der Wertepropagation und gute Heuristiken zur Auswahl der nächsten zu belegenden Variable. Außerdem entstanden um das CDCL-Solving weitere verwandte Algorithmen wie die Berechnung von Primimplikanten (wichtig für die Minimierung von Formeln) oder zur Berechnung maximaler lösbarer Teilmengen von unlösbaren Klauselmengen (wichtig für Optimierungszwecke).

Einige wesentliche industrielle Anwendungsgebiete für das SAT Solving sind heute die formale Verifikation von Mikroelektronik und von Software sowie die Lösung von Konfigurationsproblemen. Im Falle von Software übersetzt man ein Programm vollautomatisch in eine Boole´sche Formel. Im Falle von Konfigurationsproblemen (z.B. plugin Konfiguration im Eclipse Programm-Editor) codiert man die Randbedingungen („wenn A dann nur wenn auch B aber nicht C oder D") als Klauseln und lässt den Solver eine gültige Konfiguration berechnen. Auch den Konfiguratoren der Automobilindustrie, die man im Internet benutzen kann, liegen ähnliche Randbedingungen zugrunde. Viele Automobilhersteller benutzen heute SAT-Solver um ihre hochkomplexen Konfigurationsprobleme zu lösen bzw. die Formelsysteme zu analysieren und zu validieren [KS00]. In der Mikrobiologie bestehen Anwendungsmöglichkeiten dadurch, dass sowohl Proteine als auch Gene durch die Kombination von endlich vielen Grundbausteinen gebildet werden, was man durch Boole'sche Variablen codieren kann.

Moderne CDCL Solver funktionieren äußerst effizient und robust, ohne händische Hilfen, und bewältigen Anwendungsprobleme mit (Hundert-)Tausenden von Variablen und Klauseln. Da das Erfüllbarkeitsproblem ein NP-vollständiges Problem ist (es ist das prototypische NP-vollständige Problem) ist kein Algorithmus bekannt, der es auch im verzwicktesten Fall in weniger als exponentieller Zeit (in der Anzahl der Variablen) löst; außerdem ist die Existenz eines solchen Algorithmus sehr unwahrscheinlich. Andererseits können somit auch alle anderen NP-vollständigen Probleme im Kern durch SAT-Solver gelöst werden [Kn16, BG18].

Insgesamt gibt es viele wichtige Anwendungsprobleme, bei denen die Struktur der Klauseln ein praktisch effizientes CDCL Solving erlaubt. Selbst wenn also durch die KI Forschung auf dem Gebiet des Reasoning bis jetzt nicht alle Erwartungen und Hoffnungen erfüllt werden konnten, so hat schon alleine das daraus entstandene SAT-Solving zu vielen konkreten Anwendungen in Industrie und Wissenschaft geführt.

1.2 Von Logik und Beweistheorie zu maschinellen Beweisassistenten

Die moderne Beweistheorie eröffnet die Möglichkeit, interaktive und automatische Beweisassistenten zu entwickeln, die sowohl Beweise in der Mathematik als auch Softwareprogramme in der Informatik überprüfen können. Die zunehmende Komplexität von menschlichem Wissen und menschlicher Technik macht Verifikation zu einem Schlüsselproblem zukünftiger Entwicklungen, insbesondere auf dem Gebiet der Künstlichen Intelligenz. Bemerkenswert ist aber auch, wie tief diese aktuellen Entwicklungsfragen in den Grundlagen von Logik, Mathematik und Philosophie verwurzelt sind.

In der KI realisieren Algorithmen Wissensverarbeitung, indem aus Datenstrukturen (also Zeichenreihen) weitere Zeichenreihen abgleitet werden. Das entspricht dem Ideal des mathematischen Beweisens, das in der Mathematik seit der Antike vertreten wird. Euklid hatte gezeigt, wie aus als wahr vorausgesetzten Axiomen nur durch logische Schlüsse mathematische Lehrsätze abgeleitet und bewiesen werden konnten. In der KI stellt sich die Frage, ob mathematisches Beweisen auf Algorithmen übertragen und damit „automatisiert“ werden kann. Dahinter steht dann die grundlegende KI-Frage, ob und bis zu welchem Grad Denken automatisiert, also durch einen Computer ausgeführt werden kann.

Schauen wir uns dazu einen klassischen Beweis näher an: Um 300 v.Chr. bewies Euklid die Existenz unendlich vieler Primzahlen [AZ01, 3]. Euklid vermeidet den Begriff „unendlich“ und behauptet: „Es gibt mehr Primzahlen als jede vorgelegte Anzahl von Primzahlen“. Der Beweis wird in der Schule mit Widerspruch geführt. Man nimmt das Gegenteil der Behauptung an, schließt logisch unter dieser Annahme auf einen Widerspruch. Die Annahme war also falsch. Wenn wir nun voraussetzen, dass eine Aussage entweder war oder falsch ist, dann gilt das Gegenteil der Annahme: die Behauptung ist richtig.

Der Nachteil an diesem Beweis ist, dass wir die Existenz der Primzahlen konstruktiv nicht bewiesen haben. Wir haben nur gezeigt, dass das Gegenteil der Annahme zu Widersprüchen führt. Um die Existenz eines Objekts zu beweisen, benötigen wir einen Algorithmus, der ein Objekt erzeugt und beweist, dass die Aussage für dieses Beispiel richtig ist. Formal lautet eine Existenzaussage $A \equiv \exists x.B(x)$. In einer abgeschwächten Form könnten wir fordern, eine Liste von endlich vielen Zahlen $t_1, \ldots, t_n$ zu konstruieren, die Kandidaten für die Aussage B sind, so dass die Oder-Aussage $B(t_1) \vee \cdots \vee B(t_n)$ gilt, also die Aussage B für wenigstens eine der konstruierten Zahlen $t_1, \ldots, t_n$ wahr ist. Das könnte auch eine Maschine leisten. Wenn allgemein für alle x ein y mit $B(x, y)$ existieren soll, also formal $A \equiv \forall x \exists y.B(x, y)$ gilt, dann benötigen wir einen Algorithmus p, der für jeden x-Wert einen

Wert $y = p(x)$ konstruiert, so dass $B(x, p(x))$ für alle x gilt, also formal $\forall x B(x, p(x))$. In einer schwächeren Form wären wir zufrieden, wenn für den Suchprozess eines y-Wertes für einen gegeben x-Wert wenigstens eine obere Schranke $b(x)$ berechnet werden könnte, also formal $\forall x \exists y \leq b(x) B(x, y)$. Damit lässt sich der Suchprozess genau abschätzen.

Der amerikanische Logiker G. Kreisel hat deshalb gefordert, dass Beweise mehr als bloße Verifikationen sein sollen. Sie sind gewissermaßen „eingefrorene" Algorithmen. Man muss sie nur in den Beweisen entdecken und „herauswinden" (unwinding proofs) [Fe96]. Dann können sie auch Maschinen übernehmen.

Tatsächlich ist im erwähnten Beweis des Primzahlsatzes ein konstruktives Verfahren „versteckt". Es lässt sich nämlich für jede Position r einer Primzahl p_r (in der Aufzählung $p_1 = 2, p_2 = 3, p_3 = 5, \ldots$) eine obere Schranke $b(r)$ berechnen, also zu jeder vorgelegten Anzahl von Primzahlen eine weitere angeben, die allerdings unterhalb einer berechenbaren Schranke liegt. In Euklids indirektem Beweis werden ja endlich viele Primzahlen angenommen, die kleiner oder gleich einer Schranke x sind, also $p \leq x$. Damit wird dann eine Zahl $1 + \prod_{p \leq x} p$ konstruiert, aus der die Widersprüche abgeleitet werden. (Dabei bezeichnet $\prod_{p \leq x} p$ das Produkt aller Primzahlen, die kleiner als x sind.) Wir konstruieren daher zunächst die Schranke

$$g(x) := 1 + x! \geq 1 + \prod_{p \leq x} p.$$

Die Fakultätsfunktion $1 \cdot 2 \cdot \ldots \cdot x = x!$ lässt sich durch die sogenannte Stirling-Formel abschätzen. Wir zielen allerdings auf eine obere Schranke der $r + 1$-ten Primzahl p_{r+1}, die nur von der Position r in der Aufzählung der Primzahlen anstelle von der unbekannten Schranke $x \geq p_r$ abhängt. Euklids Beweis zeigt $p_{r+1} \leq p_1 \cdot \ldots \cdot p_r + 1$. Daraus lässt sich *für alle* $r \geq 1$ (durch vollständige Induktion über r) beweisen, dass $p_r < 2^{2^r}$. Die gesuchte berechenbare Schranke ist also $b(r) = 2^{2^r}$.

In Logik und Mathematik werden Formeln (also Zeichenreihen) Schritt für Schritt abgeleitet, bis der Beweis einer Behauptung abgeschlossen ist. Computerprogramme arbeiten im Grunde wie Beweise. Schritt für Schritt leiten sie nach festgelegten Regeln Zeichenfolgen ab, bis ein formaler Ausdruck gefunden ist, der für eine Lösung des Problems steht. Stellen wir uns z.B. die Montage eines Werkstücks auf einem Fließband vor. Das entsprechende Computerprogramm beschreibt, wie das Werkstück Schritt für Schritt aus vorausgesetzten Einzelteilen nach Regeln aufeinander aufbauend entsteht.

Ein Kunde wünscht von einem Informatiker ein Computerprogramm, das ein solches Problem löst. Bei einem sehr komplexen und unübersichtlichen Produktionsprozess, möchte er sicher vorher einen Beweis, dass das Programm auch korrekt arbeitet. Eventuelle Fehler wären gefährlich oder würden erhebliche Mehrkosten verursachen. Ein Informatiker beruft sich dazu auf eine Software, die den Beweis automatisch aus den formalen Eigenschaften des Problems extrahiert hat. So wie Software im „Data Mining" zur Suche von Daten oder Datenkorrelationen eingesetzt wird, so lässt sich passende Software auch zur automatischen

Suche von Beweisen einsetzen. Man spricht dann von „proof mining “ [Ko08, Kap. 2]. Das entspricht Georg Kreisels Ansatz, Algorithmen aus Beweisen herauszufiltern (unwinding proofs), nun allerdings automatisch durch Computerprogramme.

Dann entsteht allerdings die Frage, ob die Software zur Extraktion des Beweises selber zuverlässig ist. In einem genau vorgegebenen Rahmen lassen sich solche Zuverlässigkeitsbeweise für die zugrunde gelegte Software führen. Der Kunde kann dann sicher sein, dass das Computerprogramm für seine Problemlösung korrekt arbeitet. Dieses „automatische Beweisen“ hat also nicht nur erhebliche Bedeutung für die moderne Softwaretechnik. Sie führt auch zu philosophisch tiefen Fragen, wieweit nämlich (mathematisches) Denken automatisiert werden kann: Die Beweisfindung ist automatisch. Den Korrektheitsbeweis der dazu verwendeten Software führt aber ein Mathematiker. Selbst wenn wir diesen Beweis wieder automatisieren würden, entsteht eine grundlegende erkenntnistheoretische Frage: Führt uns das nicht in einen Regress, an dessen Ende immer der Mensch steht (stehen muss)?

Ein Beispiel ist das interaktive Beweissystem MINLOG, das aus formalen Beweisen automatisch Computerprogramme heraus extrahiert [Sc06, SW12, Kap. 7]. Es benutzt die Computersprache LISP. Ein einfaches Beispiel ist die Behauptung, dass für jede Liste v von Symbolen in LISP eine Umkehrliste w mit umgekehrter Anordnung der Symbole existiert. Das ist wieder eine Behauptung von der Form $A \equiv \forall v \exists w B(v, w)$. Der Beweis kann informal durch eine Induktion über den Aufbau der Listen v geführt werden. MINLOG extrahiert daraus automatisch ein passendes Computerprogramm. Aber auch für anspruchsvolle mathematische Beweise lässt sich diese Software benutzen. Ein allgemeiner Zuverlässigkeitsbeweis garantiert, dass die Software korrekte Programme liefert.

Um Widersprüche zu vermeiden, wie sie Anfang des 20. Jahrhunderts die Grundlagendiskussion der Mathematik auslösten, sollten Beweise konstruktiv sein. Hilbert wollte mathematische Theorien zunächst formalisieren und für diese Formalismen nachträglich konstruktive Widerspruchsfreiheits- und Vollständigkeitsbeweise führen. Die Mathematiker sollten also weiterhin ungestört ihre Arbeit tun und die Logiker würden sich danach um ihre Korrektheit und Widerspruchsfreiheit kümmern. Wie die Gödelschen Sätze zeigen, lässt sich das Hilbertsche Programm nur bedingt realisieren. Demgegenüber wollte Brouwers Intuitionismus von vornherein nur konstruktive Verfahren in der Mathematik zulassen. Mit Blick auf Sicherungsverfahren spielen intuitionistische Verfahren selbst bei maschinellen Computerprogrammen bis heute eine Rolle.

Die intuitionistische Interpretation der logischen Konstanten geht auf Brouwer, Heyting und Kolmogrov zurück [He34, Ko32, Ko08, 43]. Die Brouwer-Heyting-Kolmogorov (BHK)-Interpretation erklärt die Bedeutung logischer Konstanten mit Termen von Beweiskonstruktionen:

Es gibt keinen Beweis von $\perp$ („Falsch“).

i. Ein Beweis von $(A \wedge B)$ ist ein Paar (q, r) von Beweisen, wobei q ein Beweis von A und r ein Beweis von B ist.

ii. Ein Beweis $(A \vee B)$ von ist ein Paar (n, q), das aus einer ganzen Zahl n und einem Beweis q besteht, der A beweist, falls $n = 0$ und entsprechend B, falls $n \neq 0$.

iii. Ein Beweis p von $(A \rightarrow B)$ ist eine Konstruktion, die einen hypothetischen Beweis q von A in einen $p(q)$ von B transformiert.

iv. Ein Beweis p von $\forall x A(x)$ ist eine Konstruktion, die für jede Konstruktion c_d eines Elements d aus der Domäne des Allquantors einen Beweis $p(c_d)$ von $A(d)$ erzeugt.

v. Ein Beweis von $\exists x A(x)$ ist ein Paar (c_d, d), wobei c_d die Konstruktion eines Elements d aus der Domäne des Existenzquantors und q ein Beweis von $A(d)$ ist.

2 Typentheoretische Verifikationsprogramme in Mathematik und Informatik

Eine wichtige Gemeinsamkeit zwischen Computersprachen und logischen Formalismen ist die Unterscheidung von Typen unterschiedlicher Zeichen, Terme, Formeln und Beweise. B. Russell hatte die Typentheorie vorgeschlagen, um Widersprüche in logischen Formalismen und der Mengenlehre zu vermeiden. In einer Computersprache müssen die Typen der Symbole genau angegeben werden, damit die Maschine keine fehlerhaften Zordnungen durchführt.

2.1 Beweise als Programme mit intuitionistischen Typen

1969 beobachtete der Logiker W.A. Howard (wie vorher bereits H.B. Curry), dass Gentzens Beweissystem natürlichen Schließens in seiner intuitionistischen Version direkt als eine typisierte Variante der Berechnung in Form des Churchschen Lambda-Kalküls interpretiert kann [Ho69]. Man spricht daher auch vom Curry-Howard Isomorphismus. Nach Church bedeutet $\lambda a.b$ eine Funktion, die ein Element a auf den Funktionswert b mit $\lambda a.b[a] = b$ abbildet. Im Folgenden werden Beweise durch Terme $a, b, c, \ldots$, Aussagen durch $A, B, C, \ldots$ repräsentiert [Ma18, Kap. 9].

Beispiele:

$$\begin{array}{rc} & [A] \\ \lambda a.b & \vdots \\ (\rightarrow I) & \dfrac{B}{A \rightarrow B} \end{array}$$

$$\begin{array}{l} \lambda a.(\lambda b.a) \\ (\rightarrow I) \end{array} \begin{array}{c} [A] \\ \vdots \\ \dfrac{B \rightarrow A}{A \rightarrow (B \rightarrow A)} \end{array}$$

Ein Beweis ist ein Program, and die Formel, die er beweist, ist der Typ für dieses Programm. Nach Gentzens Kalkül des natürlichen Schließens lässt sich auch sein Sequenzenkalkül durch den Lambda-Kalkül charakterisieren.

Ein Beweis von $\Gamma \vdash \alpha$ bedeutet, dass man über ein Programm verfügt, das für gegebene Werte mit Typen aufgelistet in Γ ein Objekt vom Typ α herstellt. Ein Axiom korrespondiert zur Einführung einer neuen Variablen mit einem neuen und uneingeschränktem Typ, die $\rightarrow I$ Einführungsregel korrespondiert zur Funktionsabstraktion und die $\rightarrow E$ Eliminationsregel korrespondiert zur Funktionsanwendung.

$t : \alpha$ bedeutet sowohl „t beweist A“ als auch „t ist vom Typ α“.

Auch Aussagen lassen sich als Typen in der intuitionistischen Typentheorie verstehen. Nach dem Curry-Howard Isomorphismus von Aussagen als Typen ist $\Sigma x : A.B$ die disjunkte Summe der A-indizierten Familie von Typen B und $\Pi x : A.B$ ihr cartesisches Produkt [DP16].

Die kanonischen Elemente von $\Sigma x : A.B$ sind Paare (a, b) mit $a : A$ und $b : B[x := a]$ vom Typ, der durch Ersetzung aller frei vorkommenden x in B durch α entsteht. Die Elemente von $\Pi x : A.B$ sind (berechenbare) Funktionen f mit $fa : B[x := a]$, wobei $a : A$.

Wie lassen sich Theoreme als Typen nach dem Curry-Howard Isomorphismus verstehen? Als Beispiel betrachten wir das Theorem, wonach es beliebig große Primzahlen gibt, d.h. formal:

$$\forall m : \mathbb{N}.\exists n : \mathbb{N}.m < n \wedge Prime(n)$$

Nach dem Curry-Howard Isomorphismus wird daraus der Typ von Funktionen, die eine Zahl m auf ein Tripel $(n, (p, q))$ abbilden, wobei n eine Zahl ist, p ein Beweis, dass $m < n$ ist, und q ein Beweis, dass n eine Primzahl ist:

$$\Pi m : \mathbb{N}.\Sigma n : \mathbb{N}.m < n \times Prime(n)$$

Das ist ein Beispiel für das Beweis-als-Programm Prinzip: Danach wird ein konstruktiver Beweis, dass es beliebig große Primzahlen gibt, zu einem Programm, das für gegebene Zahlen eine größere Primzahl zusammen mit Beweisen erzeugt, dass sie in der Tat größer und Primzahl ist.

Zusätzlich zu den Typenformern des Curry-Howard Isomorphismus erweiterte der Logiker und Philosoph P. Martin-Löf die basale intuitionistische Typentheorie, die Heyting

Arithmetik der höheren Typen HA^{ω} und Gödels System T der primitiv-rekursiven Funktionen höheren Typs enthält, mit primitiven Identitätsttypen, wohl-geordneten Baumtypen, Hierarchien von Universen und allgemeinen Begriffen induktiver und induktiv–rekursiver Definitionen. Martin-Löfs beweistheoretische Erweiterung betrifft sowohl die Anwendung auf die Programmierung als auch auf die Formalisierung der Mathematik [Ma98].

Neben den gegebenen Regeln für Π gibt es analoge Regeln für andere Typenformer, die den logischen Konstanten der typisierten Prädikatenlogik entsprechen. In der intuitionistischen Typenarithmetik werden wie in der Peano-Arithmetik die natürlichen Zahlen mit 0 und der Nachfolgeroperation s erzeugt. Um Widersprüche mit unbegrenzten Begriffsbildungen wie dem Universum aller Typen zu vermeiden, führte Martin-Löf das Universum kleiner Typen ein. Das typentheoretische Universum U ist analog zu einem Grothendieck-Universum in der Mengenlehre. Das Grothendieck-Universum ist eine Menge von Mengen, die abgeschlossen unter allen Methoden ist, mit denen Mengen in der Zermelo-Fraenkel Mengenlehre gebildet werden können. Das Grothedieck-Universum deckt weite Möglichkeiten mathematischer Begriffsbildung ab.

In der intuitionistischen Typentheorie ist das Auswahlaxiom ein beweisbares Theorem. Es ist nämlich eine unmittelbare Folge der BHK-Interpretation der intuitionistischen Quantoren. In der Mengenlehre ist das Auswahlaxiom allerdings nicht immer konstruktiv. Typen sind im Allgemeinen keine konstruktiven Approximationen von Mengen im klassischen Sinn. Mit Blick auf Datenstrukturen in der Informatik ist bemerkenswert, wie sich wohlgeordnete Bäume und iterative Mengen in der intuitionistischen Typentheorie darstellen lassen. Wohlfundierte Bäume können auf ein typentheoretisches Modell der konstruktiven Mengenlehre erweitert werden [Ac78].

In der intuitionistischen Typentheorie werden induktive Typen durch Konstruktoren eingeführt:

Beispiele:

a) Typ $\mathbb{N}$ der natürlichen Zahlen mit Konstruktoren
   - $0 : \mathbb{N}$
   - $\mathrm{succ} : \mathbb{N} \to \mathbb{N}$

b) Typ $\mathrm{List}(A)$ der endlichen Listen von Elementen des Typs A mit Konstruktoren
   - $\mathrm{nil} : \mathrm{List}(A)$ (leere Liste)
   - $\mathrm{cons} : A \to \mathrm{List}(A) \to \mathrm{List}(A)$ (Hinzufügung eines Elements am Anfang der Liste)
   - $\mathrm{app} : \mathrm{List}(A) \to \mathrm{List}(A) \to \mathrm{List}(A)$ (Verbindung zweier Listen)

Ein Induktionsprinzip beweist eine Behauptung für einen Typ, der durch seine Konstruktoren frei erzeugt ist.

2.2 Von der Typentheorie zur univalenten Mathematik

In der Tradition von Leibnizens Mathesis Universalis stellt sich die Frage, bis zu welchem Grad das Denken von (mengentheoretischen) Unendlichkeiten durch formale Codes („Typen“) dargestellt werden kann, die durch „Maschinen“ bearbeitbar sind. In der Informatik lassen sich Typen des funktionalen Programmierens als erste Schritte in diese Richtung verstehen.

Seit ihren Anfängen spielen Datentypen eine Schlüsselrolle in Computersprachen: Wieweit können mathematische Objekte mit Typen von Computersprachen repräsentiert werden? Die Homotopie Theorie stammt aus der algebraischen Topologie und homologischen Algebra mit Beziehungen zur höheren Kategorientheorie, die als abstraktes Fundament der Mathematik verstanden werden kann. Demgegenüber ist die Typentheorie ein Zweig der mathematischen Logik und theoretischen Informatik. Die Homotopie Typentheorie (Homotopy Type Theory = HoTT) verbindet beide Ansätze und interpretiert Typen als Objekte der abstrakten Homotopie Theorie. Daher versucht HoTT, sowohl eine universale („univalente“) Grundlegung der Mathematik als auch einen Kalkül für einen Beweisassistenten zu entwickeln.

In HoTT entsprechen Terme der intuitionistischen Typentheorie solchen der Homotopie Theorie: In der intuitionistischen Typentheorie kann ein Term $a : A$ als ein Element a vom Typ A oder als ein Beweis der Aussage A oder, in der Homotopie Typentheorie, als ein Punkt des Raums A verstanden werden. Beweise p der Identität zwischen zwei Elementen a, b vom Typ A werden geometrisch als Pfade veranschaulicht, die die entsprechenden Punkte verbinden.

HoTT verbindet die Typentheorie mit Beweistheorie, Mengenlehre und Homotopie Theorie [UFP13]: Mit höheren induktiven Typen lassen sich Geometrie und Topologie charakterisieren. Wie die bisherigen induktiven Typen sind höhere induktive Typen ein allgemeines Schema, um neue Typen zu definieren, die durch einige Konstruktoren erzeugt werden (Einführungsregeln). Im Unterschied zu gewöhnlichen induktiven Typen erzeugen diese Konstruktoren nicht nur Punkte dieses Typs, sondern auch Pfade zwischen diesen Punkten, Pfade zwischen Pfaden zwischen Punkten und weiter höhere Pfade in diesem Typ [AW09].

Grundlage von HoTT ist das von dem russischen Mathematiker W. Voevodsky eingeführte Univalenz Axiom (UA). Dabei bezieht sich Äquivalenz auf höher dimensionale Objekte von der Mengenlehre bis zur Kategorientheorie (z.B. gleiche Elemente, isomorphe Mengen, äquivalente Groupoide). UA besagt, dass „alles“ durch Äquivalenz erhalten wird. Insbesondere sind äquivalente Typen (z.B. isomorphe Strukturen) identisch.

Die Homotopie Typentheorie (HoTT) eröffnet grundlegende Zusammenhänge zwischen Beweisbarkeit, Konstruktivität und Berechenbarkeit. HoTT soll ermöglichen, mathematische Beweise in eine Programmiersprache für maschinelle Beweisassistenten sogar für höhere mathematische Kategorien mit “Isomorphismen als Gleichheit” (UA) zu übertragen (z.B. Coq). Daher lautet ein wesentliches Ziel von HoTT:

Typenprüfung $\Rightarrow$ Beweisprüfung in höheren Kategorien (d.h. für „schwierige Beweise“)

HoTT ist durch höhere induktiv definierte Strukturen (z.B. induktiv definierte Räume mit Kollektionen von Punkten, Pfaden, höheren Pfaden etc.) erweitert, die durch passende Induktionsprinzipien charakterisiert werden können. HoTT ist klassisch widerspruchsfrei mit Bezug auf ein semantisches Modell in der Kategorie von Kan Komplexen [Vo12, Jo02], kann aber auch konstruktiv begründet werden (T. Coquand).

Die Motivation, sich mit HoTT zu beschäftigen, hat mit einer neuen Art von Grundlagenkrise zu tun, der sich die moderne Mathematik zunehmend stellen muss. In der gegenwärtigen Mathematik könnten nämlich Beweise so kompliziert werden, dass ein einzelner Mathematiker sie kaum in allen Details prüfen kann: Man vertraut auf die Expertise seiner Kolleginnen und Kollegen. Die Situation ist ähnlich zur modernen industriellen Arbeitswelt. Nach dem französischen Soziologen Emile Durkheim (1858-1917) ist die moderne industrielle Produktion so komplex, das sie nur nach dem Prinzip der Arbeitsteilung ("job sharing") und des gegenseitigen Vertrauens in Expertise organisiert werden kann. Niemand hat dabei aber den totalen Überblick über alle Details.

Auf dem Hintergrund gravierender Fehler, die von Experten übersehen wurden, wollte der Wladimir Voevodsky (1966-2017, IAS Princeton, Fields Medaille) aber nicht länger dem Prinzip des "job-sharing" in der Mathematik trauen. Menschen könnten in Zukunft mit der wachsenden Komplexität der Mathematik überfordert sein. Sind Computer die einzige und letzte Lösung? Voevodskys Grundlagenprogramm univalenter Mathematik ist durch die Idee einer Software für Beweisprüfung inspiriert, um Vertrauen & Verifikation in der Mathematik zu garantieren. Diese Herausforderung der Beweisprüfung lässt auf eine Grundlagenkrise der modernen Software im allgemeinen erweitern. Wie soll die Sicherheit von immer komplexer werdenden Computerprogrammen (z.B. in der Künstlichen Intelligenz) garantiert werden? [Ma18, Kap. 2] Die folgenden Abschnitte legen dafür die ersten Schritte mit Beweisassistenten, die auf dem Formalismus von HoTT aufbauen.

2.3 Von der Typentheorie zu Beweisassistenten

Der Kalkül der Konstruktionen (Calculus of Constructions = CoC) ist eine Typentheorie von T. Coquand et al., die sowohl als typisierte Programmiersprache als auch als konstruktive Begründung der Mathematik dienen kann [CH88]. Ähnlich wie Martin-Löf erweitert dieser Kalkül den Curry-Howard Isomorphismus auf Beweise im ganzen intuitionistischen Prädikatenkalkül. Coc benötigt nur sehr wenige Konstruktionsregeln für Terme. Die Objekte von CoC sind Beweise (Terme mit Propositionen als Typen), Propositionen (kleine Typen), Prädikate (Funktionen, die Propositionen zurückgeben), große Typen (Typen von Prädikaten, z.B. P), T (Typ von großen Typen). Die Ableitungsregeln von CoC lassen sich wieder in einem Sequenzkalkül zusammenstellen. Logische Operatoren und Datenypen in CoC benötigen sehr wenige Basisoperatoren. Der einzige logische Operator zur Bildung weiterer Operatoren und Datentypen ist $\forall$ (z.B. $\forall C : P.(A \Rightarrow C)$ für die Negation $\neg A$).

Der Kalkül der induktiven Konstruktionen (Calculus of inductive Constructions = CiC) basiert auf CoC und verfügt über zusätzliche induktive und ko-induktive Definitionen, die durch folgende Regeln zur Konstruktion von Termen eingeführt werden [BC04]. In CiC wird ein induktiver Typ frei durch eine bestimmte Anzahl von Konstruktoren erzeugt [Pa93].

Induktive Beweise ermöglichen es, Behauptungen für unendliche Kollektionen von Objekten zu beweisen (z.B. ganze Zahlen, Listen, binäre Bäume), weil alle diese Objekte in einer endlichen Anzahl von Schritten konstruiert werden. Das Induktionsprinzip eines induktiven Typs beweist eine Behauptung für einen Typ, der frei durch seine Konstruktoren erzeugt ist.

Neben induktive Types gibt es ko-induktive Typen, die unendliche Objekte betreffen (z.B. potentiell unendliche Listen, potentiell unendliche Bäume mit unendlichen Ästen) [Gi96]. Terme erhält man durch wiederholten Gebrauch von Konstruktoren, wie bereits für induktiven Typen gezeigt wurde. Allerdings gibt es kein Induktionsprinzip und die Äste können unendlich sein.

In praktischen Anwendungen wie Telekommunikation, Energie oder Transport lassen sich diese (potentiell) unendlichen Objekte auch als Ströme (z.B. Daten-, Energie- und Warenströme) illustrieren. Sie entstehen in unbegrenzt vielen Schritten, die durch den Konstruktor Cons definiert werden. Im Unterschied zum induktiven Typ einer Liste gibt es keinen Konstruktor der leeren Liste. Daher können endliche Listen nicht konstruiert werden.

Auf dieser Grundlage lässt sich nun der Coq Beweisassistent einführen [BC04]. Coq implementiert eine Programmspezifikation, die auf dem Kalkül der induktiven Konstruktionen (CiC) beruht und beides eine Logik höherer Ordnung und eine reich typisierte funktionale Sprache verbindet. Die Befehle von Coq erlauben,

- Funktionen oder Prädikate zu definieren (die effizient evaluiert werden können)
- mathematische Theoreme und Software-Spezifikationen zu behaupten
- formale Beweise von diesen Theoremen interaktiv zu entwickeln
- diese Beweise durch eine relativ kleine Zertifikation maschinell zu prüfen
- zertifizierte Programme zu extrahieren (e.g., Objective Caml, Haskell, Scheme).

Coq liefert interaktive Beweismethoden, Entscheidungs- und Semi-Entscheidungsalgorithmen. Interaktiv bedeutet, dass der Kalkül ständig durch die extrahierten Beweisstrategien und Routinen erweitert wird, die dann automatisch für neue Problemlösungen zur Verfügung stehen. Verbindungen mit externen Theorembeweisern sind verfügbar. Coq ist also eine Plattform für die Verifikation sowohl von mathematischen Beweisen als auch von Computerprogrammen auf der Grundlage von CiC.

2.4 Verifikation von Schaltkreisen mit typentheoretischen Beweisassistenten

Ein Hardware oder Software Programm ist korrekt („zertifiziert durch Coq“), falls verifiziert werden kann, dass es einer gegebenen Spezifikation in CiC folgt. Eine entsprechende Verifikation wird nun am Beispiel eines Schaltkreises in der Elektrotechnik erläutert [CJ96]. Dabei kann die Struktur bzw. Architektur eines Schaltkreises und sein tatsächliches Verhalten mathematisch durch verschaltete endliche Automaten modelliert werden. In Schaltkreisen sind potentiell unendlich lange Zeitfolgen von Daten (Ströme) zu berücksichtigen.

Ein Schaltkreis ist korrekt genau dann, wenn unter bestimmten Bedingungen die Output-Ströme des Automaten, der die Schaltkreisstruktur darstellt, äquivalent sind zum Automaten, der das Verhalten des Schaltkreises beschreibt. Daher muss zunächst die Automatentheorie in CiC mit den ko-induktiven Typen von Strömen implementiert werden.

Die Korrektheit eines Schaltkreises wird also durch die Äquivalenz seiner Struktur und seines Verhaltens bewiesen. Dabei werden Struktur und Verhalten durch zwei zusammengesetzte Automaten modelliert. Die Äquivalenz von zusammengesetzten Automaten kann durch ein Äquivalenz-Lemma invarianter Relationen bewiesen werden [CJ96]. Dieses Lemma kann ebenfalls in CiC implementiert werden. Dieses Lemma ist ein Beispiel für eine Beweisroutine von Coq, mit der Verifikationen von Schaltkreisen unter bestimmten Bedingungen automatisch realisiert werden können.

Wie aktuell Software-Verifikationen sind, zeigt die jüngste Technikentwicklung. Inkorrektheit von Programmen kann zu Katastrophen führen, die von falschen Angaben von Strahlungsdosierungen in der Medizin, die Patienten schädigten, über die Explosion der Rakete Ariane V aufgrund eines Programmierfehlers bis zu Software- und Systemfehlern von Boing 737max in 2019 reichen. Solche Unfälle werfen ein Schlaglicht auf die Gefahren sicherheitskritischer Systeme ohne Software Verifikation. Dramatisch wird die Situation in den Programmen der Künstlichen Intelligenz, die wegen ihrer Komplexität zu “schwarzen Kästen” werden [Ma19]. Im mathematischen Idealfall könnten Beweisassistenten, wie sie in diesem Artikel beschrieben wurden, ihren exakten Betrieb garantieren. Davon sind wir allerdings in der Praxis noch weit entfernt. Selbst in der Mathematik sind Beweisassistenten bisher nur eine große Vision auf eine zukünftige sichere Welt, die allerdings theoretisch möglich wäre.

3 Grundlagen und Verifikation des Machine Learning

Im Unterschied zu den logischen Formalismen symbolischer KI orientiert sich das moderne Machine Learning an statistischen Lernverfahren neuronaler Netze. Der Anwendungserfolg des Machine Learning beruht darauf, dass große Datenmengen mit leistungsstarken Rechnern bewältigt werden können. Wie sicher ist aber statistisches Lernen im Unterschied zu formalen Verifikationsverfahren der symbolischen KI?

3.1 Vom Bayesschen Lernen zum Machine Learning

Eine zentrale Herausforderung für KI-Anwendungen ist die Komplexität der betrachteten Systeme – von zellulären Organismen über Robotiksysteme bis zum Internet. Die Bayessche Wissenschaftstheorie erklärt statistisch, wie sich unsere Modelle und Hypothesen über die Welt durch neue Erfahrungen verändern und wie wir aus Erfahrung lernen können. Nach dem Bayesschen Theorem kommt es zur Bestimmung der Wahrscheinlichkeit $P(M|D)$ eines Modells M unter Voraussetzung der Datenmenge D zunächst darauf an, die Datenplausibilität (Likelihood) $P(D|M)$ eines Modells und seine apriorische Wahrscheinlichkeit $P(M)$ einzuschätzen.

Sie kann daher als Rahmentheorie für Lernalgorithmen verstanden werden, mit denen Lernverfahren im Machine Learning automatisiert werden. Diese Automatisierung ist z.B. in der Bioinformatik, Robotik oder im Internet von zentraler Bedeutung, da die komplexen Datenmengen und ihre Verbindung mit großen Mengen von Modellparametern immer schwieriger zu durchschauen sind [Ma20, Kap. 4]. Im Machine Learning spielen neuronale Netze nach dem Vorbild des menschlichen Gehirns eine dominante Rolle. Der Durchbruch der KI-Forschung in der Praxis hängt wesentlich mit der Fähigkeit neuronaler Netze zusammen, große Datenmengen (Big Data) z.B. bei der Mustererkennung mit effektiven Lernalgorithmen anzuwenden. Praktische Anwendungen erfordern Tausende von Neuronen und Synapsen in mehrschichtigen neuronalen Netzen (deep learning), die statistisch mit endlich vielen Datensätzen von Inputs und Outputs trainiert werden.

Vom Bayesschen Standpunkt aus lassen sich neuronale Netze als graphische Modelle $M(w)$ mit bestimmten Parametern w auffassen. Diese Modelle erinnern vereinfacht an das menschliche Gehirn. In klassischen vorwärtslaufenden (feed-forward) neuronalen Netzen sind Neuronen als Knoten eines Graphen in Schichten angeordnet. Jedes Neuron einer Schicht ist mit allen Neuronen der nachfolgenden Schicht durch gerichtete Kanten (Synapsen) verbunden, die Neuronen einer Schicht untereinander aber nicht. Ausnahmen sind die Input-Schicht, deren Neuronen keine einlaufenden Verbindungen besitzen, und die Output-Schicht, deren Neuronen keine auslaufenden Verbindungen haben. Die Schichten zwischen Input- und Output-Schicht heißen „versteckt“ (hidden). Jede Verbindung/Kante ist im graphischen Modell eines neuronalen Netzes mit einer Zahl gewichtet, die der Intensität der synaptischen Verbindung entspricht. Diese Gewichte (englisch: weights) sind die Parameter w des graphischen Modells $M(w)$ eines neuronalen Netzes. Jedes Neuron ist durch eine Aktivierungsfunktion charakterisiert, mit der die Input-Output Relation für dieses Neuron definiert wird. In Analogie zum Gehirn sagt man, dass ein Neuron „feuert“ bzw. erregt ist, wenn die Summe der gewichteten Inputs seiner Nachbarzellen einen bestimmten Schwellenwert überschreitet.

3.2 Was sind und wozu brauchen wir Kausalmodelle?

Die zentrale Herausforderung sind die gewaltigen Datenmengen D mit Tausenden von „versteckten“ (englisch: hidden) Parametern H ihrer statistischen Modelle M. Die sich daraus ergebenden globalen Wahrscheinlichkeitsverteilungen $P(D, M, H)$ sind mathematisch häufig nicht berechenbar. Was sich im Detail tatsächlich unter diesen statistischen Datenwolken abspielt, bleibt verborgen.

Statistische Datenkorrelationen können Hinweise auf kausale Zusammenhänge liefern, müssen es aber nicht. Stellen wir uns eine Testreihe vor, bei der sich eine günstige Korrelation zwischen einer verabreichten chemischen Substanz und der Bekämpfung bestimmter Krebstumore ergibt. Ein aktuelles Beispiel ist auch die Entwicklung eines Impfstoffes gegen den Corona-Virus. Auch in diesem Fall entsteht Druck des betroffenen Unternehmens, mit einem entsprechenden Medikament in die Produktion zu gehen und Gewinne abzuschöpfen. Aber auch betroffene Patienten mögen darin ihre letzte Chance sehen. Tatsächlich erhalten wir ein nachhaltiges Medikament aber nur, wenn wir den zugrunde liegen kausalen Mechanismus des Tumorwachstums, also die Gesetze der Zellbiologie und Biochemie verstanden haben.

Statistisches Lernen und Schließen aus Daten reichen also nicht aus. Wir müssen vielmehr die kausalen Zusammenhänge von Ursachen und Wirkungen hinter den Messdaten erkennen. Diese kausalen Zusammenhänge hängen von den Gesetzen der jeweiligen Anwendungsdomäne unserer Forschungsmethoden ab, also den Gesetzen der Physik oder den Gesetzen der Biochemie und des Zellwachstums im Beispiel der Krebsforschung. Wäre es anders, könnten wir mit den Methoden des statistischen Lernens und Schließen bereits die Probleme dieser Welt lösen. Tatsächlich scheinen das einige kurzsichtige Zeitgenossen beim derzeitigen Hype der Künstlichen Intelligenz zu glauben.

Statistisches Lernen und Schließen ohne kausales Domänenwissen ist aber blind – bei noch so großer Datenmenge (Big Data) und Rechenpower. Galilei und Newton erkannten grundlegende Gesetze aus nur wenigen Beobachtungsdaten. Neben der Statistik der Daten bedarf es zusätzlicher Gesetzes- und Strukturannahmen der Anwendungsdomänen, die durch Experimente und Interventionen überprüft werden. Kausale Erklärungsmodelle (z.B. das Planetenmodell Newtons oder ein Tumormodell) erfüllen die Gesetzes- und Strukturannahmen einer Theorie (z.B. die Gravitationstheorie oder die Gesetze der Zellbiologie).

Beim kausalen Schließen werden Eigenschaften von Daten und Beobachtungen aus angenommenen Kausalmodellen, d.h. Gesetzesannahmen von Ursachen und Wirkungen, abgeleitet. Kausales Schließen ermöglicht damit, die Wirkungen von Interventionen oder Datenveränderungen (z.B. durch Experimente) zu bestimmen. Kausales Lernen versucht umgekehrt, ein Kausalmodell aus Beobachtungen, Messdaten und Interventionen (z.B. Experimente) abzuleiten, die zusätzliche Gesetzes- und Strukturannahmen voraussetzen.

Ein Kausalmodell besteht aus einem System von Zuordnungen von Ursachen zu Wirkungen mit eventuellen Störvariablen. Ursachen und Wirkungen werden durch Zufallsvariablen beschrieben. Ihre funktionalen Zuordnungen (unter Berücksichtigung von Störvariablen) werden durch Gleichungen definiert, also z.B. Wirkung $X_j = f(X_i, R)$ in funktioneller Abhängigkeit von Ursache X_i und Stör- und Rauschvariable R. Anschaulich kann das Netzwerk der Ursachen und Wirkungen durch einen Graphen von Knoten und Kanten dargestellt werden. Zufallsvariablen von Ursachen und Wirkungen entsprechen Knoten. Kausale Wirkungen entsprechen gerichteten Pfeilen: $X_i \rightarrow X_j$ bedeutet, dass Ursache X_i Wirkung X_j auslöst.

Es lässt sich beweisen, dass ein Kausalmodell eine eindeutige Wahrscheinlichkeitsverteilung der Daten einschließt, aber nicht umgekehrt: Für Kausalmodelle (z.B. Planetenmodell) müssen zusätzliche Gesetze (z.B. Gravitationsgesetz) angenommen werden [MJS13]. Um kausale Abhängigkeiten und Unabhängigkeiten von Ereignissen zu erkennen, muss die Abhängigkeit und Unabhängigkeit der sie darstellenden Zufallsvariablen ermittelt werden. Statistisch lässt sich die Unabhängigkeit der Resultate x und y zweier Zufallsvariablen (anschaulich Zufallsexperimente) X und Y dadurch ausdrücken, dass ihre Verbundwahrscheinlichkeit $p(x, y)$ faktorisierbar ist, d.h. $p(x, y) = p(x)p(y)$. Man spricht in diesem Fall auch von der Markov-Bedingung.

3.3 Vom statistischen zum kausalen Lernen

Ein hochaktuelles Anwendungsbeispiel sind selbstlernende Fahrzeuge: Um das Prinzip zu erläutern, können wir uns vereinfacht ein elektrisches Spielzeugauto vorstellen, das rund herum mit Sensoren ausgestattet ist. Die Sensoren (z.B. Nachbarschaft, Licht, Kollision) seien mit den Neuronen eines neuronalen Netzwerks verbunden. Werden benachbarte Sensoren bei einer Kollision mit einem äußeren Gegenstand erregt, dann auch die mit den Sensoren verbundenen Neuronen. So entsteht im neuronalen Netz ein Verschaltungsmuster, das den äußeren Gegenstand repräsentiert. Im Prinzip ist dieser Vorgang ähnlich wie bei der Wahrnehmung eines äußeren Gegenstands durch einen Organismus – nur dort sehr viel komplexer. Wenn wir uns nun noch vorstellen, dass dieses Automobil mit einem „Gedächtnis“ (Datenbank) ausgestattet wird, mit dem es sich solche gefährlichen Kollisionen merken kann, um sie in Zukunft zu vermeiden, dann ahnt man, wie die Automobilindustrie in Zukunft unterwegs sein wird, selbst-lernende Fahrzeuge zu bauen.

Hier zeigt sich aber eine grundlegende Schwäche des derzeitigen maschinellen Lernens: Wie viele reale Unfälle sind erforderlich, um selbstlernende (“autonome”) Fahrzeuge zu trainieren? Wer ist verantwortlich, wenn autonome Fahrzeuge in Unfälle verwickelt sind? Welche ethischen und rechtlichen Herausforderungen stellen sich? Bei komplexen Systemen wie neuronalen Netzen mit Tausenden oder sogar Millionen von Elementen erlauben zwar die Gesetze der statistischen Physik, globale Aussagen über Trend- und Konvergenzverhalten des gesamten Systems zu machen. Die Zahl der Parameter ist jedoch unter Umständen so groß, dass keine lokalen Ursachen ausgemacht werden können. Die neuronalen Netze sind

also eine Black Box, die mit Big Data trainiert wird, um gewünschtes Verhalten zu erzeugen. Keiner weiß im Einzelnen, was dort in der Black Box abgeht. Wenn aber Ursachen und Wirkungen nicht klar zu unterscheiden sind, lassen sich rechtliche und ethische Fragen der Verantwortung nicht klären. *Ehe wir also über Ethik und Recht sprechen, müssen wir unsere Hausaufgaben in der Grundlagenforschung des maschinellen Lernens machen.*

Tatsächlich ist das maschinelle Lernen häufig nur Statistik mit Lernalgorithmen und neuronalen Netzen - mathematisch keineswegs spektakulär wie in den Medien suggeriert. Jeder Anfänger der Statistik weiß, dass statistische Korrelationen keine kausalen Erklärungen ersetzen können: Wenn eine günstige statistische Korrelation zwischen einer chemischen Substanz und dem Abnehmen eines Krebstumors gefunden wurde, ist das noch keine Garantie für ein nachhaltiges Medikament. Dazu muss man das Grundlagenwissen über die kausalen Wachstumsgesetze eines Tumors und biochemische Grundgesetze kennen. Dasselbe gilt für einen Impfstoff gegen den Corona Virus SARS-CoV-2.

Mit diesem Beispiel verbinde ich eine grundsätzliche Feststellung für den heutigen KI-Hype: Einige glauben damit ja bereits auf Wasser gehen und alle Probleme dieser Welt in absehbarer Zeit mit „KI“ lösen zu können. Erfolgreich sind diese KI-Methoden aber nur dann, wenn sie mit Fachwissen und Theorie aus den jeweiligen Anwendungsgebieten (wie z.B. Physik, Medizin und Ingenieurwissenschaften) verbunden werden.

Statistisches Lernen und Schließens ist jedenfalls nur schwache KI, die jeder einfache Organismus in der Natur auch ohne statistische Formeln bewältigt: Selbst ein Wurm wird nach gehäuften Erfahrungen von gefährlichen Situationen davor zurückschrecken. Was über statistische Mustererkennung in Daten hinausgeht, ist die Fähigkeit zu kausalem Lernen und Schließen. Gibt es Algorithmen, mit denen sich kausale Modelle unter geeigneten Bedingungen finden lassen? Dieses kausale Lernen wäre ein erster Schritt in Richtung einer starken KI. Tatsächlich ist kausales Lernen mittlerweile Thema theoretischer Grundlagenforschung. Aber es bedarf noch vieler Forschung, bis einmal eine Software z.B. in einem biochemischen Datensatz automatisch ein kausales Erklärungsmodell entdecken und damit eine begründete medizinische Diagnose geben kann. Wie sicher ist jedoch ein Softwareprogramm, wenn es zunehmend mehr oder weniger intelligente Entscheidungen selbstständig treffen soll?

3.4 Sicherheit und Vertrauen von Machine Learning durch logische Beweise?

Ein Computerprogramm heißt korrekt bzw. zertifiziert, falls verifiziert werden kann, dass es einer gegebenen Spezifikation folgt. Praktisch angewendet werden Verifikationsverfahren mit unterschiedlichen Graden der Genauigkeit und damit der Verlässlichkeit [TB03]. Aus Zeit-, Aufwands- und Kostengründen begnügen sich viele Anwender allerdings nur mit Stichprobentests. Im Idealfall müsste ein Computerprogramm aber so sicher sein wie ein mathematischer Beweis. Dazu wurden Beweisprogramme („Beweisassistenten“) entwickelt, mit denen ein Computerprogramm automatisch oder interaktiv mit einem Nutzer auf Korrektheit überprüft wird.

Künstliche neuronale Netze sind zwar äußerst effektiv, um komplexe Probleme (real world problems) zu bearbeiten [Ma19]. Was aber fehlt, sind Spezifikationen und Standards für die Sicherheit ihrer Outputs. Dazu muss die Black Box neuronaler Netze besser verstanden, kontrolliert und verifiziert werden. Die Verifikation neuronaler Netze ist allerdings ein hartes Erkenntnisproblem: Selbst der Nachweis einfacher Eigenschaften erweist sich im Rahmen der Komplexitätstheorie als NP-vollständig. Gründe dafür sind die Größe der praktisch angewendeten Netze (Skalierung) und die nichtlinearen Aktivierungsfunktionen ihrer Neuronen, die von Menschen in diesem Umfang und mit dieser Geschwindigkeit nicht nachvollzogen werden können. Da neuronale Netze zudem der Dynamik komplexer Systeme unterliegen, sind sie häufig empfindlich gegen kleine Störungen und Veränderungen ihrer Inputs, die sich zu unkontrollierbaren Effekten aufschaukeln können. Robustheit und Stabilität der Netze hängt also mit ihrer Sicherheit eng zusammen.

Für unterschiedliche Klassen neuronaler Netze lassen sich unterschiedliche Verifikationen angeben, die aus verschiedenen Theorien der Logik und Mathematik ableitbar sind. Dazu wurden Verifikationsverfahren vorgeschlagen, die auf der Erfüllbarkeit von Formeln der Booleschen Aussagenlogik (SAT = Satisfiability Theories), Erfüllbarkeit von Formeln der Prädikatenlogik 1. Stufe (SMT = Satisfiability Modulo Theores), Reduktion auf lineare Probleme (MIP = Mixed Integer Linear Programming) und Robustheit von mehrschichtigen Perzeptron-Netzen (MLP = Multi-Layer Perceptron) beruhen. Bei SAT- und SMT-Verifikationen wird die klassische KI (symbolic AI) des automatischen Beweisens (automated reasoning) mit dem Machine Learning verbunden. MIP beruht auf der Logik und Algebra linearen Programmierens. Robustheitsuntersuchungen von MLP wenden Erkenntnisse aus der Theorie komplexer dynamischer Systeme im Machine Learning an.

SAT-Verifikationen stehen in der Tradition der klassischen KI. Sie sind mittlerweile nicht nur von theoretischem Interesse, sondern wurden für industrielle Anwendungen ausgebaut [FH07]. Neuronale Netze lassen sich durch verschiedene Rand- und Nebenbedingungen (constraints) charakterisieren. Logisch liegt es daher nahe, die Verifikation eines neuronalen Netzes auf die Verifikation der Formeln zurückzuführen, mit denen sich ihre Rand- und Nebenbedingungen formalisieren lassen. Bei der SAT-Verifikation werden solche neuronalen Netze untersucht, deren Rand- und Nebenbedingungen sich durch Formeln der Booleschen Aussagenlogik charakterisieren lassen. Verifikation besteht dann in dem Nachweis der logischen Erfüllbarkeit dieses Formelsystems. Damit sollen seine Widerspruchsfreiheit und der sichere Ablauf im entsprechenden neuronalen Netz garantiert werden.

Hier zeigt sich sehr klar, wie aktuelle Fragen der Sicherheit moderner Software und KI in Grundlagenfragen der Logik und Philosophie verwurzelt sind. Derzeit beschäftige ich mich mit der Frage, wie das moderne maschinelle Lernen durch solche Beweisassistenten kontrolliert werden kann [MSS18]. Am Ende geht es um die Herausforderung, ob und wie man KI-Programme zertifizieren kann, bevor man sie auf die Menschheit loslässt. Statistisches Lernen, wie es heute praktiziert wird, funktioniert zwar häufig in der Praxis, aber die kausalen Abläufe bleiben oft unverstanden und eine Black Box. Statistisches Testen und Probieren reicht für sicherheitskritische Systeme nicht aus. Daher plädiere ich in der

Zukunft für eine Kombination von kausalem Lernen mit zertifizierten KI-Programmen durch Beweisassistenten, auch wenn das für Praktiker aufwendig und ambitioniert erscheinen mag.

4 Verifikation und Technikgestaltung

KI-Programme treten mittlerweile aber nicht nur in einzelnen Robotern und Computern auf. So steuern bereits lernfähige Algorithmen die Prozesse einer vernetzten Welt mit exponentiell wachsender Rechenkapazität. Ohne sie wäre die Datenflut im Internet nicht zu bewältigen, die durch Milliarden von Sensoren und vernetzten Geräten erzeugt wird. Aufgrund der Sensoren kommunizieren nun also auch Dinge miteinander und nicht nur Menschen. Daher sprechen wir vom Internet der Dinge (Internet of Things: IoT). Im industriellen Internet („Industrie 4.0“) wird das Internet der Dinge auf die Industrie- und Arbeitswelt angewendet. Künstliche Intelligenz und Machine Learning werden dazu in den Arbeitsprozess integriert. Werkstücke kommunizieren untereinander, mit Transporteinrichtungen und beteiligten Menschen, um den Arbeitsprozess flexibel zu organisieren. Produkte können so individuell zur gewünschten Zeit nach Kundenwünschen erstellt werden. Technik, Produktion und Markt verschmelzen zu einem soziotechnischen System, das sich selbst flexibel organisiert und sich verändernden Bedingungen automatisch anpassen soll.

Die sicherheitskritischen Herausforderungen, die wir eben erörtert haben, werden sich in solchen Infrastrukturen noch einmal potenzieren. Darüber hinaus stellt sich aber die Frage nach der Rolle des Menschen in einer mehr oder weniger automatisierten Welt. Ich plädiere daher für Technikgestaltung, die über Technologiefolgenabschätzung hinausgeht. Die traditionelle Sicht, die Entwickler einfach werkeln zu lassen und am Ende die Folgen ihrer Ergebnisse zu bewerten, reicht aus Erfahrung nicht aus. Am Ende kann das Kind in den Brunnen gefallen sein und es ist zu spät. Nun lässt sich zwar Innovation nicht planen. Wir können aber Anreize für gewünschte Ergebnisse setzen. Ethik wäre dann nicht Innovationsbremse, sondern Anreiz zu gewünschter Innovation. Eine solche ethische, rechtliche, soziale und ökologische Roadmap der Technikgestaltung für KI-Systeme würde der Grundidee der sozialen Marktwirtschaft entsprechen, nach der ein Gestaltungsspielraum für Wettbewerb und Innovation gesetzt wird. Maßstab bleibt aber die Würde des einzelnen Menschen, wie sie im Grundgesetz der Verfassung als oberstes Axiom der parlamentarischen Demokratie festgelegt ist.

Literaturverzeichnis

[Ac78] Aczel, P.: The type theoretic interpretation of constructive set theory., in: A. Macintyre, L. Pacholski, J. Paris (eds.), Logic Colloquium ’77. North-Holland : Amsterdam-New York, 55–66, 1978.

[AZ01] Aigner, M. und Ziegler, G.M.: Proofs from The Book, Berlin 2. Aufl., 2001.

[AW09] Awodey, S. Warren, M.A.: Homotopy theoretic models of identity type, in: Mathematical Proceedings of the Cambridge Philosophical Society 146 (1), 45-55, 2009.

[BC04] Bertot, Y. und Castéran, P.: Interactive Theorem Proving and Program Development: Coq'Art: CiC, Springer, 2004.

[BG18] Bringsjord, Selmer; N. S. Govindarajulu, Naveen Sundar. Artificial Intelligence, The Stanford Encyclopedia of Philosophy (Fall 2018 Edition), Edward N. Zalta (ed.), https://plato.stanford.edu/archives/fall2018/entries/artificial-intelligence.

[CM81] Clocksin, William F.; Mellish, Christopher S.: Programming in Prolog. Berlin: Springer, 1981.

[CJ96] S. Coupet-Grimal; L. Jakubiec: Coq and Hardware Verification: a Case Study (TPHOLs , 1996, LCNS 1125, 125-139).

[CH88] Coquand, T; Huet, G.: The calculus of constructions (Coc), in: Information and Computation 76(2–3), 95–120, 1988.

[DP60] Davis, M.; Putnam, H.: A computing procedure for quantification theory, in: J.ACM 7(3): 210-215, 1960.

[Da+62] Davis, M.; Logemann, G.; Loveland, D.: A machine program for theorem-proving, in: Commun. ACM 5(7): 394–397, 1962.

[DP16] Dybjer, P. und Palmgren, E.: Intuitionistic Type Theory, in: The Stanford Encyclopedia of Philosophy, The Metaphysics Research Lab (CSLI), Stanford University: Stanford (open access), 2016.

[Fe96] Feferman, S.: Kreisel's „unwinding" Program, in: P. Odifreddi (Ed.) Kreisleriana. About and Around Georg Kreisel, Review of Modern Logic, 1996, 247-273.

[FH07] Fränzle, Martin und Christian Herde: HySAT: An efficient proof engine for bounded model checking of hybrid systems, in: Formal Methods in System Design 30:179-198, 2007.

[Gi60] Gilmore, P. C.: A proof method for quantification theory: its justification and realization, in: IBM J. Res. Dev. 4(1): 28–35, 1960.

[Gi96] Giménez, E.: Un calcul de constructions infinies et son application à la vérification de systèmes communicants (PhD thesis Lyon), 1996.

[He34] Heyting, A.: Mathematische Grundlagenforschung. Intuitionismus, Beweistheorie, Springer, Berlin, 1934, repr. 1974.

[Ho69] Howard, W.A.: The formulae-as-types notion of construction, in: J. P. Seldin, J.R. Hindley (Hrsg.), To H.B. Curry: Essays on Combinatory Logic, Lambda Calculus and Formalism, Academic Press: Boston, MA, 479–490, 1969.

[Jo02] Joyal, A.: Quasi-categories and Kan complexes, in: Journal of Pure and Applied Algebra 175, 207-222, 2002.

[Kn16] Knuth, D. E.: Satisfiability. The Art of Computer Programming, Vol 4, Fasc. 6. Boston: Addison Wesley, 2016.

[Ko08] Kohlenbach, U.: Applied Proof Theory: Proof Interpretations and Their Use in Mathematics, Berlin, 2008.

[Ko32] Kolmogorov, A.N.: Zur Deutung der intuitionistischen Logik, in: Math. Z. 35, 58-65, 1932.

[KS00] Küchlin, Wolfgang; Sinz, C.: Proving Consistency Assertions for Automotive Product Data Management, in: J. Automated Reasoning 24:145-163, 2000.

[KM20] Küchlin, W.; Mainzer, K.: Logische Grundlagen der klassischen KI, in: K. Mainzer (Hrsg.), Philosophisches Handbuch der Künstlichen Intelligenz, Springer, 2020.

[Ma20] Mainzer, K.: Leben als Maschine: Wie entschlüsseln wir den Corona-Kode? Von der Systembiologie und Bioinformatik zu Robotik und Künstlicher Intelligenz, Brill Mentis: Paderborn 2. erweiterte Auflage, 2020.

[Ma19] Mainzer, K.: Künstliche Intelligenz. Wann übernehmen die Maschinen? Springer: Berlin 2. erweiterte Auflage (engl. Übersetzung), 2019.

[Ma18] Mainzer, K.: The Digital and the Real World. Computational Foundations of Mathematics, Science, Technology, and Philosophy, World Scientific Singapore, 2018.

[MSS18] Mainzer, K; Schuster, P.; Schwichtenberg, H. (Eds.): Proof and Computation. Digitization in Mathematics, Computer Science, and Philosophy. World Scientific Singapore, 2018.

[Ma95] Marques-Silva, J. P.: Search Algorithms for Satisfiability Problems in Combinatorial Switching Circuits. PhD Thesis, U. Michigan, 1995.

[MS96] Marques-Silva, J. P.; Sakallah, Karem A.: GRASP-A new search algorithm for satisfiability, in: Proceedings of International Conference on Computer Aided Design, 220-227, 1996.

[MS99] Marques-Silva, J. P.; Sakallah, Karem A.: GRASP: A search algorithm for propositional satisfiability, in: IEEE Transactions on Computers 48 (5): 506-521, 1999.

[Ma98] Martin-Löf, P.: An intuitionistic theory of types. Twenty-five years of constructive type theory (Venice, 1995), in: Oxford Logic Guides 36, Oxford University Press, New York, 127-172, 1998.

[MJS13] Mooij, J.M. ; Janzing, D.; B. Schölkopf, B.: From ordinary differential equations to structural causal models: The deterministic case, in: Proceedings of the 29th Annual Conference on Uncertainty in ArtificiaL Intelligence (UAI), 440-448, 2013.

[Ni71] Nilsson, Nils J.: Problem Solving Methods in Artificial Intelligence. New York: Mc Graw Hill, 1971.

[Ni80] Nilsson, N. J.: Principles of Artificial Intelligence. Palo Alto: Tioga Publishing Co., 1980.

[Pa93] Paulin-Mohring, C.: Inductive Definition in the System Coq: Rules and Properties (Research Report 92-49, LIP-ENS Lyon), 1993.

[Qu55] Quine, W. V. O.: A Proof Procedure for Quantification Theory, in: J. Symbolic Logic 20: 141-149, 1955.

[Sc06] Schwichtenberg, H.: Minlog, in: F. Wiedijk (ed.), The Seventeen Provers of the World. Lecture Notes in Artificial Intelligence vol. 3600, Berlin, 151-157, 2006.

[SW12] Schwichtenberg, H. und Wainer, S. S.: Proofs and Computations, Cambridge, 2012.

[UFP13] The Univalent Foundations Program: Homotopy Type Theory: Univalent Foundations of Mathematics. Princeton, NJ: Institute for Advanced Study, 2012.

[TB03] J. Tretmans; E. Brinksma: TorX: Automated model-based testing, in: A. Hartman and K. Dussa-Zieger (Eds.), Proceedings of the First European Conference on Model-Driven Software Engineering, 2003.

[Vo12] Voevodsky, W.: A universe polymorphic type system. http://uf-ias-2012.wikispaces.com/file/view/Universe+polymorphic+type+sytem.pdf, 2012.

Primzahlen als Herausforderung

Reinhard Kahle[1]

Abstract: Wir diskutieren, in welchem Maße zahlentheoretischen Fragestellungen von der neuen, statistikbasierten Künstlichen Intelligenz gelöst werden könnten und welche Auswirkungen das z.B. in der Kryptographie hätte. Dabei sehen wir die wesentliche Herausforderung darin, Methoden zu finden, die zu zeigen erlauben, was diese KI *nicht* leisten kann.

Keywords: Zahlentheorie; Neue KI; Diskrete Probleme; RSA

1 Primzahlen

Schon Euklid konnte beweisen, daß es unendlich viele Primzahlen gibt. Doch ist es Mathematikern bis heute nicht gelungen, eine eingängige Struktur in der Primzahlverteilung zu finden, die es über eine rein statistische Vorhersage über die Anzahl der Primzahlen in einem Intervall erlauben würde, natürliche Zahlen „einfach“ auf ihre Primzahleigenschaft hin zu prüfen. Das *Sieb des Erathostenes* liefert zwar eine sichere Methode, um zu überprüfen, ob eine Zahl prim ist, nur ist dieses Verfahren bei hinreichend großen Zahlen hoffnungslos aufwendig; und auch wenn wir heute bessere Methoden kennen, gibt es keine, die nach den in der Komplexitätstheorie entwickelten Kriterien als *schnell* eingestuft werden kann. Das gilt auch nicht für die Methode, die dem bedeutenden Resultat zugrunde liegt, daß die Primzahleigenschaft *polynomiale Berechnungskomplexität* hat [AKS04]; dabei erweist sich nicht nur der Grad des Polynoms für die Laufzeit als problematisch, sondern auch die Platzkomplexität. In der Praxis läßt sich aber zum Glück mit probabilistischen Methoden für Zahlen in der aktuell relevanten Größenordnung der Kryptographie in ausreichender Zeit und mit hinreichender Sicherheit überprüfen, ob diese prim sind (*Miller-Rabin-Test*, [Ra76]) .

Die Bedeutung, die Primzahlen insbesondere in der Kryptographie haben, legt nahe, daß es eine interessante Frage sein sollte, ob es der modernen, auf statistischen Methoden basierenden KI gelingen könnte, einen *schnellen* und zuverlässigen Primzahltest zur Verfügung zu stellen. Dazu gibt es Untersuchungen (siehe z.B. [ES06]), die aber – soweit wir wissen – zur Zeit noch nicht „in Konkurrenz“ zu den etablierten Algorithmen treten können.

[1] Carl Friedrich von Weizsäcker-Zentrum, Universität Tübingen, Keplerstr. 2, 72074 Tübingen, Deutschland, und CMA, FCT, Universidade Nova de Lisboa, 2829-516 Caparica, Portugal, kahle@mat.uc.pt

doi:10.18420/inf2020_64

Aus theoretischer Sicht ist die Suche nach einem Primzahltest durch die „neue KI“ eine Aufgabe, die sich in einem größeren Kontext stellt: In welchem Maße kann die KI *Strukturen entdecken*, die Mathematikern (bisher) verborgen geblieben sind.

Es gilt als Stärke der neuen KI, daß sie in der Lage ist, bei geeignetem Training in großen Datenmengen („Big Data“) *Strukturen* zumindest implizit ausfindig zu machen, die es erlauben, anschließend auch Einzeldaten im Hinblick auf verschiedene Eigenschaften hin zu klassifizieren.

2 Teilbarkeit

In unserem Kontext wollen wir derartige Strukturen innerhalb der Arithmetik ausnutzen. Wenn wir hier von Strukturen sprechen, meinen wir Gesetzmäßigkeiten, die sich in mathematischen Objektbereichen, wie z.B. den natürlichen Zahlen, aufweisen lassen.

Die Eigenschaft, gerade zu sein, läßt sich für eine natürlich Zahl bereits dadurch nachprüfen, daß man nachsieht, ob die letzte Ziffer dieser Zahl (in Dezimaldarstellung) 0, 2, 4, 6 oder 8 ist; für die Teilbarkeit durch drei ist es ausreichend, zu prüfen, ob die Quersumme der untersuchten Zahl durch drei teilbar ist. Im Fall der Primzahlen, für die zwar das Sieb des Erathostenes als Algorithmus zur Verfügung steht, ist es aber eine offene Frage, ob es eine in den natürlichen Zahlen verborgene Struktur gibt, die einen „einfachen“ Primzahltest ermöglichen würde.

Ohne daß wir das effektive durchgeführt hätten, gehen wir hier davon aus, daß die statistikbasierte KI in der Lage sein sollte, zu *lernen*, wann eine natürliche Zahl gerade oder durch drei teilbar ist.

Wenn eine KI-Software die Gradzahligkeit tatsächlich so gelernt hat, daß nur die letzte Ziffer getestet wird, sollte sich das im Prinzip durch Laufzeitüberprüfung feststellen lassen – ohne daß dafür die *Black Box* geöffnet werden müßte.

Im Fall der Teilbarkeit durch drei ist nicht klar, ob sich die Form, in der der Test von einer KI-Software durchgeführt wird, rekonstruieren ließe. Aber es auch nicht zu sehen, warum die Software die Idee der Quersumme, wenn auch nur implizit, verwenden sollte.

Allgemein können die Teilbarkeitstests für verschiedene Zahlen sehr unterschiedlich ausfallen, so daß man fragen kann, wie die KI hier eigentlich vorgehen dürfte. Hat sie z.B. eine Möglichkeit, aus den trivialen Teilbarkeitstest für 2 und 5 *selbständig* den trivialen Test für 10 zu erlernen? Kann Sie *bessere* Tests für die Teilbarkeit durch 17 finden, als eine *brute-force* Methode, die man auf Grundlage der Schulmathematik verwenden würde?

In diesem Zusammenhang sei auch kurz auf ein „Problem“ der Mathematik in der Aufstellung

ihrer Hilfsformeln verwiesen: Für die Determinantenberechnung einer gegebenen Matrix mit Hilfe des Laplaceschen Entwicklungssatzes werden gerne Formeln der Form:

$$\det A = \sum_{i=1}^{n} (-1)^{i+j} \cdot a_{ij} \cdot \det A_{ij}$$

benutzt. Der Ausdruck $(-1)^{i+j}$ bestimmt dabei das Vorzeichen eines jeden Summanden. In einer naiven algorithmischen Ausführung würde er $i + j$ viele Multiplikationen nach sich ziehen, die offensichtlich unnötig sind, da man lediglich eine Fallunterscheidung braucht, die prüft, ob $i + j$ gerade oder ungerade ist. Da die übliche Formelsprache der Arithmetik aber keine Fallunterscheidung kennt, bedient man sich eines formal korrekten Ausdrucks, der aber einen „algorithmischen Overkill" darstellt. Wenn wir entsprechende funktionale Zusammenhänge von einer KI-Software lernen lassen, ist zu erwarten, daß diese sich nicht an tradierten Darstellungsformen orientiert, sondern beliebige algorithmische Lösungen finden kann. Wie der benutzte Algorithmus konkret aussieht, würde man erst wissen, wenn man die der KI zugrundeliegende *Black Box* öffnen könnte. Doch könnten Laufzeitbetrachtungen eventuell gewisse Hinweise bringen, auch wenn sich die *Black Box* nicht öffnen läßt.

Dementsprechend geht es uns hier nicht darum, die von der KI implizit verwendeten Strukturen bzw. Algorithmen, explizit zu machen (also die *Black Box* zu öffnen); im Gegenteil, es geht gerade um die Beurteilung der (potentiellen) Leistungsfähigkeit der KI, wenn man ihre detaillierte Vorgehensweise *nicht* kennt.

Was die Primzahlen betrifft, könnte die Hoffnung sein, daß die KI einen *einfachen* Test erlernen könnte, der uns einen *schnellen* Primzahltest zur Verfügung stellen würde. Eine solche Hoffnung wäre mit der Überzeugung verbunden, daß die Primzahleigenschaft auf einer *einfachen* strukturellen Eigenschaft beruht, die die Mathematiker bisher nicht entdeckt haben, die aber die KI entdecken könnte. Allerdings kann es natürlich auch der Fall sein, daß eine solche einfache Eigenschaft gar nicht existiert und alle Versuche daher umsonst wären. Zunächst wollen wir auf ein „Standardargument" eingehen, warum die neue KI nicht spezifisch für Primzahltests ausgelegt sein sollte, und das auf den Unterschied von diskreten und kontinuierlichen Fragestellungen verweist.

3 Diskrete versus kontinuierliche Probleme

Hermann Weyl gab dem für uns wichtigen Unterschied von diskreten und kontinuierlichen Fragestellungen eine sehr schöne Einkleidung, allerdings unter der Annahme, daß die zahlentheoretischen Eigenschaften im wesentlichen nur einer Zahlenmagie dienen – die folgenden Zeilen wurden vor der Entdeckung der Bedeutung der Zahlentheorie für die Kryptographie geschrieben [We71, S. 36f]:

> PLATO übernimmt ein gut Stück der pythagoreischen Zahlenweisheit; aber die Zahl der Bürger einer idealen Stadt, welche er zu 5040 = 7! ansetzt . . . schein[t] seine eigene numerlogische Erfindung zu sein. . . .
>
> Ich möchte nur auf einen Zug hinweisen, der für diese Denkweise kennzeichnend zu sein scheint: was in der Zahlenmagie gilt, sind die *zahlentheoretischen* Eigenschaften der Zahlen; was in der Naturwissenschaft gilt, sind ihre *Größen*eigenschaften. Vom Größen-Standpunkt macht es wenig Unterschied, ob die Zahl der Bürger einer Stadt 5040 beträgt oder 5039, vom zahlentheoretischen Standpunkt ist da ein himmelweiter Abstand; z.B. besitzt $5040 = 2^4 \cdot 3^2 \cdot 5 \cdot 7$ viele Teiler, während 5039 eine Primzahl ist. Wenn in PLATOS idealer Stadt *ein* Bürger über Nacht stirbt und dadurch die Bürgerzahl auf 5039 erniedrigt, ist sie sogleich völlig korrumpiert.

Aber nicht nur Platos ideale Stadt hätte unter einem solchen kleinen Unterschied zu leiden, jedes moderne kryptographische Protokoll würde in gleicher Weise korrumpiert.

Wenn nun die neue KI ganz wesentlich auf kontinuierliche (stetige) Eigenschaften der von ihr (implizit) benutzten Funktionen aufbaut, könnte dies durchaus ein Argument sein, warum zahlentheoretische Fragestellungen dieser KI nicht unmittelbar zugänglich sind.

Allerdings hat auch die Zahlentheorie erkannt, daß sich tiefliegende Erkenntnisse über natürliche Zahlen gewinnen lassen, wenn man reelle Zahlengrößen mit hinzunimmt. Euler begründete die *analytische Zahlentheorie*, in der Methoden der Differential- und Infinitesimalrechnung erfolgreich in der Zahlentheorie zur Anwendung kommen. Insbesondere eröffnete sich damit die Möglichkeit zur Formulierung und zum Beweis des Primzahlsatzes, der besagt, daß sich die Anzahl der Primzahlen unterhalb von x asymptotisch an $\frac{x}{\ln(x)}$ annähert.

Im Kontext der analytischen Zahlentheorie läßt sich zumindest nicht mehr mit dem Unterschied von diskreten und kontinuierlichen Fragestellungen argumentieren. Trotzdem ist es – nach unserem Wissen – eine offene Frage, welche zahlentheoretische Probleme der neuen KI zugänglich sind. Im Hinblick auf bekannte Resultate wäre es durchaus interessant zu sehen, ob sich der Primzahlsatz auch durch reine KI-Analyse gewinnen ließe, ebenso wie viele der überraschenden zahlentheoretischen Zusammenhänge, die vor allem Euler und der geniale indische Mathematiker Ramanujan gefunden haben. Noch spannender wäre es natürlich, wenn die KI zur Lösung von offenen Fragen beitragen könnte, z.B. ob es nur endlich viele oder unendlich viele Primzahlzwillinge gibt.

Unser Interesse hier richtet sich jetzt aber auf die Frage, was sich *nicht* berechnen, oder zumindest *nicht effizient* berechnen läßt. Diese Frage läßt sich im Diskreten sehr detailliert untersuchen.

4 Berechenbarkeits- und Komplexitätstheorie

Die formale Berechenbarkeitstheorie für Funktionen auf den natürlichen Zahlen wurde vor allem im Anschluß an Hilberts *Entscheidungsproblem* sowie das für die Klärung der Grundlagen der Mathematik entwickelte *Hilbertsche Programm* ausgearbeitet. Dabei hat sich das von Turing aufgestellte Berechenbarkeitsmodell der *Turing-Maschine* besonders bewährt. Die Beobachtung, daß alle alternativen Berechenbarkeitsmodelle äquivalent zu Turings Modell sind (soweit sie sich nicht als schwächer herausgestellt haben), hat ihren Niederschlag in der *Churchen These* gefunden, daß der (informale) Berechenbarkeitsbegriff durch die Turing-Maschinen (oder jedes äquivalente Modell) bereits erschöpfend charakterisiert ist.

Damit haben wir ein Werkzeug an der Hand, mit dem die Grenzen der Berechenbarkeit formal erfaßt werden können. Insbesondere ließ sich mit Hilfe eines Diagonalisierungsarguments zeigen, daß das *Halteproblem*, d.h. die Frage, ob ein gegebenes Programm zur Berechnung einer partiellen Funktion auf den natürlichen Zahlen für einen gegebenen Eingabewert ein Ergebnis liefert oder nicht, unentscheidbar ist.

Bemerkung Die in der Churchen These zum Ausdruck kommende Grenzziehung war den an der Entwicklung der Berechenbarkeitstheorie beteiligten Protagonisten nicht von vorneherein klar. So hatte Gödel im Rahmen des Beweises seiner Unvollständigkeitssätze den (Turing-vollständigen) Berechenbarkeitsbegriff der *rekursiven Funktionen* herausgearbeitet. Über die Tragweite der Unvollständigkeitssätze war er sich aber bei der Abfassung seiner epochalen Arbeit [Gö31, S. 197] noch nicht im klaren, als er schrieb:

> Es sei ausdrücklich bemerkt, daß Satz XI [*der zweite Gödelsche Unvollständigkeitssatz*] ... in keinem Widerspruch zum Hilbertschen formalistischen Standpunkt steh[t]. Denn dieser setzt nur die Existenz eines mit finiten Mitteln geführten Widerspruchsfreiheitsbeweises voraus und es wäre denkbar, daß es finite Beweis gibt, die sich in *P* [die von Gödel benutzte Axiomatisierung der Arithmetik] *nicht* darstellen lassen.

In einer Bemerkung, die Gödel 1963 einer englischen Übersetzung seines Artikels hinzufügte, gesteht er dann zu, daß es (erst) Turings Analyse des Berechenbarkeitsbegriff war, die ihn von der uneingeschränkten Gültigkeit seines Resultates überzeugte [Gö67, S. 616]:

> In consequence of later advances, in particular of the fact that due to A. M. Turing's work a precise and unquestionably adequate definition of the general notion of formal system can now be given, a completely general version of Theorems VI and XI is now possible.

Wir haben diese historische Bemerkung hier aufgenommen, um zu zeigen, daß sich die Entwicklung einer neuen Theorie, wie wir sie hier als Herausforderung für die KI formulieren,

sicherlich nicht gradlinig und direkt erfolgen wird, sondern erst durch ein Wechselspiel von unterschiedlichen Ansätzen gewonnen werden kann. ⊣

Über ihre Funktion für Unentscheidbarkeitsresultate hinaus hat sich das Konzept der Turing-Maschine vor allem auch dadurch bewährt, daß es eine eingängige Definition von Berechenbarkeitskomplexität ermöglichen. Für einen gegebenen Eingabewert kann man die Schritte zählen, die eine gegebene Turing-Maschine benötigt, um ein Ergebnis zu liefern. Die Komplexität für den durch die Turing-Maschine bestimmten Algorithmus zur Berechnung einer zahlentheoretischen Funktion ergibt sich dann, salopp formuliert, durch eine Funktion, die die entsprechende Schrittzahl majorisiert. Damit kann man z.B. die Klasse *Ptime* der *polynomialen* Funktionen definieren, die Algorithmen besitzen, deren Schrittzahl sich polynomial beschränken läßt. Bekannt ist auch die Klasse *NP*, bei der die Lösung eines mit n parametrisierten Problems *geraten* werden kann, die Überprüfung der Richtigkeit der Lösung dann aber nur polynomiale Laufzeit in n benötigen darf. Wenn mit P die Klasse der polynomial berechenbaren Entscheidungsproblemen bezeichnet wird, ist die Frage $P \stackrel{?}{=} NP$ das zentrale offene Problem der theoretischen Informatik. Für die Verläßlichkeit der im folgenden diskutierten kryptographischen Protokolle geht man aber von der allgemein vermuteten Ungleichheit $P \neq NP$ aus. Ungeachtet unserer Bemerkung zum polynomialen Primzahltest werden in diesem Fall Probleme und Funktionen in P beziehungsweise *Ptime* als (im Prinzip) effizient lösbar beziehungsweise berechenbar betrachtet, während Probleme in *NP* zu den (mit heutigen Mitteln) nicht effizient lösbaren Problemen gehören.

Wir haben diese wohl allgemein bekannten, grundlegenden Konzepte der Komplexitätstheorie hier ausgeführt, weil wir im abschließenden Kapitel die Frage aufwerfen werden, was die entsprechende Komplexitätstheorie für die KI sein könnte, beziehungsweise, was das Fehlen einer solchen Komplexitätstheorie für die KI bedeutet.

5 Sind unsere kryptographischen Protokolle vor der neuen KI sicher?

Die modernen *public-key*-Protokolle der Kryptographie basieren zentral auf zahlentheoretischen *Einbahnstraßenfunktionen*: $f(x) = y$ soll einfach berechenbar sein (polynomial); die inverse Funktion $f^{-1}(y) = x$ soll mit öffentlich zugänglicher Information nur schwer (exponentiell) berechenbar sein; und allein mit Hilfe eines privaten Schlüssel ist auch sie „durch eine Hintertür“ einfach berechenbar.

Im bekannten RSA-Protokoll nutzt man an zentraler Stelle aus, daß die Multiplikation zweier Zahlen *einfach* ist, die Zerlegung einer Zahl in seine Primfaktoren aber *schwer*. Mit Hilfe einer schnellen Faktorisierung großer Zahlen würde man dieses Protokoll sofort kompromittieren.

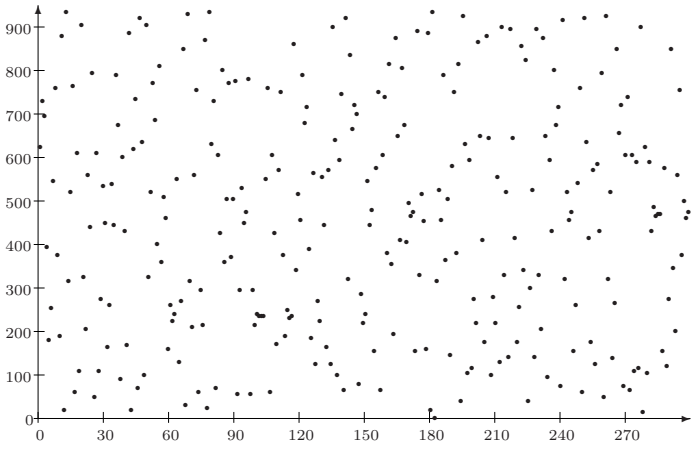
Damit stellt sich die Frage, ob die moderne KI unter Umständen eine solche schnelle Faktorisierung liefern könnte. Nach dem aktuellen Kenntnisstand ist das nicht der Fall und „es sieht auch nicht so aus“, daß die KI hierfür einen besonderen Ansatzpunkt hätte.

Aber haben wir eine Möglichkeit, den *Eindruck*, daß RSA – unter der vorausgesetzten Annahme von $P \neq NP$ – von Seiten der KI keine Gefahr drohe, durch formale Argumente zu untermauern? Anders gefragt: Was wäre die adäquate Berechenbarkeitstheorie für die neue KI, die es erlaube, formale Unentscheidbarkeits- und Komplexitätsresultate zu zeigen, wie wir sie aus der Turing-Berechenbarkeit kennen?

Hierbei sollte eine solche *KI-Berechenbarkeitstheorie* nicht von einer „Öffnung" der *Black Box* abhängen (denn dann sollte sich die traditionelle Berechenbarkeitstheorie auf den „verstandenen" Algorithmus anwenden lassen), sondern sie sollte Ergebnisse liefern, selbst wenn man den internen Ablauf der Entscheidungsfindung nicht kennt.

Eine konkrete Aufgabenstellung läßt sich am *Diskreten Logarithmus-Problem* (DLP) illustrieren, das dem *Diffie-Hellman-Schlüsselaustausch* und *ElGamal-Protokoll* zugrunde liegt. In einem endlichen Körper der Charakteristik p (p prim), läßt sich für eine gegebene primitive Wurzel g und gegebene Zahl x der Wert h in $g^x \equiv h \pmod{p}$ *schnell* berechnen. Umgekehrt ist die Berechnung von x bei gegebenen g und h aber – nach aktuellem Kenntnisstand – bei ausreichend großem p nicht mehr praktisch durchführbar.

In der folgenden Tabelle (entnommen aus [HPS08, S. 65]) sind die Werte von $627^i \bmod 941$ für $i = 1, 2, 3, \ldots$ eingetragen:



Während sich 627^i durch eine „übersichtliche" Exponentialkurve darstellen läßt, führt das „Herunterbrechen" der Funktionswerte an den ganzzahligen Punkten durch die mod 941-Operation zu einem Bild, in dem sich unmittelbar keine einfache Struktur mehr erkennen läßt. Würde eine KI-Software hier (und für anderen Zahlen g und p) aber doch eine einfache Struktur erlernen können, hätte sie auch eine Handhabe, um DLP zu lösen (und das unabhängig davon, ob diese Struktur uns zugänglich gemacht werden kann oder nicht).

Wie im Fall der Primzahlen ist es offen, ob eine solche Struktur existiert; aber um sich auf die Sicherheit der DLP-basierten kryptographischen Protokolle verlassen zu können, wäre es wünschenswert, daß man eine formale Analyse dessen hätte, was die KI in diesen (und ähnlichen) Fällen *nicht* leisten kann.

Schließlich wollen wir noch auf einen Problemkreis aufmerksam machen, der sich selbst der klassischen Komplexitätstheorie entzieht. Aus einem Resultat von Pollard [Po74] ergibt sich, daß RSA dann unsicher ist, wenn das dabei verwendete Produkt zweier großer Primzahlen eine Primzahl p benutzt, deren „Vorgänger" $p - 1$ ein Produkt viele kleiner Primzahlen ist. In einem Lehrbuch zur Kryptographie schreiben die Autoren dazu [HPS08, S. 136]:

> From a cryptographic perspective, the importance of Pollard's method lies in the following lesson. Most people would not expect, at first glance, that factorization properties of $p-1$ and $q-1$ have anything to do with the difficulty of factoring pq. The moral is that even if we build a cryptosystem based on a seemingly hard problem such as integer factorization, we must be wary of special cases of the problem that, for subtle and nonobvious reasons, are easier to solve than the general case.

Im Allgemeinen läßt sich nicht ausschließen, daß eine intelligente KI-Software eine „ELIZA $p + 2$"-Methode findet, bei der RSA dann kompromittiert ist, wenn $p + 2$ spezielle Eigenschaften aufweist.

Zusammenfassend argumentieren wir hier nicht dafür, daß die KI neue Wege finden soll, zahlentheoretische Fragestellungen effizient zu lösen, sondern, ganz im Gegenteil, daß die Herausforderung darin besteht, Methoden zu finden, die zu zeigen erlauben, was die KI *nicht* leisten kann. Das könnte zumindest für diejenigen von Interesse sein, die auch weiterhin Internet-Banking ohne Beunruhigung benutzen wollen.

Literatur

[AKS04] Agrawal, M.; Kayal, N.; Saxena, N.: PRIMES Is in P. Annals of Mathematics 160/2, S. 781–793, 2004.

[ES06] Egri, L.; Shultz, T. R.: A Compositional Neural-network Solution to Prime-number Testing. Proceedings of the Annual Meeting of the Cognitive Science 28, 2006.

[Gö31] Gödel, K.: Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme. Monatshefte für Mathematik und Physik 38, S. 173–198, 1931.

[Gö67] Gödel, K.: On formally undecidable propositions of *Principia Mathematica* and related systems I. In (van Heijenoort, J., Hrsg.): From Frege to Gödel: A Source Book in Mathematical Logic, 1879–1931. Harvard University Press, S. 596–616, 1967.

[HPS08] Hoffstein, J.; Pipher, J.; Silverman, J. H.: An Introduction to Mathematical Cryptography. Springer, 2008.

[Po74] Pollard, J. M.: Theorems on factorization and primality testing. Proceedings of the Cambridge Philosophical Society 76/3, S. 521–528, 1974.

[Ra76] Rabin, M. O.: Probabilistic algorithms. In (Traub, J. F., Hrsg.): Algorithms and complexity. Academic Press, S. 21–39, 1976.

[We71] Weyl, H.: Über den Symbolismus der Mathematik und mathematischen Physik. In (Reidemeister, K., Hrsg.): Hilbert. Springer, S. 20–38, 1971.

Diese Arbeit wurde u.a. von der Udo Keller-Stiftung und, in eine frühen Phase, von der VolkswagenStiftung im Rahmen des Projekts *Can Software be 'responsible'*? sowie durch die Portugiesische Forschungsgemeinschaft FCT über das *Centro de Matemática e Aplicações*, UID/MAT/00297/2020, gefördert.

Laßt hundert Blumen blühen

Wolfgang Bibel[1]

Abstract: Der Beitrag skizziert im ersten Teil einschlägige Aspekte der historischen Entwicklung der Künstlichen Intelligenz (KI) als Wissenschaft und daraus resultierender Technologie. Im zweiten Teil werden beispielhaft Synergien zwischen Verfahren des „deep learning“ (DL) und dem klassischen automatischen Beweisen erläutert. Aus diesen beiden Analysen werden im dritten Teil Empfehlungen für die weitere Entwicklung von Wissenschaft und Technologie der KI in Deutschland und Europa abgeleitet; spezifische Fragestellungen vor allem im Zusammenhang mit DL finden dabei eine besondere Berücksichtigung.

Keywords: Künstliche Intelligenz; Wissenschaftsgeschichte; Maschinelles Lernen; Deep Learning; Automatisches Beweisen

1 Einleitung

Der Workshop „Konzeptionelle Herausforderungen für die KI“ auf der Jahrestagung der Gesellschaft für Informatik, INFORMATIK2020, spricht in seinem Begleittext von einer „neue[n], auf statistischer Analyse großer Datenmengen aufbauende[n] KI“. Der vorliegende Beitrag zu diesem Workshop diskutiert zunächst die Frage, inwieweit man hierbei überhaupt von einer „neuen KI“ sprechen sollte. Dazu wird diesbezüglich auf die Entwicklung des Gebiets der Künstlichen Intelligenz (KI) eingegangen. Dieser Teil versteht sich insgesamt als ein Appell an alle in der KI und um sie herum, die zugrundeliegende Wissenschaft und die sich aus ihr nährende Technologie als ein großes Ganzes zu verstehen, jedes Teilgebiet darin erblühen und die so erwachsende Gesamtheit hin zu künstlicher allgemeiner Intelligenz synergetisch zu großer Stärke gedeihen zu lassen.

Im zweiten Teil wird dann an einem konkreten Forschungsthema illustriert, wie sich zwei Teilgebiete der KI gegenseitig befruchten und verstärken können. Es handelt sich zum einen um das heute so erfolgreiche Gebiet des maschinellen Lernens (ML), beispielsweise in der Ausprägung von „deep learning“ (DL), und zum anderen um das vor mehr als siebzig Jahren in Deutschland begründete Gebiet des automatischen Beweisens (ATP). Nur mit grundlegenden Beiträgen – mindestens auch – von diesen beiden lassen sich die wünschenswerten Ziele erreichen, die sich mit einer KI der Zukunft verknüpfen.

Im letzten Teil werden aus der vorangegangenen Analyse die Konsequenzen für eine künftige Entwicklung der KI gezogen. Es gilt, sich auch der Schwächen des DL anzunehmen, die

[1] Technische Universität Darmstadt, bibel@gmx.net

doi:10.18420/inf2020_65

übrigen Gebiete der KI gleichberechtigt im Fokus zu behalten und synergetisch zusammenzuführen. Als Lehre aus einer hierzulande nur bedingt erfolgreichen KI-Geschichte wird für Europa und Deutschland die Orientierung der Förderung von KI an einer wohlüberlegten Vision von Anwendungen auch auf Bereiche empfohlen, deren Handlungs- und Entscheidungsprozesse aus KI-Sicht bis heute in völlig inadäquater Weise angelegt sind. Eine solche Rationalisierung gesellschaftlichen Handelns würde tunlichst einhergehen mit entsprechenden institutionellen Neustrukturierungen, nicht zuletzt auch einer Neuordnung des Gefüges der Wissenschaften unter Berücksichtigung der seit Mitte des letzten Jahrhunderts erzielten Erkenntnisse.

Der Titel soll die vom Autor vertretene Strategie hin zu einer fruchtbaren wissenschaftlichen Entwicklung bildlich zum Ausdruck bringen. Dabei sei nicht verschwiegen, daß er einem Zitat von Mao Zedong aus dem Jahre 1956 entlehnt ist, das in seiner Gänze lautet: „Laßt hundert Blumen blühen, laßt hundert Schulen miteinander wetteifern“. Dieser historische Hinweis mag auch zur Erinnerung daran dienen, daß es mit einem schönen Appell allein nicht getan ist, sondern daß das anschließende Tun sich immer an den Worten orientieren und messen lassen muß, um Fehlentwicklungen schon im Keim ersticken zu können.

2 Anmerkungen zur Historie der KI

Das Gadamer- bzw. Heidegger-Zitat „Zukunft ist Herkunft“ komprimiert in drei Worten die historische Erfahrung der großen Bedeutung der Geschichte für die kluge Wahl eines Weges in die Zukunft. Um dem Anliegen dieses Workshops nach richtungsweisenden Konzepten für die KI Genüge tun zu können, beginnt dieser Beitrag daher mit einschlägigen Erinnerungen an die Historie der KI. Aufgrund dieser Historie wird sich auch erweisen, inwieweit es gerechtfertigt ist, von einer „neuen KI“ im Zusammenhang mit den derzeit so aktuellen Lernverfahren zu sprechen.

Eine international umfassende Darstellung der Geschichte der KI gibt es bis heute leider nicht. Das Nilssonsche Buch [Ni09] fokussiert fast ausschließlich auf die Entwicklung in den USA und erweist sich damit in entscheidenden Aspekten als unvollständig. So wird darin im Kapitel 3 auf die einschlägigen Inhalte von drei Konferenzen in den USA und einer in UK eingegangen, die in den Jahren 1948, 1955, 1956 und 1958 stattgefunden haben und als Beginn der KI eingeschätzt werden. Die Arbeiten von Konrad Zuse aus den Jahren 1945 bis 1950, repräsentiert durch die Veröffentlichungen [Zu48, Zu49, Zu50a, Zu50b], die die KI schon Jahre davor inhaltlich begründet haben, sind damit dort unerwähnt geblieben. Unverständlicherweise haben es auch die deutschen Begründer der Informatik bis heute versäumt, diesen historisch herausragenden Beitrag Zuses angemessen zu würdigen. So ist dieser bis heute so gut wie unbekannt geblieben. Ich möchte daher auch an dieser Stelle auf die mir in diesem Zusammenhang bislang bekannten folgenden Sachverhalte hinweisen (s. dazu auch [Bi20a]).

Aufgrund der besonderen Umstände der Nachkriegszeit fokussierte Zuse ab 1945 seine Arbeit auf eine umfassendere Darstellung seiner grundlegenden Erkenntnisse zum „allgemeinen Rechnen“. Hierzu sind im „Konrad Zuse Internet Archive“ umfangreiche handschriftliche Manuskripte wie beispielsweise seine „Theorie der angewandten Logistik“ verfügbar. Als unveröffentlichte Manuskripte begründen sie nur bedingt einen Anspruch auf urheberrechtliche Anerkennung. Völlig anders steht es jedoch mit den aus den Inhalten dieser Texte hervorgegangenen Publikationen, von denen es aus den Jahren 1948 bis 1950 mindestens zwei in Zeitschriften [Zu48, Zu49] sowie zwei Patentschriften [Zu50a, Zu50b] gibt. Diese vier Publikationen dokumentieren eindeutig die Zuseschen Forschungsergebnisse bis zum Jahre 1950, die sich wie folgt kurz zusammenfassen lassen.[2]

Danach war Zuse die Universalität seiner Rechner bereits voll bewußt. Er beschreibt beispielsweise detailliert, wie ein Rechner die Wohlgeformtheit einer logischen Formel nachweisen kann, dh. er gibt den Algorithmus (bei ihm „Vorschrift“ genannt) dafür an. Als formale Sprache zur Beschreibung des Algorithmus legt er seinen Plankalkül zugrunde, der die weltweit erste Programmiersprache darstellt. Insgesamt präsentiert er damit das weltweit erste nichtnumerische Programm. Der Plankalkül selbst erscheint nach einer groben Einsicht als eine logische Programmiersprache, vom grundlegenden Konzept her vergleichbar mit dem mehr als 25 Jahre später entstandenen PROLOG. Um die Implementierbarkeit des Plankalküls nachzuweisen, gibt Zuse zudem Algorithmen an, die den automatischen Beweis von Formeln in der Prädikatenlogik erbringen. Damit begründet er zugleich das Gebiet des Automatischen Beweisens (automated theorem proving, ATP) und zwar Jahre vor Newell, Shaw und Simon [NSS56] bzw. Davis [Da57], die bislang international als dessen erste Pioniere gelten. Da dieses eines der Kerngebiete der KI darstellt, begründet er inhaltlich damit implizit zugleich die KI [Bi20b]. Eine eingehende und sachkundige Analyse dieser herausragenden Beiträge steht bislang noch immer aus.

Vor dem Hintergrund dieser wissenschaftlich belegten Beiträge dürfen dann natürlich auch seine unveröffentlichten Manuskripte ergänzende Berücksichtigung in der Beurteilung Zuses historischer Beiträge finden. Beispielsweise ist seine Formalisierung des Schachspiels im Plankalkül bereits ab 1945 hier zu erwähnen, die seine breite KI-Sicht bestätigt. Insgesamt rundet sich damit das Bild von Zuse nicht nur als Erfinder des modernen Komputers sondern inhaltlich zugleich als Mitbegründer einer Wissenschaft ab, die hinter der KI steht (vgl. [Bi18]), und zwar von vergleichbarem Rang wie Alan Turing. Daß Zuses Beiträge weder auf die weitere Entwicklung des Komputers noch auf die der KI nachweisbaren Einfluß nehmen konnten, steht auf einem anderen Blatt und bedürfte einer zusätzlichen historischen Analyse.

In etwa zeitgleich mit Zuses beschriebenen Arbeiten ist Wieners Kybernetik-Buch [Wi48] entstanden, der in diesem Kontext nicht ungenannt bleiben sollte. Zudem gab es erste Überlegungen zur maschinellen Sprachübersetzung, beispielsweise durch Andrew Booth und Warren Weaver (s. [Bi14]). Etwas später publizierte auch Claude Shannon die Be-

[2] Die Hinweise auf diese Schriften Zuses verdanke ich Herrn Dr. Ralf Bülow, der auf deren Inhalte beispielsweise in [Bü20] aufmerksam gemacht hat.

schreibung von Algorithmen zum Schach [Sh50] in einer Weise, die über die analogen Beschreibungen von Zuse Jahre vorher nicht hinausgeht, sondern hinter diesen teilweise sogar zurückbleibt.[3] Damit entfällt die in [Bi14, Footnote 12] angebrachte Rechtfertigung für die Nichterwähnung von Zuse in diesem Kontext, dh. Zuse war auch der weltweit erste Begründer der Schachprogrammierung. Vor allem ist dann auch Alan Turing zu nennen, der mit seinem Mind-Artikel [Tu50] eine Leit-Perspektive für die KI schuf. Ihm war Zuse in Bezug auf das akademische Knowhow, vor allem in Bezug auf Publikationen unterlegen, während in ihren Visionen beide durchaus vergleichbar sind.

Nach diesen – bei weitem nicht vollständig aufgezählten – individuellen und inhaltlichen Ansätzen zur KI fanden dann die eingangs erwähnten Konferenzen, vor allem die Dartmouth Konferenz [Mc55] im Jahre 1956 statt, die als formeller Start der institutionalisierten KI angesehen wird. Sie hat als Kristallisationspunkt dazu geführt, daß sich die KI dann vor allem in den USA rasch entwickeln konnte. Eine Darstellung der geschichtlichen Entwicklung der KI ist offensichtlich aber sehr unvollständig, wenn sie erst mit diesen Konferenzen beginnt und die hier aufgelisteten Forscher neben weiteren unerwähnt läßt. In diesem Sinne ist die eingangs erwähnte Feststellung zu verstehen, daß eine solche umfassende Darstellung bis heute aussteht. Wie wir an einzelnen Punkten im Folgenden sehen werden, sind auch die Jahre danach historisch bislang nur lückenhaft aufgearbeitet.

Als Begründer des Gebietes von lernfähigen Neuronalen Netzen (NNs) und damit auch des Teilgebietes des maschinellen Lernens (ML) in der KI gilt Frank Rosenblatt mit seinem Perzeptron-Modell [Ro58]. In [Bi14] wurde jedoch darauf hingewiesen und detaillierter ausgeführt, daß Karl Steinbuch schon vorher die wesentlichen Ideen zu seiner Lernmatrix als Patent angemeldet hatte (s. auch [KIT06], wo es bestätigend heißt: „Während seiner Industrietätigkeit meldete Steinbuch mehrere Dutzend Patente an, darunter das für die "Lernmatrix"." – Steinbuchs reine Industrietätigkeit endete 1958 mit der Berufung als Professor). Wenn diese Aussagen in einer detaillierten Überprüfung bestätigt würden, dann gebührte der Lernmatrix damit nicht nur die zeitliche Priorität in Bezug auf NNs, sondern sie ist zusätzlich dem Perzeptron in Bezug auf Lernfähigkeiten weit überlegen.

Nach meinem auf Originalliteratur wie [St60] beruhenden eigenen Kenntnisstand hatte Steinbuch seine Lernmatrix spätestens 1960 entwickelt und als Patent publiziert, in dem Rosenblatt nicht zitiert ist.[4] Im Gegensatz zu Rosenblatts einfachen NNs mit zwei Eingabe- und einem Ausgabeneuron weist die Steinbuchsche Matrix $m \times n$ Schaltelemente bzw. Neuronen auf und ist daher wesentlich allgemeiner in dem Sinne, daß es wesentlich mehr Funktionen lernen kann.

Wie im Falle von Zuse wurde auch Steinbuch international so gut wie nicht zur Kenntnis genommen, obwohl er seine Ergebnisse 1963 zusätzlich in einem angesehen Journal der IEEE veröffentlichte. In einschlägigen Publikationen sucht man seinen Namen in diesem

[3] So schreibt Cannon noch 1948: „Our problem is to represent chess as numbers and operations on numbers“, während Zuse schon Jahre vorher zum allgemeinen Rechnen mit nichtnumerischen Objekten gelangt war.

[4] Den Hinweis auf die zitierte Patentschrift verdanke ich Christian Vater.

Kontext aber auch heute noch vergeblich. Aufgrund des Vorangehenden gebührt ihm in Bezug auf die Erfindung der NNs jedoch mindestens der gleiche Rang wie Rosenblatt, wenn nicht sogar ein höherer.

Die NNs haben seither eine wechselvolle Geschichte zu verzeichnen, die von den Fortschritten der Forschung auf diesem Gebiet vor allem in Nordamerika geprägt war. Sie ist, beginnend mit 1943, zB. in [AD20] beschrieben. Etwa um 2012 erreichte die Entwicklung der NNs beispielsweise mit der Publikation [KSH12], fußend auf ihrem Vorläufer [Le98], einen Höhepunkt, der auf der Grundlage der bis dahin insgesamt erreichten Ergebnisse seitdem zu unzähligen und tief beeindruckenden Anwendungen führte, die auch in der Öffentlichkeit großes Aufsehen erregen. Zu diesen gehören die Erkennung von Bildern, Gesichtern, Videos, Verstehen akustischer Sprache, Übersetzung von natürlicher Sprache, Generierung von Texten, Einsatz zur Steuerung autonomer Fahrzeuge, Automatisierung von Finanzinvestitionen uvam. Diese revolutionäre Leistungsverbesserung der NNs ist nun als „deep learning“ (DL) bekannt geworden. Der Erfolg von DL ist der Kombination von drei entscheidenden Faktoren zu verdanken: die zwischenzeitliche Verfügbarkeit wesentlich größerer Datenmengen, die zugrundeliegenden algorithmischen Innovationen, mittels derer diese komplexen NNs auf der Grundlage von großen Datenmengen trainiert werden können, und die Implementierung dieser Algorithmen auf graphischen Prozessoren (GPU), die die erforderliche Rechenleistung erbringen.

Dieser kurze Abriß der historischen Entwicklung der NNs macht deutlich, daß es sich bei DL um eine ganz normale Fortentwicklung eines wissenschaftlichen Teilgebiets innerhalb des ML handelt, wenn auch einem technologisch aktuell extrem erfolgreichen. Und ML bleibt auch weiterhin nur eines der vielen Teilgebiete innerhalb der KI. DL als „neue KI“ zu bezeichnen ist daher sachlich völlig unbegründet.

Diese und andere Fehleinschätzungen im Zusammenhang mit KI haben auch damit zu tun, daß generell der Status der KI innerhalb der Wissenschaften bis heute unsachgemäßen Vorstellungen unterliegt. Die Pioniere der KI, von denen oben die Rede war, haben in ihren Publikationen klar zu erkennen gegeben, daß die Erfindung des Komputers zugleich den Start für die Entwicklung einer neuen Naturwissenschaft vom Range der Physik und der Biologie ausgelöst hat.

Diese beiden letztgenannten Disziplinen samt den jeweiligen Clustern von Disziplinen um sie herum waren extrem erfolgreich darin, Erklärungen für die Phänomene der Welt in naturwissenschaftlich präziser Weise mittels experimentell begründeten Theorien zu erarbeiten. Ihrem Zugriff entzogen sich jedoch bis zur Mitte des letzten Jahrhunderts ein großes Bündel von unsichtbaren und ungreifbaren Phänomenen, für die bis dahin daher nur spekulative Erklärungen aufgestellt werden konnten. Dazu gehören der Austausch und die Auswertung von Signalen und Informationen, Wahrnehmungen, Kommunikation, mentale Einstellungen, Intensionen, Zielvorgaben, Empfindungen, Emotionen, Aktionssteuerung, Bewußtsein, Denken, Psyche, Geist, Intellekt usw. Zwar ließen sich derartige Phänomene in ihren Auswirkungen teilweise beobachten und die Beobachtungen konnten beschrieben wer-

den. Zu brauchbaren Erklärungen konnte das verfügbare wissenschaftliche Instrumentarium jedoch nicht ausreichen. Genau das hat sich mit der Erfindung des universellen Komputers grundlegend geändert und die Entwicklung einer neuen Naturwissenschaft ausgelöst, die in den ersten Jahrzehnten vor allem von den KI-Pionieren und Informatikern vorangetrieben wurde. Diese Entwicklung hat zugleich zu einer technologischen Entwicklung von bis dahin ungekanntem Ausmaß geführt, die noch immer mit einem stetig zunehmenden Tempo andauert.

Die hiermit beschriebene neue Naturwissenschaft wird bis heute nicht als einheitliche Wissenschaft verstanden; vielmehr wird sie innerhalb unterschiedlichster Disziplinen in verzettelter Weise betrieben. Zu diesen gehören KI, Informatik, Neurowissenschaften, Kognitionswissenschaft, (Kognitions-) Psychologie, Philosophie, Verhaltensforschung uvam. Sie alle fokussieren in ihren einschlägigen Zielrichtungen auf das oben beschriebene Bündel natürlicher Phänomene, leider oft ohne Kenntnisnahme der entsprechenden oder einschlägigen Erkenntnisse zu den analogen Phänomenen in anderen dieser Disziplinen. Das folgende Beispiel möge diese bedauerliche Zersplitterung illustrieren.

In [HFB20] wird über biologische Experimente berichtet, die in der Beobachtung des Verhaltens von Ameisen (temnothorax albipennis ants) bestanden. Aufgrund dieser Beobachtungen wurde das Verhalten in einem einfachen algorithmischen Modell simuliert. Wie man den Referenzen der Veröffentlichung entnehmen kann, haben die Autoren offenbar nicht die geringste Kenntnis von der Tatsache, daß Ameisenalgorithmen seit Jahrzehnten in der KI intensivst studiert werden [DS04], mittels derer die Simulation des Ameisenverhaltens schon weit fortgeschritten ist. Vielmehr beginnen sie mit ihren Studien wieder bei Null, wenn auch unter Verwendung informatischer Systeme, anstatt an den in der KI inzwischen diesbezüglich erreichten Stand anzuknüpfen und ihn weiter voranzutreiben. Das ist reine und bedauerliche Verschwendung von Forschungskraft und Forschungsgeld. Das Beispiel ist dabei leider nur eines unter Tausenden analoger Forschungen.

Die Ursache für diese exorbitante Verschleuderung von Forschungspotenzial liegt eindeutig in der beschriebenen Zersplitterung einer Wissenschaft, die durch den gemeinsamen Forschungsgegenstand und die mit dem Komputer nun verfügbare Methodik charakterisiert ist, wie beispielsweise in [Bi18] genauer beschrieben wurde. Dort wird auch erläutert, daß sich infolge dieser Zersplitterung bis heute auch keine Bezeichnung dieser Wissenschaft eingebürgert hat. Um sie benennen zu können, wird dort die Kunstbezeichnung IPsI-Wissenschaft (kurz für *I*nformation, *Ps*ychologie, *I*ntelligenz) verwendet, die anstelle von KI auch hier gelegentlich als Notbehelf dienen mag, wenn die Betonung auf KI vor allem als allgemeiner Wissenschaft – und nicht nur als daraus hervorgegangener Technologie – liegt.

Die Methode von IPsI besteht in der Modellierung auf dem universellen Komputer, beispielsweise zur Simulierung des Verhaltens von Ameisen. Die Modellierung erlaubt dann Experimente, mathematische Theoriebildung und deren experimentelle Bestätigung. Dies ist das grundlegende und charakteristische Vorgehen einer jeden Naturwissenschaft und Kern des bereits Jahrhunderte anhaltenden Erfolgs der Naturwissenschaften. IPsI

spielt in den aufgezählten und vielen weiteren Disziplinen zunehmend eine zentrale Rolle. Vormals geisteswissenschaftlich oder vorwissenschaftlich betriebene Forschung wird dabei nunmehr mit präzisen Methoden naturwissenschaftlich durchgeführt und erzielt so fundierte und jederzeit nachprüfbare Erkenntnisse analog wie in Physik, Biologie etc. anstelle der vormals rein spekulativen Vorstellungen und deren Tradierungen. Dies ist die eigentliche wissenschaftliche Revolution, die mit der Erfindung des Komputers ausgelöst wurde.

Es gibt unterschiedliche Ansätze, um zu einer Modellierung eines natürlichen Phänomens zu gelangen. Sie mag auf der Grundlage von Beobachtungen von Hand programmiert und die dabei entstehenden Programme dann schrittweise verbessert werden. Mittels Lernverfahren wie DL kann dieser iterative Prozeß erheblich beschleunigt und automatisiert werden. Anstelle derartiger phänomenologisch geprägter Ansätze kann man auch versuchen, die Maschinerie der Natur unmittelbar künstlich zu simulieren und so beispielsweise die neuronale Struktur der Ameisen auf den Rechner abzubilden. Um welches natürliche Phänomen es sich bei einer solchen Simulation handelt, spielt für das methodische Vorgehen eine zweitrangige Bedeutung. Die gleiche – hier für das Verhalten von Ameisen illustrierte – Methode hat sich bei jedem der natürlichen Phänomene aus dem oben beschriebenen Bündel erfolgreich bewährt und es so jeweils erstmalig einer naturwissenschaftlichen Untersuchung zugänglich gemacht.

3 Integrationsschritte am Beispiel von ATP und ML

Eines der natürlichen Phänomene aus dem im letzten Abschnitt beschriebenen Bündel ist das mathematische Beweisen auf der Grundlage präziser logischer Schlüsse. Diese Fähigkeit des menschlichen Geistes hat seit Jahrtausenden die Grundlagen für ein wissenschaftlich fundiertes Zurechtfinden in der Welt gelegt. Zudem gilt das mathematische Denken als Modell und Vorbild für logisches Denken allgemein. Es war daher auch kein Zufall, daß Zuse genau dieses Phänomen für die Demonstration der Potenz der neuen Rechner in den frühen Pionierarbeiten [Zu48, Zu49] ausgewählt hat. Seither hat das automatische Beweisen (ATP) innerhalb der KI eine beeindruckende Entwicklung erleben dürfen und in einer Reihe von Fällen bereits die entsprechenden menschlichen Fähigkeiten übertreffen können.

In den ersten Jahrzehnten vollzog sich diese Entwicklung ausschließlich längs des ersten der beiden im letzten Abschnitt beschriebenen Ansätze, nämlich der Programmierung von Hand unter Zugrundelegung von Logikkalkülen, die die Logiker aus der Beobachtung des menschlichen Schließens im Verlauf von mehr als zwei Jahrtausenden entwickelt haben. Bereits Ende der 1980er Jahre haben Mitglieder aus der Forschungsgruppe des Autors international wohl erstmals Versuche in Richtung eines Einsatzes von NNs mittels darauf basierender Lernverfahren zur Verbesserung der Leistungsfähigkeit automatischer Beweiser unternommen [ESS89]. Der damit begründete Forschungszweig einer synergetischen Integration von ATP und ML hat in den drei Jahrzehnten seither eine wiederum erfolgreiche Entwicklung hin zu immer besseren Systemen erfahren, die sich in Arbeiten

wie beispielsweise [PU20] widerspiegelt, um eine der neuesten Arbeiten dazu aus der inzwischen vorhandenen Fülle herauszugreifen.

In dieser Arbeit wird auf der Grundlage von 13.822 verfügbaren Beweisen mathematischer Theoreme ein rekurrentes neuronales Netz (RNN) daraufhin trainiert, die bestmögliche Wahl unter vorhandenen Alternativen bei jedem Beweissuchschritt zu treffen. Nach dem Training, sprich der Lernphase, zeigt das System eine signifikant bessere Performanz bei der Beweissuche auch für Theoreme, die nicht in der Trainingsmenge enthalten waren.

Experimente und Veröffentlichungen dieser Art gibt es inzwischen unzählige, von denen einige sich beispielsweise in den Referenzen von [BO19, FKU20] finden. Viele dieser Experimente präferieren als Grundlage Beweisverfahren, die auf der Konnektionsmethode (KM) basieren, während noch immer die Mehrzahl der international verfügbaren Beweissysteme die Resolutionsmethode (RM) verwenden. Zum Verständnis dieser beiden differierenden Präferenzen muß man sich die wesentlichen Unterschiede dieser beiden Beweismethoden vor Augen halten.

Die RM ist ein exhaustives Verfahren, das bei der Suche nach einem Beweis für ein gegebenes Theorem im Prinzip in erschöpfender Weise alle denkbaren, für den Beweis potenziell nützlichen Lemmata in Form von Resolventen explizit generiert. Die Generierung von vielen Millionen solcher Lemmata bei der Suche nach dem Beweis eines einzigen Theorems sind dabei eher die Regel als die Ausnahme, was angesichts heutiger Rechengeschwindigkeiten von den meisten ATP-Spezialisten als hinnehmbar eingeschätzt wird. Der Vorteil, der sich mit diesem Vorgehen ergibt, liegt darin, daß im Erfolgsfall dann infolge des in der Flut von Lemmata schließlich gefundenen Einsatzes eines oder mehrerer möglichst guter Lemmata auch relativ kurze Beweise resultieren. Denn ein gutes Lemma kann im Beweis beliebig oft angewandt werden, ohne daß es jedesmal neu bewiesen werden muß, was genau dadurch zu kürzeren Beweisen führt.

Die Beweissuche nach der KM findet auf einem um Größenordnungen kompakteren Suchraum statt, der allein vom Umfang der zu beweisenden Formel und ihrer inneren Struktur bestimmt ist [BO20]. In ihrer ursprünglichen Form und ohne besondere Vorkehrungen weist diese Methode jedoch keinerlei zusätzliche Lemmagenerierung wie in der RM auf, was sich in manchen Fällen sehr nachteilig auf die Länge der resultierenden Beweise und damit natürlich auch auf den Aufwand bei der Beweissuche auswirkt, sodaß aus genau diesem Grund Resolutionsbeweiser bei entsprechend vorhandener Rechenkapazität und in entsprechenden Fällen von Theoremen Konnektionsbeweisern überlegen sind, obwohl ihre Suche aus den soeben genannten Gründen grundsätzlich wesentlich aufwändiger ist.

Die Identifikation von Lemmata ist aber auch in der KM ohne Weiteres leicht realisierbar.[5] Wie zudem bereits in [BE93] mit dem dort dargestellten Konnektionsstrukturkalkül (KSK) gezeigt wurde, läßt sich der beschriebene Nachteil der KM in Bezug auf die Beweislängen

[5] Im Gegensatz dazu ist es für die KM genauso wie für die RM grundsätzlich eine schwierige Aufgabe, unter den unzähligen denkbaren Lemmata die für die Beweissuche wirklich nützlichen aufzuspüren.

sofort beheben, wenn Lemmata in geeigneter Form Berücksichtigung finden. Dies ist in bestehenden Beweisern bislang nur ansatzweise realisiert worden. In [WB20] wird daher ein dem KSK vergleichbarer, jedoch pragmatischer Einsatz von Lemmata bei der Beweissuche beschrieben, wobei die Vorteile der KM nicht geschmälert und die Nachteile gegenüber der RM eliminiert werden.

Im Kontext der KM besteht ein Lemma aus einem Teil der zu beweisenden Formel samt einer Menge zugehöriger Konnektionen und Substitutionen, wobei zu den Details auf [WB20] verwiesen sei. Damit stehen wir auch hier vor der Aufgabe, eine möglichst gute Auswahl von Lemmata in die Beweissuche mit einzubeziehen, wobei die genannten Substitutionen als Träger von Qualitätsmerkmalen vermutet werden. Der Vorteil gegenüber der RM besteht hier nun darin, daß dieser Auswahlprozeß unabhängig von der Beweissuche – und nicht innerhalb der Beweissuche wie bei der RM – separat durchgeführt werden kann. Dies ermöglicht dann auch für diesen Auswahlprozeß den Einsatz von entsprechend konfigurierten Lernmechanismen auf der Basis von NNs. Da es sich bei diesem Auswahlprozeß um etwas spezifisch anderes als bei dem Beweissuchprozeß handelt, können die beiden Lernprozesse auch zielgenauer und mit insgesamt besseren Ergebnissen optimiert werden. Der experimentelle Nachweis dieser Einschätzung steht aber bislang noch aus.

Das hier beschriebene Vorgehen kommt dem menschlichen Beweisen jedenfalls viel näher als das der RM. Denn auch Mathematiker wählen Lemmata aufgrund ihrer Erfahrungen und einem Bauchgefühl aus und suchen dann unabhängig davon den Beweis zu erbringen, wobei dann auch die Beweissuche von ihren einschlägigen Erfahrungen im Beweisen unterstützt wird. Es ist genau dieser Aspekt des Einbringens von Erfahrungen und von Bauchgefühl, der mit NNs modelliert wird, wodurch die Iteration der Verfeinerungen einer Programmierung von Hand automatisiert und dem NN-Mechanismus übertragen wird. Die bisher damit erzielten experimentellen Erfolge zeigen, daß diese ersten Schritte einer synergetischen Integration von zwei bislang unabhängig betriebenen Teilgebiete der KI, nämlich ATP und ML, zu einer weiteren Leistungssteigerung der resultierenden Systeme führt. Wie diese Darstellung zeigt, schlummert hier noch ein erhebliches und bislang nur ansatzweise ausgeschöpftes Potenzial im Hinblick auf eine wesentlich verbesserte Performanz der ATP-Systeme.

4 Quo vadis, KI

Unsere Welt ist aus menschlicher Sicht in jeder Hinsicht extrem komplex. Ihr wissenschaftlich erzieltes Verständnis wird dem Menschen daher auch weiterhin nur teilweise möglich sein. Dies gilt für alle beobachtbaren Phänomene, seien es physikalische, biologische oder geistige. Die Letztgenannten mögen dabei das Bündel von Phänomenen repräsentieren, die wir im zweiten Abschnitt als IPsI-Phänomene apostrophiert haben und um die es in der KI in erster Linie geht. In Bezug auf ihr Verständnis stehen wir in einem gewissen Sinn noch immer erst am Anfang und im Hinblick auf weitere und tiefere Einblicke vor wahrhaft großen Herausforderungen.

Verständlicherweise sind Wissenschaftler wie alle Menschen stolz auf ihre Leistungen. Zudem ist die Erbringung geistiger Leistungen durch Maschinen auch nach Jahrzehnten von KI-Forschung noch immer ein Ereignis, das Aufsehen erregt und uns in besonderer Weise berührt bzw. manche unter uns auch beunruhigt. Es ist daher nicht verwunderlich, daß die mittels DL erbrachten technischen Leistungen, von denen im Abschnitt 2 die Rede war, ein derartiges Aufsehen und Interesse in der Öffentlichkeit erregen, wie wir es in den letzten Jahren erlebt haben. Täglich von Werbung berieselt und beispielsweise auf ein etwas verbessertes Waschmittel gleich mit „das neue Persil“ hingewiesen ist es daher nicht unverständlich, daß im Zusammenhang mit DL auch von der „neuen KI“ die Rede ist. Wie im Abschnitt 2 dargelegt, kann sachlich von einer neuen Wissenschaft KI im Zusammenhang mit DL aber definitiv keine Rede sein. DL hat diese Wissenschaft lediglich ein kleines Schrittchen weiter vorangebracht.[6]

Gleichwohl stellen sich im Zusammenhang mit NNs durchaus eine Reihe von interessanten Fragen, von denen wir einige hier diskutieren wollen. NNs modellieren ja die Arbeitsweise von Neuronen in Gehirnen. Wie bei jeder Modellierung stellt sich daher auch hier die grundsätzliche Frage, wie nahe diese Modellierung dem Original wirklich kommt? Die Ergebnisse der Hirnforschung legen tatsächlich den Schluß nahe, daß NNs nur eine relativ grobe Annäherung an die tatsächlichen Prozesse im Gehirn darstellen.

In der Euphorie über die Erfolge von DL haben sich KI-Forscher, wie dem Vernehmen nach beispielsweise Andrew Ng von der Stanford University, zu der Meinung verstiegen, NNs seien alles, was man zur Erreichung einer allgemeinen KI letztlich benötigen würde. Nun ist keiner von uns in der Lage, die Zukunft vorauszusehen. Angesichts der vorherigen Frage und ihrer Antwort darauf, erscheint eine derartige Prognose aber ziemlich abwegig. Die heutige Wissenschaft KI umfaßt einen großen Strauß von unterschiedlichsten Blumen, sprich Teilgebieten oder Technologien, von denen DL nur eine unter vielen ist. Aufgrund meiner Erfahrung von mehr als einem halben Jahrhundert in KI kann ich nur wärmstens empfehlen, die Blüte all dieser Blumen zu pflegen und sich nicht nur auf die einer einzigen zu beschränken. In genau diesem Sinne sollte der Titel dieser Arbeit verstanden werden. Nur so werden wir nach meiner Überzeugung dem Verständnis und der Realisierung allgemeiner Intelligenz näherkommen.

Grundsätzlich ist es nicht völlig abwegig daraufhin zu spekulieren, daß NNs einmal so weiterentwickelt sein werden, daß sie dann die Prozesse im Gehirn quasi originalgetreu nachbilden. Dann wäre es grundsätzlich denkbar, daß alle geistigen Fähigkeiten, beispielsweise auch die von Mathematikern, allein durch Lernprozesse von NNs erreicht werden könnten. Beim gegenwärtigen Stand der KI ist ein derartiger zukünftiger Entwicklungsstand jedoch völlig unabsehbar. Umso mehr macht es daher beispielsweise Sinn, auf eine Kombination von logischem Schließen im Sinne von ATP und Lernverfahren auf der Grundlage von NNs zu setzen, wie wir es im letzten Abschnitt illustriert haben. Denn auch in der menschlichen Evolution haben die Mathematik und die Naturwissenschaften erst *nach* einer langen Ent-

[6] In den Medien wird, beiläufig erwähnt, gelegentlich auch von einer „neuen KI“ in dem sprachlich salopp formulierten Sinne einer neuen, mit Methoden der KI entwickelten Software gesprochen.

wicklungszeit der Fähigkeit zum logischen Denken in den letzten Jahrtausenden überhaupt entstehen können. Um ein weiteres Beispiel einer derartigen Kombination anzuführen, wäre es denkbar, die menschliche Assoziation von sensorischen Eindrücken und den dafür von uns gebrauchten Begriffen mittels NNs nachzubilden. So könnten syntaktisch formulierte Aussagen mit einer Form von Semantik in Systemen ähnlich versehen werden, wie wir Menschen sie erleben. Experimente in dieser Richtung sind mir jedoch bislang nicht bekannt.

Aktuell sind, abgesehen von ethischen Gesichtspunkten, die in diesem Aufsatz außen vor gelassen werden, eine Reihe von weiteren erkennbaren Schwächen der bislang erreichten Technologie von NNs und DL tatsächlich sehr wohl bekannt und Gegenstand intensiver Forschungen. Dazu gehören beispielsweise das Fehlen von Erklärungen im Falle von DL-generierten Entscheidungen, die Fokussierung des Lernprozesse auf thematikfremde Merkmale, die Integration von Weltwissen in den Lernprozeß sowie das Lernen aus wenigen Beispielen, um nur vier wichtige derartige Themen beispielhaft anzuführen, die im Folgenden jeweils kurz erläutert werden.

Wenn ein menschlicher Schachspieler einen geschickten Schachzug macht, so kann er ihn in der Regel auch begründen, jedenfalls bis zu einem gewissen Grad. Ein mit DL trainiertes Schachprogramm ist dazu absolut nicht in der Lage. Bei der Entscheidung über Schachzüge mag dieser Unterschied nicht allzu gravierend sein. In anderen Anwendungen von DL, wie beispielsweise in der Medizin, erweist sich dieser Mangel an Transparenz von DL-trainierten Systemen als höchst problematisch. Sowohl ein Arzt wie auch sein Patient möchte in jedem Fall die Argumente erfahren, warum eine Behandlungsentscheidung so und nicht anders ausfällt. Erklärbare KI (explainable AI, xAI oder auch Explainable Interactive Learning, XIL) ist daher ein hochaktuelles Teilgebiet der KI von weltweitem Interesse (s. zB. [Gu18, HCM20] sowie [Sc20] und die dort genannten Referenzen). Erklärungen bestehen aus deklarativen Beschreibungen von Wissenszusammenhängen, bedürfen daher zur Integration in Systemen der Formalismen, die auf dem Gebiet der Wissensrepräsentation (knowledge representation and reasoning, KRR) entwickelt wurden. KRR läßt sich in gewisser Weise als verallgemeinertes ATP verstehen. KRR – und damit in gewissen Sinne auch ATP – ist in diesem Sinne für DL daher genauso von Bedeutung wie umgekehrt DL für ATP, wovon im Abschnitt 3 die Rede war. Tatsächlich läßt sich KRR als ein ML, DL ebenso wie ATP umfassendes Teilgebiet der KI verstehen. In [Bi18] wird zudem darauf hingewiesen, daß die KI-Technik einer Spezifikationsextraktion aus einem gegebenen Programm im Hinblick auf Erklärungen eine Rolle spielen könnte, was hier nur am Rande zusätzlich erwähnt sei.

In einem Interview mit der Frankfurter Allgemeinen Zeitung (FAZ) illustrierte einer der führenden DL-Forscher, Sepp Hochreiter, jüngst den Mangel einer Integration von Weltwissen in DL-Systemen mit dem Beispiel einer Plastiktüte, die vom Wind in die Bahn eines autonom gesteuerten Autos flattert. Ein menschlicher Fahrer weiß, daß von einer solchen Tüte keine Gefahr ausgehen kann und setzt daher seine Fahrt unbeeindruckt fort. Ein DL-trainiertes Steuerungssystem kann dagegen den entscheidenden Unterschied einer derartigen Plastiktüte und – sagen wir – eines herabfallenden Dachziegels nicht erkennen

und bremst daher die Fahrt in beiden Fällen ab. Das Generierungssystem GPT-3 (generative pretrained transformer) von OpenAI [Gr20], um noch ein zweites Beispiel anzuführen, generiert zwar erstaunlich perfekt empfundene Texte in Englisch oder anderen symbolischen Kommunikationssprachen quasi durch Kombination von Textschnipseldaten. Der resultierende Text entsteht durch Optimierung von sage und schreibe 175 Milliarden Parametern auf der Grundlage von 570 Gigabyte an Trainingstext (entspricht etwa einer Billion Wörter) mittels des zugrundeliegenden NNs. Da es aber eben reine Schnipselkonfiguration leistet und nichts, aber auch gar nichts in irgendeiner Weise „versteht“, führt dies dann gelegentlich zu lächerlichen Inhalten. Auch diese beiden Beispiele demonstrieren, daß beim gegenwärtigen Stand der Kunst DL ohne KRR wohl nicht umfassend erfolgreich sein kann, weil bislang nur KRR Weltwissen zu repräsentieren imstande ist.

Menschen lernen auch aus einem oder wenigen Beispielen. DL dagegen ist immer auf riesige Datenmengen angewiesen, um zu einem einigermaßen erfolgreichen Verhalten zu gelangen, und es ist beim gegenwärtigen Stand der Kunst nicht zu erkennen, wie DL allein die umfassendere menschliche Lernfähigkeit insoweit erreichen könnte. Vielmehr spricht Vieles dafür, daß auch hier nur eine Synergie zwischen KRR und ML zu einem entsprechenden Erfolg führen wird.

Wie diese drei ausgewählten Beispiele von durchaus auch bei DL erkennbaren Schwächen illustrieren, steckt auch hinter DL trotz seiner großartigen Erfolge doch nur eine menschgemachte Technologie mit entsprechenden Grenzen ihrer Leistungsfähigkeit. Bei allem Verständnis für die menschlich verständliche Begeisterung über diese Erfolge stimme ich mit Steven H. Walker, einem früheren DARPA Direktor, in folgender Einschätzung daher voll überein: „*DARPA believes current research and development investments around the world are much too focused on second-wave AI or machine learning*“ [FCL20, p.4]. Dabei sollte man in diesem Zitat auch auf seine Erwähnung der zweiten vom ML geprägten Welle der Fokussierung in der KI achten, deren Anfänge in den USA bereits in die 1970er Jahre zurückreichen, sodaß auch dieser Teil der KI so neu gar nicht ist. Nur in einer Synergie verschiedenster Stränge der KI-Forschung und anderer Forschungsgebiete werden sich erfolgreiche Entwicklungen zum Wohle aller ergeben. In diesem Sinne wollen wir uns daher im Rest dieses Abschnitts wieder ausschließlich auf die KI als Ganzes konzentrieren. Angesichts des unüberschaubaren Umfangs der KI wäre dabei jegliches Eingehen auf irgendwelche technischen Details im Kontext dieser Arbeit völlig willkürlich und damit definitiv unangebracht.

Vielmehr stellt sich mir die grundsätzliche Frage, wie wir in Europa und Deutschland die KI als Wissenschaft und Technologie, auch im internationalen Wettbewerb, besser voranbringen könnten als dies in den letzten 75 Jahren gelungen ist. Die historische Erfahrung lehrt uns dabei, daß für eine gedeihliche Entwicklung eine realistische Vision unabdingbar ist. Denn es waren genau solche Visionen, beispielsweise von Vannevar Bush im Jahr 1945 [Bu45], John McCarthy 1956 [Mc56] und Josef Licklider 1960 [Li60], die in den USA die Leitlinien für die Entwicklung der KI dort vorgegeben und zum großen Erfolg geführt haben. Dazu gehörte natürlich vor allem auch, daß diese Visionen von den maßgeblichen

Entscheidern auch wahrgenommen und beachtet wurden. In den USA ging man 1962 sogar soweit, den Visionär Licklider das US Information Processing Techniques Office (IPTO) in der Advanced Research Projects Agency (ARPA) gründen zu lassen. Die vergleichbaren deutschen Visionäre Zuse und Steinbuch, von denen im Abschnitt 2 die Rede war, hatten hierzulande dagegen keinerlei Chance zu einer vergleichbaren Realisierung ihrer Visionen erhalten, obwohl sie ihren amerikanischen Kollegen an Durchblick in nichts nachstanden (s. zB. [St61]).

Dabei haben die Visionen aus diesen Anfängen der KI bis zu einem gewissen Grad auch heute noch unverändert Gültigkeit. So schrieb Bush: „*Professionally our methods of transmitting and reviewing the results of research are generations old and by now are totally inadequate for their purpose.*“ Diese damals für die Wissenschaft formulierte Einschätzung hat heute vor allem Gültigkeit für die Methoden, die unsere gesellschaftlichen Entscheidungsprozesse in vielen Bereichen wie beispielsweise in Politik, Verwaltung, Recht, Wirtschaft usw. prägen, die aus KI-Sicht recht antiquiert erscheinen und deren Neukonzeptionen auf der Grundlage wissenschaftlicher Erkenntnisse längst überfällig sind (vgl. dazu beispielsweise die diesbezüglichen Ausführungen in dem Buch [Bi03] sowie speziell zum Recht in [Bi05]). Denn Menschen sind naturgemäß nur sehr eingeschränkt fähig, globalere Gesichtspunkte ihren Entscheidungen zugrundezulegen, sondern sie denken, wie alle Lebewesen, in Bezug auf ihre Handlungsentscheidungen in allererster Linie zeitlich und räumlich lokal und vorwiegend an ihr eigenes Wohlergehen und das ihrer Sippe (man denke beispielsweise nur an das aktuell so oft gehörte „America first“). Auch demokratisch herbeigeführte Entscheidungen sind aus dem gleichen Grund von erratischen Einflüssen geprägt und daher ähnlich bedingt tauglich wie diejenigen von einzelnen Personen. Nur mittels KI könnte man eine objektivere Entscheidungsbasis aufbauen und darauf beruhende Mechanismen entwickeln, die die unterschiedlichen Interessen global, fair und auch zukunftsorientiert berücksichtigen.

Es wäre daher sehr wünschenswert, daß sich die Entwicklung und Förderung der KI an einer Vision dieser Art orientieren würde. Die plan- und visionslose Ausschüttung von Milliarden an Fördergeldern nach dem Gießkannenprinzip oder ähnlich oberflächlichen Prinzipien wird Europa und Deutschland dagegen nicht sehr weit voranbringen, weder in der KI noch als Gesellschaft. Leider ist im Rahmen erstarrter organisatorischer und institutioneller Strukturen kaum zu erhoffen, daß sich an dem bisher geübten Weiterwursteln und reflexartigen, kurzfristigen Reagieren auf Anstöße von außen Grundlegendes ändern wird. Weil man die Hoffnung aber nie aufgeben sollte, seien dazu dennoch einige Hinweise gegeben.

In meiner Jugend wurde oft das Sprüchlein zitiert: „Wenn das Wörtchen *wenn* nicht wär, wär mein Vater ein Millionär.“ Es macht danach wenig Sinn, der Vergangenheit nachzutrauern und vorwurfsvoll in einer Rückschau darüber zu spekulieren, wie die Entwicklung der Informatik und KI wohl gelaufen wäre, *wenn* Zuse nach 1945 und Steinbuch nach 1960 in Positionen berufen worden wären, in denen sie entscheidenden Einfluß auf die Weichenstellungen in der Wissenschafts- und Technologieentwicklung hätten nehmen können.

Im Gegensatz dazu macht es aber eine Menge Sinn, die damaligen organisatorischen und institutionellen Strukturen daraufhin zu untersuchen, warum sie die adäquate Positionierung von weitsichtigen Größen wie Zuse und Steinbuch nicht zwangsläufig befördert hatten. Mögen diese beiden Persönlichkeiten auch besonders ins Auge springen, könnte hier eine erstaunlich lange Liste ähnlicher Fälle erstellt werden. Beispielsweise könnte man auch Andreas von Bechtolsheim erwähnen, der Deutschland frustriert verließ und dann in den USA als Entwickler und Unternehmer höchst erfolgreich reüssierte, oder Ernst Dickmanns, dessen autonome Fahrzeuge schon Mitte der 1980er Jahre der internationalen Konkurrenz weit voraus waren, dessen Technologie von den deutschen Autobauern aber nicht weitergeführt wurde, oder Sebastian Thrun, der leider auch in die USA abgewandert ist und dort schließlich die auf dem autonomen Fahren international führende Firma Waymo mitbegründet hat, oder die vielen deutschen Spitzenkräfte in ATP, die sich nach 2000 mangels verfügbarer Positionen in alle Himmelsrichtungen international zerstreuen mußten, oder der einstige europäische Vorsprung im maschinellen Verstehen und in der Verarbeitung natürlicher Sprache, dessen Vermarktung dann doch in den USA erfolgte, und viele bzw. vieles andere mehr. Die große Anzahl solcher Fälle muß ins Auge stechen und suggeriert, daß hier organisatorische Mängel in den institutionellen Strukturen ebenso wie mentale Denkmuster eine entscheidende Rolle spielen dürften. Eine wissenschaftliche Ergründung dieser Rolle und Konsequenzen aus den sich daraus ergebenden Erkenntnissen wären daher im Hinblick auf eine Revision dieser Strukturen und Denkmuster überfällig.

In der Tat bin ich davon überzeugt, daß alle einschlägigen Institutionen durchgreifender Reformen bedürften: Angefangen von dem föderalen Hochschulwesen, in dem sich wünschenswerte Spitzenuniversitäten wie Stanford oder MIT hierzulande prinzipiell nicht entfalten können, über eine von Gremien dominierte DFG, die zB. bei größeren Entscheidungen mehr auf Gruppenkonsens als auf individuelle und in Eigenverantwortung schriftlich begründete Beurteilungen setzt, bis hin zu den unzähligen Ministerien auf Bundes- und Landesebene, die sich im Wettbewerb um die Förderung von prestigeträchtigen Projekten gegenseitig eifersüchtig beäugen und zu übertreffen suchen. Für das *normale* Gedeihen von Forschung und Entwicklung scheint dieses System durchaus gut zu funktionieren. Zur Hervorbringung von wirklicher Spitzenforschung und echten Pionierleistungen samt deren zügiger Umsetzung in wirtschaftlichen Erfolg ist es aber nur sehr bedingt in der Lage, wie die letzten Jahrzehnte jedermann erkennbar gezeigt haben. Und im weltweiten Wettbewerb scheint nur der zu reüssieren, der genau darin die Nase vorne hat.

Zudem scheint mir in diesem System eine weitsichtig orientierte, steuernde Institution wie das DARPA zu fehlen. Wenn man einmal davon absieht, daß DARPA immer eine Institution des amerikanischen Verteidigungsministeriums gewesen ist, dann besticht seine jahrzehntelange visionsgeprägte und personell hervorragend vertretene Rolle allemal, in der es ganz gezielt Exzellenz in großzügiger Weise förderte und zwar unabhängig von den kurzfristig erreichbaren Zielsetzungen [FCL20]. In geschickter Weise brachte es Wissenschaftler in Grundlagenforschung mit Anwendern zusammen, wodurch der Wissenstransfer ungehindert und ohne Zwischenbarrieren erfolgen konnte. Nirgendwo in

Europa oder Deutschland ist eine vergleichbar erfolgreiche Institution wie das DARPA auszumachen.

Auf einer noch grundlegenderen Ebene sehe ich längerfristig die Notwendigkeit einer Neuordnung des gesamten Wissenschaftssystems. Der Abschnitt 2 hat im Zusammenhang mit der dort beschriebenen IPsI-Wissenschaft schon einige Hinweise hierzu gegeben, nicht zuletzt den Hinweis auf die einschlägige Arbeit [Bi18]. Wie der Prozeß einer solchen Neuordnung planvoll in Gang gesetzt werden könnte, dazu fehlen mir angesichts der festgezurrten Strukturen allerdings die Ideen.

Auch glaube ich zu erkennen, was aber wissenschaftlich genauer untersucht und präziser formuliert werden müßte, daß wir hierzulande in einer Gesellschaft leben, die den Durchschnittsmenschen generell zum Maßstab von Bewertungen präferiert. Geistesgenies genießen in dieser Denke keinen besonderen Status als Vorbilder oder besonders zu fördernde Zeitgenossen, sondern werden bestensfalls wegen ihrer Eigenheiten karikaturiert. Nur Genies, deren geniale Fähigkeiten man mit den Augen oder Ohren bestaunen kann, wie beispielsweise Fußballer oder SängerInnen, werden bewundert und dafür mit Geld überschüttet, auch wenn ihr Tun auf lange Sicht für die Gesellschaft bei weitem nicht so nutzbringend wie das von erfolgreichen Wissenschaftlern sein mag.

5 Konklusion

Die Wissenschaft der Künstlichen Intelligenz (KI), hier in einem umfassenden Sinne als IPsI-Wissenschaft apostrophiert, und die aus ihr hervorgehende Technologie hat in den letzten Jahren aufgrund aufsehenerregender Erfolge international eine große Beachtung und Förderung erfahren dürfen. Da Europa und Deutschland hierbei bislang nur relativ bescheidene Rollen spielen konnten, werden in dem vorliegenden Beitrag einige grundlegende Hinweise daraufhin zusammengestellt, wie diese Situation verbessert werden und Europa zurück in eine Führungsrolle auf dem Gebiet der KI gebracht werden könnte. Die im Abschnitt 4 zusammengestellten Hinweise resultieren dabei teilweise aus einer Sicht auf ausgewählte, im Abschnitt 2 dargestellte Aspekte der Geschichte der KI sowie aus dem im Abschnitt 3 behandelten Beispiel einer Synergie der Gebiete des automatischen Beweisens und des maschinellen Lernens. Aus allem ergibt sich die im Titel zum Ausdruck gebrachte Empfehlung, daß man allen in der KI betriebenen Teilgebieten genügend Raum lassen und sich nicht auf eines dieser beschränken sollte, was besondere Gewichtungen natürlich keineswegs ausschließen soll.

Dank. Ich danke Ralf Bülow für seine Hinweise auf die frühen Zuseschen Arbeiten, Christian Vater für hilfreiche Hinweise zu Steinbuch, Hannes Bibel für seine technische Unterstützung sowie Reinhard Kahle und Klaus Mainzer für die Einladung zu diesem Workshop.

Literaturverzeichnis

[AD20] Alspector, J., Dietterich, T.G.: DARPA's role in machine learning. AI Magazine 41(2), 36–48 (2020)

[An64] Anderson, A.R. (ed.): Minds and Machines. Contemporary Perspectives in Philosophy, Prentice Hall, Englewood Cliffs NJ (1964)

[Bi05] Bibel, L.W.: AI and the conquest of complexity in law. Artificial Intelligence and Law Journal 12, 159–180 (2005)

[BE93] Bibel, W., Eder, E.: Methods and calculi for deduction. In: Gabbay, D.M., Hogger, C.J., Robinson, J.A. (eds.) Handbook of Logic in Artificial Intelligence and Logic Programming, vol. 1, chap. 3, pp. 71–193. Oxford University Press, Oxford (1993)

[Bi03] Bibel, W.: Lehren vom Leben – Essays über Mensch und Gesellschaft. Sozialwissenschaft, Deutscher Universitäts-Verlag, Wiesbaden (2003)

[Bi14] Bibel, W.: Artificial Intelligence in a historical perspective. AI Communications 27(1), 87–102 (2014)

[Bi18] Bibel, W.: On a scientific discipline (once) named AI. In: Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence (IJCAI 2018). pp. 5143–5149. IJCAII (2018), https://www.ijcai.org/proceedings/2018/0713.pdf

[Bi20a] Bibel, W.: On the development of AI in Germany. Künstliche Intelligenz 34(2), 251–258 (2020), open Access https://rdcu.be/b3oxS

[Bi20b] Bibel, W.: Der unterschätzte Konrad Zuse. Frankfurter Allgemeine Zeitung (226), 18 (28 September 2020), https://www.faz.net/aktuell/wirtschaft/digitec/der-unterschaetzte-konrad-zuse-16974194.html

[BO19] Bibel, W., Otten, J.: Experiments with connection method provers. In: 4th Conference on Artificial Intelligence and Theorem Proving, AITP 2019, April 7-12, 2019, Obergurgl, Austria (2019), http://aitp-conference.org/2019/slides/WB.pdf

[BO20] Bibel, W., Otten, J.: From Schütte's formal systems to modern automated deduction. In: Kahle, R., Rathjen, M. (eds.) The Legacy of Kurt Schütte, chap. 13, pp. 215–249. Springer, Cham (2020)

[Bü20] Bülow, R.: Konrad Zuse und der Plankalkül. Heinz Nixdorf Museums Forum, HNF Blog Neues von gestern aus der Computergeschichte (22 Juni 2020), https://blog.hnf.de/konrad-zuse-und-der-plankalkuel/

[Bu45] Bush, V.: As we may think. Atlantic Monthly 176(7), 101–108 (1945)

[Da57] Davis, M.: A computer program for Presburger's algorithm. In: Summaries of talks presented at the Summer Institute for Symbolic Logic. pp. 215–233. Institute for Defense Analysis, Princeton NJ (1957), also contained in [SW83, 41–48].

[DS04] Dorigo, M., Stützle, T.: Ant Colony Optimization. MIT Press, Boston MA (2004)

[ESS89] Ertel, W., Schumann, J.M., Suttner, C.B.: Learning heuristics for a theorem prover using back propagation. In: Retti, J., Leidlmair, K. (eds.) 5. Österreichische Artificial-Intelligence-Tagung: Igls/Tirol, 28.-31. März. pp. 87–95. Springer, London (1989)

[FKU20] Färber, M., Kaliszyk, C., Urban, J.: Machine learning guidance for connection tableaux. Journal of Automated Reasoning (2020), https://doi.org/10.1007/s10817-020-09576-7

[FF63] Feigenbaum, E.A., Feldman, J. (eds.): Computers and Thought. McGraw-Hill, New York NY (1963)

[FCL20] Fouse, S., Cross, S., Lapin, Z.J.: DARPA's impact on Artificial Intelligence. AI Magazine 41(2), 3–8 (2020)

[Gr20] Graf, A.: Multitalent für Sprache. Spektrum.de (11 August 2020), https://www.spektrum.de/news/kuenstliche-intelligenz-der-textgenerator-gpt-3-als-sprachtalent/1756796 (Zugriff 17.8.20)

[Gu18] Guidotti, R., et al.: A survey of methods for explaining black box models. ACM Computer Surveys 51, 1–42 (2018)

[HCM20] Holzinger, A., Carrington, A., Müller, H.: Measuring the quality of explanations: The system cusability scale. KI - Künstliche Intelligenz 34(2), 193–198 (2020)

[HFB20] Hunt, E.R., Franks, N.R., Baddeley, R.J.: The bayesian superorganism: externalized memories facilitate distributed sampling. Journal of the Royal Society Interface 17 (2020), 20190848

[KIT06] KIT-Archiv: 27048 Nachlass Karl Steinbuch. Findbuch (2006/2007), http://xputer.de/steinbuch/Steinbuchs-Nachlass.htm (Zugriff 13.8.2020)

[KSH12] Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Proceedings of the 25th International Conference on Neural Information Processing Systems (NIPS 2012. vol. 1, pp. 1097–1105. Association of Computing Machinery (ACM), New York NY (2012)

[Le98] Lecun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. In: Proceedings of the Institute of Electrical and Electronics Engineering (IEEE). vol. 86, pp. 2278–2324. IEEE (1998), doi.org/10.1109/5.726791

[Li60] Licklider, J.C.R.: Man-computer symbiosis. IRE Transactions on Human Factors in Electronics HFE-1(1), 4–11 (1960)

[Mc56] McCarthy, J.: Artificial intelligence (1956), a project proposal

[Mc55] McCarthy, J., Minsky, M.L., Rochester, N., Shannon, C.E.: A proposal for the Dartmouth summer research project on artificial intelligence (August 31 1955)

[NSS56] Newell, A., Shaw, J., Simon, H.: The logic theory machine. In: IRE Trans. Information Theory. vol. IT-2, pp. 61–79 (1956), also contained in [SW83, 49–73].

[Ni09] Nilsson, N.J.: The Quest for Artificial Intelligence: A History of Ideas and Achievements. Cambridge University Press (2009)

[PU20] Piotrowski, B., Urban, J.: Guiding inferences in connection tableau by recurrent neural networks. In: Benzmüller, C., Miller, B. (eds.) CICM 2020: Intelligent Computer Mathematics, Proceedings of the International Conference on Intelligent Computer Mathematics. LNCS, vol. 12236, pp. 309–314. Springer (2020), https://doi.org/10.1007/978-3-030-53518-6_23

[Ro58] Rosenblatt, F.: The perceptron. a probabilistic model for information storage and organization in the brain. Psychological Reviews 65, 386–408 (1958)

[Sc20] Schramowski, P., Stammer, W., et al.: Making deep neural networks right for the right scientific reasons by interacting with their explanations. Nature Machine Intelligence 2, 476–486 (August 2020)

[Sh50] Shannon, C.E.: Programming a digital computer for playing chess. Philosophy Magazine 41, 356–375 (1950)

[SW83] Siekmann, J., Wrightson, G. (eds.): Automation of Reasoning – Classical Papers on Computational Logic 1957–1966, vol. 1. Springer, Berlin (1983)

[St61] Steinbuch, K.: Automat und Mensch – Über menschliche und maschinelle Intelligenz. Springer (1961)

[St60] Steinbuch, K.: Elekrischer Zuordner mit Lerncharakter. Deutsches Patent 179409 (Anmeldetag 23 Sept 1960)

[Tu50] Turing, A.M.: Computing machinery and intelligence. Mind 59, 433–460 (1950), also in [FF63, 11–35] and [An64, 4–30].

[WB20] Wernhard, C., Bibel, W.: Comparative studies of complex proofs, (2020 in preparation)

[Wi48] Wiener, N.: Cybernetics or Control and Communication in the Animal and the Machine. MIT Press, Cambridge MA (1948)

[Zu48] Zuse, K.: Über den Plankalkül als Mittel zur Formulierung schematisch kombinativer Aufgaben. Archiv der Mathematik 1(6), 441–449 (1948)

[Zu49] Zuse, K.: Die mathematischen Voraussetzungen für die Entwicklung logistisch-kombinativer Rechenmaschinen. Zeitschrift für angewandte Mathematik und Mechanik 29(1/2), 36–37 (1949)

[Zu50b] Zuse, K.: Kombinierte numerische und nichtnumerische Rechenmaschine. Patentschrift Nr.926449, Deutsches Patentamt (13 Mai 1950), 11 Seiten

[Zu50a] Zuse, K.: Schlüssel-Programmwerk. Patentschrift Nr.977282, Deutsches Patentamt (25 April 1950), 9 Seiten

Original oder Plagiat? – Der schnelle Weg zur wissenschaftlichen Arbeit im Zeitalter künstlicher Intelligenz(en)

Original oder Plagiat?

Der schnelle Weg zur wissenschaftlichen Arbeit im Zeitalter künstlicher Intelligenz

Doris Weßels[1] Eike Meyer [2]

Abstract: Hochschulen müssen trotz der immer wieder artikulierten Ressourcen- und Kapazitätsprobleme zukunftsfähige Leitlinien und Praktiken für den Umgang mit studentischen Leistungen in Form schriftlicher Haus- und Abschlussarbeiten entwickeln – bis hin zur Entwicklung alternativer Konzepte als Bestandteil neuer Lernarchitekturen. Für die Hochschulleitungen und Lehrenden ergeben sich zwei Fragestellungen. Zum einen: Welchen Impact haben KI-basierte Werkzeuge des „Natural Language Generation bzw. Processing" (NLG/NLP) für Prüfungsleistungen in Form schriftlicher Haus- und Abschlussarbeiten? Zum zweiten: Wie ist das „System Hochschule" anzupassen, um seinem Bildungs- und Qualitätsanspruch im digitalen Zeitalter gerecht zu werden? In einem Online-Workshop im Rahmen der INFORMATIK2020 am 1.10.2020 wurden ausgewählte Werkzeuge in einem spielerisch anmutenden Team-Wettbewerb eingesetzt, um auf dieser Grundlage den schwierigen Grat zwischen Original und Plagiat bei der Erstellung wissenschaftlicher Arbeiten selbst zu erleben, das Verhalten als Lehrender und Forschender kritisch zu reflektieren und das Problembewusstsein für diese neue Herausforderung zu schärfen.

Keywords: Künstliche Intelligenz; Plagiate; Natural Language Generation; Natural Language Processing; GPT-3; Plagiatsprüfung

1 Einleitung

Plagiate und der Einsatz von Ghostwritern dürfen heute als eine weit verbreitete Problematik in der Hochschulwelt bewertet werden, ohne das Ausmaß wegen fehlender Transparenz quantitativ bewerten zu können. Der Umgang mit nicht eigenständig erstellten schriftlichen Arbeiten von Studierenden stellte Lehrende bereits in der Vergangenheit vor große Herausforderungen. Nun zeigt sich aber eine noch größere Herausforderung durch das wachsende Angebot leistungsstarker KI-basierter Werkzeuge für die automatisierte Überarbeitung („Rewriting) vorhandener und Generierung neuer Texte [PWW20]. Bei dem hier vorzustellenden INFORMATIK2020-Workshop hatten die Teilnehmenden nach einem einführenden Impulsvortrag und einer Vorstellung ausgewählter Tools die Möglichkeit,

[1] Fachhochschule Kiel, Fachbereich Wirtschaft, Sokratesplatz 2, 24149 Kiel, doris.wessels@fh-kiel.de
[2] Fachhochschule Kiel, Lehrbeauftragter, Sokratesplatz 2, 24149 Kiel, eike.meyer@fh-kiel.de, https://orcid.org/0000-0003-4381-1615

doi:10.18420/inf2020_66

selbst Erfahrungen mit diesen frei verfügbaren Werkzeugen zu machen. Anschließend wurden im Expertenkreis die Potenziale und Herausforderungen für die Hochschul- und Forschungslandschaft diskutiert. Der Ablauf des Workshops inklusive der Ergebnisse der Arbeitsphase und der anschließenden Diskussion sollen im Folgenden vorgestellt und erläutert werden

2 Standortbestimmung und KI-Implikationen - Impulsvortrag

2.1 Plagiate gestern und heute

Für das Selbstverständnis von Hochschulen sind die von Studierenden zu erstellenden schriftlichen Haus-, Studien- oder auch Abschlussarbeiten ein zentrales Ergebnis ihres Bildungsauftrags, tief verankert in einer jahrhundertealten Universitätskultur. Dahinter verbirgt sich das Ziel der Schreibkompetenz-Förderung der Lernenden – im Sinne einer akademischen Basisdisziplin. Im Jahr 2018 hat die Gesellschaft für Schreibdidaktik und Schreibforschung in ihrem Positionspapier dieses Ziel präzise formuliert:

„Schreibkompetenz im Studium ist die Fähigkeit, Texte zum Lernen und als Anknüpfungspunkte für eigene Textproduktionen zu nutzen und sich schriftsprachlich angemessen auszudrücken. Diese Fähigkeit setzt sich aus fachübergreifenden und fachspezifischen Komponenten zusammen und kann in drei Dimensionen beschrieben werden:

- Kompetente Schreibende nutzen Schreiben zum kritischen Denken,
- steuern produktiv den eigenen Schreibprozess und
- kommunizieren entsprechend den Textkonventionen der jeweiligen Fachgemeinschaft angemessen“ [Ge18].

Die Schlussfolgerung lautet somit, dass der erfolgreiche und damit nicht identifizierte Einsatz von im Hintergrund agierenden Ghostwritern und die Verwendung von Plagiaten mit allen Mitteln unterbunden werden sollten.

Bei der Diskussion der Ghostwriter-Problematik wird häufig argumentiert, dass es sich um ein Nischenproblem und damit selten auftretende Einzelfälle handelt. Neben der Fülle (mehr oder weniger) professioneller Agenturen [Ha19] finden sich die Anbieter aber bereits in Facebook-Gruppen von Studierenden oder sie bieten ihre Dienste in der Region über Ebay-Kleinanzeigen an, siehe den Screenshot in Abbildung 1.

Leider offenbart auch der Umgang mit klassischen Plagiaten ein defizitäres Verhalten an vielen Hochschulen. Bei der von Doris Weßels im Februar 2020 durchgeführten „Spotlight“-Umfrage unter 25 Hochschullehrenden von mehr als 20 Hochschulen aus Deutschland, Schweiz und Österreich zeigte sich, dass nur 15 der 25 befragten Hochschullehrenden

Abb. 1: Drei beispielhafte Angebote bei eBay-Kleinanzeigen für den Suchbegriff “Ghostwriter” in Schleswig-Holstein (`https://www.ebay-kleinanzeigen.de/s-schleswig-holstein/ghostwriter/k0l1408`, Aufruf: 10.10.2020)

(entspricht einer Quote von 60 %) der Befragten eine Plagiatserkennungs-Software nutzen (siehe Abbildung 2). Dieser Quote stehen 40 % der Lehrenden gegenüber, die ohne technische Unterstützung arbeiten und sich auf ihre Erfahrung und Intuition bei der Entdeckung von Plagiaten verlassen müssen oder möchten.

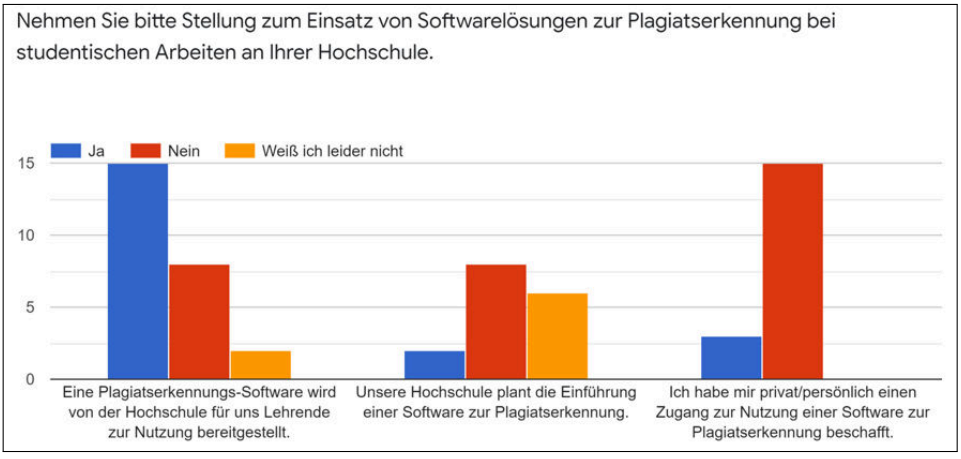


Abb. 2: “Spotlight”-Umfrage zum Einsatz von Plagiatserkennungs-Softwarelösungen bei 25 Hochschullehrenden aus Deutschland, Schweiz und Österreich (“Spotlight”-Umfrage Weßels 2020)

Bei der Frage nach den bereits entdeckten Plagiaten erhärtet sich der Verdacht, dass es eine Korrelation zwischen der Verwendung von Plagiatserkennungs-Software und dem Entdecken

von Plagiaten gibt, siehe hierzu die Quote von 40 % Hochschullehrende ohne Auftreten von Plagiaten und 60 % mit teilweise sogar gehäuften Plagiatsfällen ihrer Studierenden in Abbildung 3.

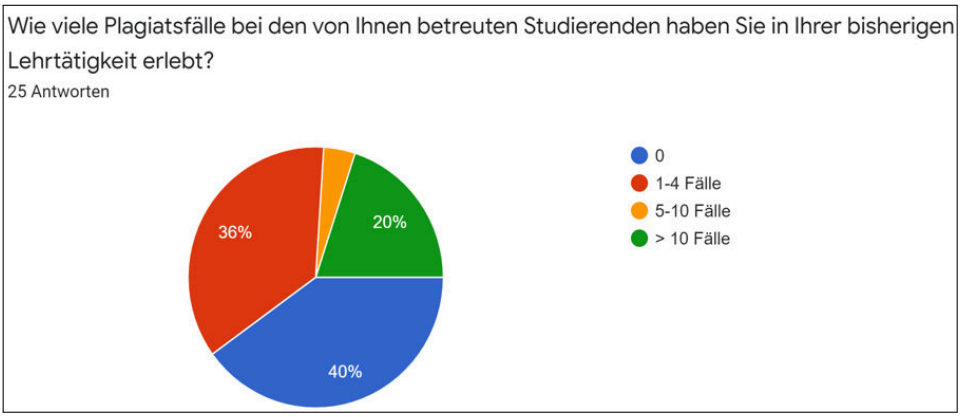


Abb. 3: Häufigkeit von Plagiaten in der eigenen Praxis der Lehrenden (“Spotlight”-Umfrage Weßels 2020)

Die Untersuchung der Gründe für dieses Verhalten gestaltet sich schwierig. Die “Stress”-Vermeidung bei der Entdeckung von Plagiaten darf als ein relevanter Einflussfaktor bewertet werden, siehe Abbildung 4.

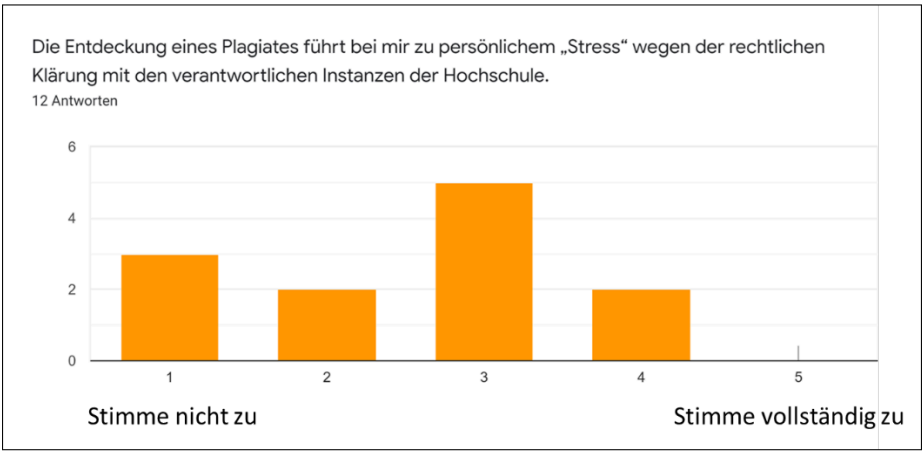


Abb. 4: Der „Stress“-Faktor bei der Entdeckung von Plagiaten für Lehrende (“Spotlight”-Umfrage Weßels 2020)

Außerdem stellt der Einsatz derartiger Software-Lösungen für die Hochschulen eine IT-Investition dar und für die Lehrenden einen Arbeitsaufwand in der täglichen Nutzung. Über die Sinnhaftigkeit wird dann kontrovers diskutiert, wenn Lehrende argumentieren und (vermeintlich?) glauben, dass sie Plagiate mit hoher Wahrscheinlichkeit ohnehin erkennen

würden und somit eine Plagiatserkennungs-Software überflüssig sei [MW20]. Erschwerend für den Einsatz von Plagiatserkennungs-Software kommt hinzu, dass die Leistungsfähigkeit der heutigen Softwarelösungen begrenzt ist bzw. deutliche Schwächen aufweist [Gr20, FD20].

Trotz dieser Vorbehalte bewertet der Großteil der Lehrenden in der „Spotlight"-Umfrage die abschreckende Wirkung der Plagiatserkennungs-Softwarelösungen bei Studierenden als sehr hoch, siehe Abbildung 5.

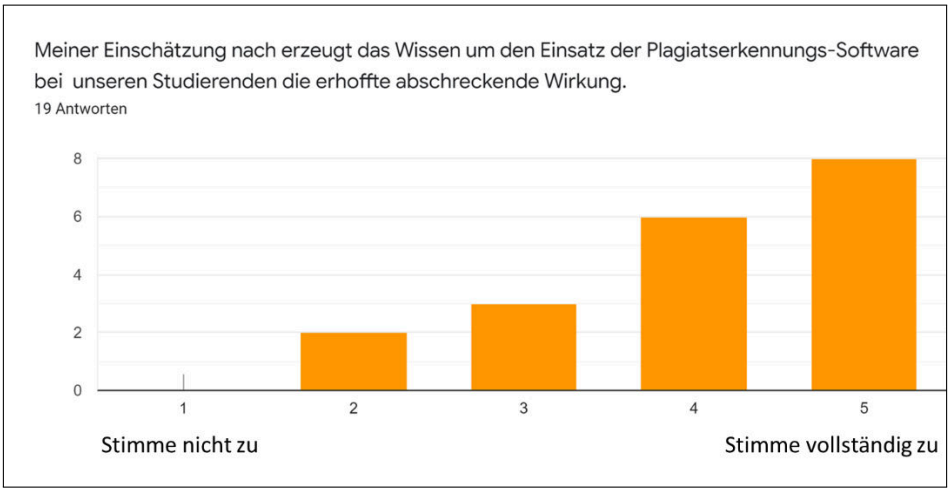


Abb. 5: Die abschreckende Wirkung von Plagiatserkennungs-Softwarelösungen bei Studierenden ("Spotlight"-Umfrage Weßels 2020)

Im Zeitalter der künstlichen Intelligenz muss der Begriff des „Plagiats" jedoch neu interpretiert oder durch einen alternativen Begriff ersetzt werden, wenn er als Synonym für die nicht eigenständig erstellte Leistung (mit Täuschungsabsicht) steht. Studierende versichern derzeit an deutschen Hochschulen mit einer eidesstattlichen Erklärung (häufig auch Selbstständigkeitserklärung oder Eigenständigkeitserklärung genannt) bei der Abgabe ihrer schriftlichen Abschlussarbeit, dass sie die wissenschaftliche Arbeit selbstständig und ohne fremde Hilfe erstellt haben.

Was verbirgt sich im juristischen Sinne hinter dem Terminus der „fremden Hilfe"? Ist damit die Unterstützung durch menschliche Akteure gemeint? Oder umfasst sie auch die Unterstützung durch „künstliche Intelligenzen"? Wenn ja, bis zu welchem Grad ist diese Form der Unterstützung erlaubt?

Bei Diskussionen zu dieser neuen und sehr komplexen Fragestellung wird häufig auf bereits heute bestehende „Grauzonen" verwiesen, die auch ohne Nutzung von KI-Technologien bei Peer-Reviews in der akademischen Praxis oder bei der Nutzung von externen Lektoratsanbietern durch Studierende entstehen. Es muss aber nicht nur dieser professionelle

Support sein, der Fragen zum Grad der Eigenständigkeit der schriftlichen Arbeit aufwirft. Gerade Studierende aus Akademikerfamilien genießen häufig einen intensiven Schreib- und Lektoratsservice durch die Familienmitglieder, ohne dass diese Unterstützungsfunktion bei der Abgabe der schriftlichen Arbeit angegeben wird. Selbstverständlich wirkt in der Regel auch der Freundes- und Kommilitonenkreis mit oder insbesondere bei dualen Studierenden das Umfeld der Kolleginnen und Kollegen. Nicht nur die Mitwirkung dieses Personenkreises, sondern auch die Nutzung von Suchmaschinen und gängiger Textverarbeitungslösungen wird gerne als Argument für die gesellschaftlich akzeptierte „Supportfunktion" angeführt, die nicht die eidesstattliche Erklärung in Frage stellt und daher auch nicht strafrechtliche Folgen befürchten lässt.

Die zuvor skizzierte Komplexität bei der Bewertung der Eigenständigkeit von schriftlichen Arbeiten, die im häuslichen Umfeld der Studierenden entstehen, erhöht sich signifikant durch den Einsatz von KI-gestützten Werkzeugen (NLP bzw. NLG). Die Unterscheidung und Bewertung von Original und Plagiat in Verbindung mit der Autorenschaft und dem Urheberrecht stellen eine neue Governance- sowie Compliance-Herausforderung mit vielfältigen rechtlichen und ethischen Fragestellungen für die Hochschulen und letztlich unsere Gesellschaft dar.

2.2 „Plagiate morgen" - Produktion von Texten mit leistungsstarken KI-gestützten Werkzeugen

Die hohe Verfügbarkeit und wachsende Leistungsstärke KI-gestützter Werkzeuge und computerlinguistischer Algorithmen spiegelt bereits heute eindrucksvoll wider, dass die Generierung wie auch die Überarbeitung von Texten ungeahnte Möglichkeiten bieten und zu einem neuartigen „Kontinuum" zwischen Original und Plagiat für wissenschaftliche Arbeiten führen [We20]. Im Rahmen des Workshops wurden größtenteils Werkzeuge eingesetzt, die auf den beiden Sprachmodellen GPT-2 und GPT-3 basierten.

Der Algorithmus GPT-2 wurde 2019 von der Non-Profit-Forschungsorganisation OpenAI aus Kalifornien veröffentlicht. Er ist darauf trainiert, das nächste Wort vorherzusagen, wenn zuvor alle vorherigen Wörter in einem Text berücksichtigt wurden [Op19]. Die Besonderheit von GPT-2 ist, dass es sich an den Stil des zuvor erfassten Inputs, der in der Regel von einem Menschen formuliert wird, anpasst und dann schrittweise eine „Geschichte" eigenständig wie ein Mensch weiterformulieren kann [Op19].

Im März 2020 wurde bereits der Nachfolger GPT-3 veröffentlicht, der seit Herbst auch kommerziell vertrieben wird. Die exklusiven Lizenzrechte wurden von der Firma Microsoft erworben, welche den Algorithmus vor allem über die hauseigene Cloud anbietet [Sc20]. Das GPT-3-Sprachmodell bietet eine deutliche Leistungssteigerung mit einem sehr breit gefächerten Einsatzspektrum. Das Sprachmodell arbeitet mit 175 Milliarden Parametern, d.h. der zehnfachen Menge im Vergleich zu den bisherigen KI-Sprachmodellen [Br20]. Erste Anwendungsbeispiele zeigen, dass neben der automatischen Generierung von Texten [Mo20]

auch lauffähiger Software-Code generiert wird, indem der Anwender sprachgesteuert seine Anforderungen an die Software formuliert [Be20].

In der nachfolgenden Tabelle 1 werden die Werkzeuge dargestellt, die im Rahmen des Team-Wettbewerbs im Workshop vorgestellt und zur Nutzung empfohlen wurden.

Tab. 1: Übersicht der Werkzeuge

Übersetzungen mit DeepL https://www.deepl.com/translator - auch durch die Verwendung von Synonymen und Glossar vielfältig für das Rewriting einsetzbar
Generierung von Text (NLG) und anderen Objekten:
https://transformer.huggingface.co/doc/distil-gpt2
https://transformer.huggingface.co
https://app.inferkit.com/generate
https://bellard.org/textsynth/
https://philosopherai.com/ - GPT-3-basierte Textgenerierung, zur Illustration der Leistungsstärke die Beispielantwort: https://philosopherai.com/philosopher/which-techniques-methods-and-procedures-are-suita-9ee429
https://thispersondoesnotexist.com/ - für fiktive Personen
https://www.thisworddoesnotexist.com/ - neue fiktive Begriffe inklusive Definition mit Spaßfaktor!
Textbasiertes GPT-3-Dungeon-Spiel: https://play.aidungeon.io/main/home, Einstellung Single Player und „Custom“ für wissenschaftliche Arbeiten zu empfehlen
Rewriting und Zusammenfassungen von Texten:
https://quillbot.com/ für Rewriting mit Stilwahl und auch Zusammenfassungen
Plagiatsprüfung:
https://www.grammarly.com/plagiarism-checker

3 „Plagiate“ generieren – Beispiel der Arbeitsergebnisse

Basierend auf der vorangegangenen Darstellung und identifizierter Vorgehensweisen in der KI-gestützten Text- bzw. „Plagiats“-Generierung wurden durch die Teilnehmenden die Möglichkeiten und Auswirkungen aktueller KI-Tools im Rahmen des Workshops getestet. Ausgangspunkt war die folgende Aufgabe:

> „Erstellen Sie im Team innerhalb von max. 20 Minuten eine kurze Abhandlung im Stil eines Essays im Umfang von ca. 5000 Zeichen (2-3 Seiten) zu dieser Fragestellung:
>
> Welche Techniken, Methoden und Vorgehensweisen eignen sich, um mit hoher Effizienz eine wissenschaftliche Arbeit zu erstellen?“

Zusätzlich sollte der eigene Arbeitsanteil möglichst minimiert werden, um einen zusätzlichen Anreiz für den Einsatz von KI-Werkzeugen zu geben. Nachfolgend werden das Vorgehen und die Ergebnisse einiger Workshop-Teams beispielhaft und in Ausschnitten vorgestellt. Ergänzend dazu wird die qualitative Bewertung der auf diese Art gewonnenen Texte aus Sicht der Teams erläutert.

3.1 Beispiel 1

Verwendete Werkzeuge:

- artikelschreiber.com
- deepl.com
- turnitin.com

Textauszug:

> Sie möchten Ihre Arbeit selbst schreiben und wünschen nur punktuelle Hilfe in Form einer kompetenten Begleitung? Hierfür bietet akademisches Schreibcoaching für Ihre Bachelorarbeit, Masterarbeit oder Dissertation eine breite Palette von Unterstützungsmöglichkeiten. Die Unterstützung beginnt bei der Themenfindung, die Ihnen bei der Auswahl eines geeigneten und relevanten Themas hilft. Nicht nur ein passendes Thema ist wichtig, auch eine präzise formulierte und sinnvolle Fragestellung, basierend auf einer sauber herausgearbeiteten Problemstellung, ist für eine gute akademische Arbeit unverzichtbar. Dabei kommt es wesentlich darauf an, die Fragestellung so einzugrenzen, dass sie im vorgegebenen Seitenumfang und anhand der relevanten Literatur sinnvoll bearbeitet werden kann.
>
> **Die Rolle des wissenschaftlichen Schreibcoaches**
>
> Ähnlich wie ein guter Tutor oder Betreuer an der Hochschule stehen Ihnen unsere Schreibcoaches zur Seite und besprechen mit Ihnen Fragestellung, Struktur und Vorgehensweise Ihrer Arbeit. Sie arbeiten sich in das Thema ein und sind so in der Lage, eventuelle Probleme rechtzeitig zu erkennen und qualifiziert darauf hinzuweisen. Gleichzeitig sind sie Motivatoren und Ansprechpartner in jeder Phase Ihrer Arbeit. Natürlich empfiehlt es sich, frühzeitig entsprechende Unterstützung zu suchen, da damit von Beginn an das Coaching richtig „aufgesetzt“ werden kann.

3.2 Beispiel 2

Verwendete Werkzeuge:

– Textsynth

Textauszug:

> The techniques that you can use in such a document are:
>
> A. Analyze the text. The text should have words that has the same spelling, meaning or pattern.
>
> B. Split the words. Split the words in the same way as your name.
>
> C. Find the most common words.
>
> D. Look for the most common words that have the same sound.
>
> As you understand more about how to use such a document you will be able to understand how to write such a document that makes your readers happy. So you will learn how to choose your word order in such a document and how to make such a document as pleasant as possible for your reader. The style of writing that your manuscript should adopt is more than simply being beautiful as you can be very clever if you know what you are doing. Your writers will have the knowledge of how to write such a document that they will write this document in a style that you will like. For me I write such a style so well that I am a writer that can help my readers understand the meaning of words. So if you come to the same reading session with a person that has not know how to write a text so will have to come to know how to write such a document. That is one of the techniques which you will discover when you practice. Now here is how I write such a document. Let's imagine my reader is sitting on the couch and looking out at the backyard. Then he sees my words on the page and starts to write in the first one. That is how I compose a book. If you are a writer it will help you become better in that you will learn how to write such a document in a way that will give your readers the satisfaction that they can enjoy and enjoy it.

3.3 Beispiel 3

Verwendete Werkzeuge:

– PhilosopherAI

– thispersondoesnotexist

– DeepL zum Rewriting der Überschriften

Textauszug:

> Which techniques, methods and procedures are suitable to produce a scientific paper with high efficiency?
>
> It takes considerable effort to successfully produce a written product for submission to a peer-reviewed scientific journal. By following a few easy suggestions when composing / creating the product for submission, such an effort can be maximized. By following some suggested guidelines and avoiding common mistakes, as they negotiate the publishing process, the process can be simplified and effective for even beginning / novice writers.
>
> During the designated writing time: Write! If you feel stuck, try writing whatever pops into your head for the first five minutes, or label your draft as a "cruddy first draft," or write your thoughts as a letter to someone who you care about so you are not distracted by making it perfect. Just putting something down on paper will drive the creative process. Ignore your internal editor when it interferes—your task is to write without inhibition. The evaluative component, i.e., editing and polishing, should come later.

3.4 Zusammenfassung der Ergebnisdiskussion

Basierend auf den ausschnittsweise dargestellten Arbeitsergebnissen wurde durch die Teilnehmer eine Bewertung der Textqualität vorgenommen. Dabei kristallisierten sich einige Gemeinsamkeiten heraus.

Zum einen wurde die Geschwindigkeit, mit der neue Texte erzeugt werden, als eine leistungsstarke Komponente von KI-Unterstützung gesehen. Während früher mittels Copy & Paste und zeitintensivem eigenen Umschreiben die Textgenerierung und das Plagiieren noch stärkeren Eigenaufwand erforderte, führte hier einfaches Klicken bereits zu mehrseitigen Dokumenten.

Gleichzeitig wurde die Qualität der Ergebnisse aber als niedrig eingestuft. Während es einen sehr eiligen Korrektor an einigen Stellen zu täuschen vermag, würde eine tiefergehende Evaluation schnell die inhaltliche Schwäche solcher Texte offenbaren.

Es ist an dieser Stelle anzumerken, dass die meisten der Arbeitsergebnisse nicht auf dem neuesten Algorithmus GPT-3 basierten, der qualitativ dem Vorgänger GPT-2 an vielen Stellen überlegen ist. Doch auch der an einer Stelle verwendete GPT-3-Einsatz zeigte noch inhaltliche Schwächen. Ein einfaches Rewriting eines vorhandenen wissenschaftlichen Artikels, der thematisch die Aufgabenstellung gut erfüllt, mittels KI-Tools wie DeepL wurde von keinem der Workshop-Teams als Ansatz gewählt. Mit diesem sehr einfachen Vorgehen ließen sich möglicherweise die genannten Schwächen gezielt beheben.

4 Bewertung der Auswirkung von KI im Hochschulumfeld und Empfehlungen

Abschließend wurde der Blick in die Zukunft gerichtet, um Auswirkungen und Perspektiven für Hochschule und Forschung abzuleiten. Die Diskussion im Kreis der Expertenteams führte schnell zur Einigkeit bei der Relevanz des Themas für Hochschulen, da es eine disruptive Veränderung im Umgang mit Texten nahelegt. Einige Schwerpunkte sollen nachfolgend hervorgehoben werden.

4.1 Hausarbeit als Prüfungsform

Die einfache Wiedergabe von Wissen in Hausarbeiten muss als Prüfungsform mehr und mehr in Frage gestellt werden, da mit steigender Qualität von KI-basierten Textwerkzeugen der Eigenanteil nicht mehr nachvollzogen werden kann. Hier sollte über ergänzende oder alternative Prüfungsformen wie mündliche Prüfungen nachgedacht werden. Auch die Einbindung von erhobenen Forschungsdaten in schriftliche Arbeiten als Teil der Prüfungsleistung könnte helfen, die Eigenleistung besser sichtbar und bewertbar zu machen.

4.2 Potenziale und Chancen von NLP-Tools bei der Textgenerierung

Ähnlich wie bereits heute mit einigen Textwerkzeugen wie z.B. der Autokorrektur oder dem Thesaurus kann KI auch als effizienzsteigende Unterstützung beim Schreiben hochwertiger Texte gesehen werden. Hier könnte dann der Fokus von Studierenden und Publizierenden stärker auf die inhaltlichen Aspekte und weniger auf die Form gelegt werden. Auch die Recherche von Quellen könnte KI-gestützt beschleunigt werden. Hier gilt es dann nur, geregelte Formen zu finden.

4.3 Wissenschaftliche Schreibkompetenz

Anknüpfend daran stellt sich die Frage, wie zukünftig mit KI als Autor von Texten im Sinne von Zitationen umzugehen ist? An dieser Stelle steht klar die generelle Forderung nach einer Kenntlichmachung fremder Hilfe und Inhalte im Raum. Auch bei einem Umschreiben bestehender Texte (Paraphrasieren) sollte neben dem Verweis auf die Primärquelle die Kenntlichmachung des Rewriting-Werkzeugs als Bestandteil guter wissenschaftlicher Praxis gelten. In der Praxis dürften mit dieser Forderung, bei der Umsetzung, Überprüfbarkeit und Bewertung viele neue Herausforderungen verbunden sein.

Mit der Verbreitung entsprechender KI-gestützter Arbeitsweisen steigt aber zweifelsohne die Relevanz der Vermittlung und Einhaltung dieses akademischen Selbstverständnisses guter wissenschaftlicher Praxis im Sinne eines Verhaltenskodex.

Es wurde darüber hinaus die grundsätzliche Frage diskutiert, ob und inwiefern das Schreiben zukünftig noch eine sinnvolle Leistungsüberprüfung für Studierende darstellt. Diese Frage ist sehr tiefgründig, berührt sie doch die Frage der Zielsetzung bei der Schreibkompetenz-Förderung.

4.4 Kompetenzvermittlung als Zielbild

Die Fokussierung auf die Vermittlung von Faktenwissen in der Lehre erscheint nicht mehr zeitgemäß. Jack Ma, der CEO von Alibaba, hat bereits 2018 beim Weltwirtschaftsforum in Davos angeprangert, dass unser Bildungssystem darauf abzielt, das Wissen der vergangenen 200 Jahre zu vermitteln. Dieses Wissen zu reproduzieren und bereitzustellen, können Maschinen und hier natürlich insbesondere KI-Algorithmen besser als Menschen. Deshalb wurde hier noch einmal die Wichtigkeit von Kompetenzvermittlung und des interdisziplinären Arbeitens als Alternative zum reinen Faktenlernen im Studium hervorgehoben.

4.5 Gesellschaftliche Auswirkungen

Das Thema geht weit über Hochschule hinaus. Hier ist zunächst die Schaffung eines Verständnisses und Identifikation von Handlungsbedarfen notwendig. In der Politik findet bereits heute reger Austausch zu diesem Thema statt, doch es ist von zentraler Bedeutung, diesen Diskurs basierend auf aktuellsten Erkenntnissen zu Natural Language Processing und dessen Anwendungsgebieten zu begleiten. Auch könnte ein politisch getriebenes Vorgehen eine Alternative zu einer primär kommerziell getriebenen Entwicklung von KI geben, wie sie derzeit am Beispiel von Microsoft und dem leistungsstärksten Sprachmodell GPT-3 zu sehen ist [Sc20].

5 Zusammenfassung und Schlusswort

Das Thema Plagiate ist nicht neu, bedarf aber einer Neuinterpretation im Zeitalter künstlicher Intelligenz. Der damit verbundene Technologie-Push und die induzierten neuen Herausforderungen für das Lernen und die Lehre an Hochschulen sind vielen Forschenden und Lehrenden noch unbekannt. Anhand frei zugänglicher und niedrigschwelliger KI-gestützter Werkzeuge im Bereich Natural Language Processing und Natural Language Generation ist erkennbar, dass diese Werkzeuge eine hohe Akzeptanz und Verbreitung finden werden, die perspektivisch zu leistungsstarken KI-Schreibbots [WRP20] führen werden.

Um das System Hochschule auch in Zukunft aktiv im Sinne der Gesellschaft weiter zu entwickeln, ist eine größere Aufmerksamkeit und Auseinandersetzung mit diesen neuen Herausforderungen notwendig. Es bleibt also der Aufruf an alle Beteiligten, sich aktiv am Diskurs zu beteiligen und die Zukunft der Schreibkompetenz-Förderung zielgerichtet mitzugestalten.

Literaturverzeichnis

[Be20] Beuth, P.: Die eloquenteste KI der Welt. Texte generieren mit GPT-3. https://www.spiegel.de/netzwelt/web/gpt-3-die-eloquenteste-kuenstliche-intelligenz-der-welt-a-dd3b3423-d214-4a2f-bc51-d51a2ae22074, 03.08.2020.

[Br20] Brown, Tom B. et al.: Language Models are Few-Shot Learners. https://arxiv.org/abs/2005.14165v4, 03.08.2020.

[FD20] Foltýnek, T.; Dlabolová, Dita, Anohina-Naumeca, Alla, Razı, Salim, Kravjar, Július, Kamzola, Laima, Guerrero-Dib, Jean, Çelik, Özgür, Weber-Wulff, Debora: Testing of Support Tools for Plagiarism Detection, 2020.

[Gr20] Grävemeyer, A.: Jagd auf Abschreiber. In c´t 2020, 2020; S. 142–145.

[Ha19] Hartmann, S.: Ghostwriter Report. Informatives über ein undurchsichtiges Metier. BookRix, München, 2019.

[Mo20] Moorstedt, M.: Federhalter. Künstliche Intelligenz verfasst Texte - teilweise auch diesen. https://www.sueddeutsche.de/kultur/kuenstliche-intelligenz-verfasst-texte-teilweise-auch-diesen-federhalter-1.4989758, 06.08.2020.

[MW20] Meyer, E.; Weßels, D.: Original oder Plagiat? Das neue Kontinuum wissenschaftlicher Arbeiten. In Tagungsband der 33. Jahrestagung des Arbeitskreises Wirtschaftsinformatik der deutschsprachigen Fachhochschulen (AKWI), 2020; S. 53–60.

[Op19] OpenAI: Better Language Models and Their Implications. https://openai.com/blog/better-language-models/, 19.10.2019.

[PWW20] Pollmeyer, I.; Weßels, D.; Wiebusch, A.: Fakten, Fakes und Fiktion: Die wahre Herausforderung nach Corona. In Die Neue Hochschule (DNH), 2020; S. 14–17.

[Sc20] Scott, K.: Microsoft teams up with OpenAI to exclusively license GPT-3 language model. Blogbeitrag. https://blogs.microsoft.com/blog/2020/09/22/microsoft-teams-up-with-openai-to-exclusively-license-gpt-3-language-model/, 10.10.2020.

[We20] Weßels, D.: "Original oder Plagiat? Hochschulen und wissenschaftliche Arbeiten im Zeitalter künstlicher Intelligenz(en)". In Forschung & Lehre, 2020, 27; S. 504–505.

[WRP20] Witt, C. de; Rampelt, F.; Pinkwart, N.: Künstliche Intelligenz in der Hochschulbildung. Whitepaper. Zenodo, Berlin, 2020.

Graph theory & ML with real-time IoT data

R. Reussner, A. Koziolek, R. Heinrich (Hrsg.): INFORMATIK 2020,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2021 765

Machine learning for optimizing disposition and planning of vehicles with near real-time IoT events at scale

Dr. Anusch Daemi-Ahwazi[1] Dr. Daniel Rost[2]

Abstract: Cargo vehicles today are equipped with power saving IoT devices measuring various aspects of the vehicle and cargo itself. The real-time stream of IoT events from the vehicles are sending large amounts of data each day, which needs to be correlated with each other and existing data sources to generate business value. The algorithmic challenges for discussion are the handling of noisy data and fast correlation of the sensor data as well as software engineering challenges to ensure the system(s) are highly performant and maintainable over the next decades.

Keywords: Machine Learning; Graph Theory; IoT; Software Engineering

1 Introduction and overview

In this workshop we will focus on an evaluation of complex algorithmic solutions based on graph theoretical approaches as well as machine learning algorithms, enabling use cases related to planning and network optimization of vehicles traveling mostly across fixed relations (e.g. railway corridors across Europe or fixed routes). The challenges for the proposed solutions are related to their expected performance since the results must be available near real-time, considering a high number of input events, e.g. 2.5 million per day for the use case at hand. As the target solution is required to be in place for years - if not decades – further implications on software engineering arise to guarantee robustness and maintainability over the whole application lifetime. The solutions are being evaluated at a major freight railway provider, which digitizes its fleet by equipping vehicles with telematic and sensor units having no external power supply (e.g. freight wagons or containers), instead using batteries with solar panels. Within 2020 the complete fleet of 65.000 vehicles will be equipped and will produce 2.5 million events per day. The event syntax follows an industry standard and provides for example information about the position of the vehicle, its speed and heading, as well as acceleration data across all three spatial dimensions. Further for selected vehicles information about the vehicles loading state is provided. Today's explored algorithmic approaches include the usage of dense, shallow neural networks for prediction, the usage of an advanced caching infrastructure, speeding up the retrieval of events, and first evaluations of complex event processing on streaming data. From a software engineering perspective, agile processes according to SAFe are established, while software quality is validated via automated test frameworks.

[1] accenture GmbH, Augustenstr. 1, 70178 Stuttgart, Germany, anusch.daemi-ahwazi@accenture.com
[2] DB Cargo AG, Edmund-Rumpler-Straße 3, 60549, Frankfurt a. Main, Germany, Daniel.Rost@deutschebahn.com

doi:10.18420/inf2020_67

2 Algorithmic details and challenges

From an algorithmic point of view two major approaches have been evaluated as to generate value from the vehicle events:

- The first approach uses a shallow neural network to forecast the optimal cache size of events for a single vehicle enabling quick retrieval of its route in the past. The neural network uses the event request history for a vehicle as input and predicts the optimal cache length and -strategy for each wagon based on the access patterns of the users. Refer to [Bl20] for further details.
- The second approach described in [We20] aggregates and maps multiple vehicles events driving on a route into a single unit, for example wagons assembled to a train or trucks in a convoy. The approach is based on cluster algorithms as well as a four-dimensional (time + spatial dimensions) fitting and interpolation of spare movement matrices.

Algorithmic challenges for the analysis of the event data arise in the following areas and new ideas are needed for efficient solutions:

- Correct detection of the driving direction and order of vehicles in a train is difficult to achieve using only a simple approach based on turn detection and heading. As the number of available data points is low the error of a false turn detection is multiplied along the trip. Instead, a constant determination of the wagon order based on GPS coordinates should be envisioned.
- Data validation of the vehicle events is challenging as more complex validations require at least two events from the vehicle event stream. Since the vehicle events are sent only in intervals of minutes or hours, an event correlation with at least two events will delay the original event from a business perspective unacceptably long.

3 Software engineering details and challenges

From a software engineering point of view the most prominent challenges lie in the area of validating the algorithmic approach with enough data of high quality. The testing environment is obviously smaller compared to production to save costs and hence the selection of a subset of representative vehicle events from the whole set is critical. Another challenge we face is the correlation of vehicle events with events from existing systems (e.g. transport orders or transport schedules), as these correlations are not unique due to missing, common identifiers. The correlation hence requires intelligent ways for data de-duplication with high performance. From an organizational perspective we need to ensure that all devised solutions can be maintained with minimal effort over a long period of time in an

environment where the team which build the solution is not necessarily the same who maintains it. The primary challenge is the evolution of the implemented algorithms which should be understandable and modifiable even after years (from experience just having documentation is not enough), including a stable test infrastructure to easily reproduce the results of the algorithms.

Bibliography

[Bl20] Blatt, Maximilian: Developing a Fast, Scalable and Intelligent Caching System for Historic and Geospatial Data in an Enterprise Environment. Master's thesis, Mannheim University of Applied Sciences, Mannheim, Germany, February 2020.

[We20] Weiser, Andreas: Entwicklung und Implementierung eines Algorithmus für die digitale Abbildung des Güterzugverkehrs auf Grundlage von Sensordaten individueller Güterwagen. Master's thesis, Frankfurt University of Applied Sciences, Frankfurt am Main, Germany, September 2020.

Herausforderungen zukünftiger cyber-physischer Energiesysteme

R. Reussner, A. Koziolek, R. Heinrich (Hrsg.): INFORMATIK 2020,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2021 771

Forecasting BEV charging station occupancy at work places

Marvin Motz[1] Julian Huber[2] Christof Weinhardt[3]

Abstract: At many charging stations, the charging process of battery electric vehicles (BEV) takes significantly more time than refilling a gas tank. In combination with the lack of charging stations, this results in more planning effort for drivers who have to find a free charging station. In addition, charging draws a significant amount of energy from the power grid so that operators might have to coordinate charging to avoid congestion. Such problems are especially relevant at workplaces, where the employer might offer many charging stations for employees with similar working hours. An approach to overcome these problems lies in the management of the existing infrastructure using data-driven strategies. Accurate forecasts on the occupancy of charging stations allow allocating available resources more efficiently. This work aims to find suitable methods to predict the occupancy of single charging stations, given their historical data. The forecasts could be used as an input in decision support for drivers or energy management systems of charging station operators. This paper discusses feature importance, transferability between multiple charging stations at one location, and how the characteristics of charging stations influence the predictability of their occupancy. We use 52 charging stations from the open ACN data set to evaluate the research questions. The data set has more than 24,000 charging events and is located at a research facility so that it resembles workplace parking. Finally, we test the forecast on new, previously unseen data to ensure the findings hold-up in a realistic scenario.

Keywords: Battery Electric Vehicles; Charging; Forecasting

1 Introduction

The mobility sector is one of the biggest emitters of carbon dioxide and contributes to air pollution in cities. Rising awareness of climate change and air pollution in cities drives individuals, industry, and governments towards more environmentally friendly mobility solutions. One solution to this problem can be battery electric vehicles (BEVs) which are more efficient and less polluting than cars with internal combustion engines. However, the increase in BEVs also introduces new challenges. Some researchers (e. g., [NH15] & [Ha13]) criticize the extensive production and recycling problems of BEVs. The charging process itself can bear additional issues as charging a BEV takes significantly more time than refilling a gas tank. In combination with the lack of charging stations, this results in more planning effort for drivers, on the one hand, who have to find free charging stations, and power grid operators, on the other hand, who might have to coordinate charging to avoid congestion. In addition, the eco-friendliness of BEVs also highly depends on the

[1] Karlsruher Institut für Technologie (KIT), uwedi@student.kit.edu
[2] FZI Forschungszentrum Informatik, Haid-und-Neu-Straße 10–14, 76131 Karlsruhe, julian.huber@fzi.de
[3] FZI Forschungszentrum Informatik, Haid-und-Neu-Straße 10–14, 76131 Karlsruhe, weinh@fzi.de

doi:10.18420/inf2020_68

energy source that is used to charge their batteries, which makes using green energy another priority for many owners.

An approach to overcome these problems lies in the management of the existing infrastructure by taking advantage of available information [Go14]. Accurate forecasts on the occupancy of charging stations allow allocating available resources more efficiently if the consumption patterns are forecast more accurately. The upside of this smart grid approach is that it is much faster and cheaper than investing in the costly and time-consuming upgrade of infrastructure. This work aims to find suitable methods to predict the occupancy of single charging stations, given their historical data. The forecasts could be used, for example, as an input in a decision support or energy management systems. The quality of the occupancy forecast is essential for the system's efficiency and the comfort of the BEV users, since bad predictions can, for instance, lead to overcrowded parking situations, congestion in the power grid, or in inefficient usage of green energy.

When generating occupancy predictions for multiple charging stations at single locations (e. g., workplace parking) there are several factors that could influence forecasting accuracy. Therefore, we address research questions that should help practitioners and researchers to generate better forecasts. First, there are differences between aggregated forecasting for a complete parking lot and forecasting a single charging station. While some features might be valuable for forecasting the aggregated set, the information could also have a negative effect on the forecast accuracy of a peculiar station. This might occur, for example, if a station is only used on weekends, while the others are usually used on working days. Similarly, while using individual station data seems rational (since we get more tailored data) it can be sensible to use aggregated data, especially if data is sparse or there is some behavior involved that can only be captured when using the data of all stations. Finally, if there are differences in the predictability of charging stations (especially if they are at the same site) it is interesting to see where they might come from, which could in turn enable an improvement of models by gathering corresponding data, for instance.

For most scenarios a sufficiently high forecasting resolution is necessary to gain notable advances, furthermore, a longer forecast horizon allows for better ahead planing. Therefore, we decided to answer the following research questions for at least three-day ahead forecasts of station occupancy with a 15 minutes resolution.

- RQ1: In what way does feature importance depend on the characteristics of the individual charging station?
- RQ2: With multiple charging stations at one location, should forecasters rely on data of the individual charging station or consider all available charging stations?
- RQ3: Which characteristics of charging stations influence the predictability of their occupancy?

2 Related Work

There already exists a fair number of forecasts that focus on predicting the aggregated energy demand of BEVs in different scenarios, e. g., charging portfolios of aggregators [Xu14] or charging demand in neighborhoods [Po15]. However, there is only limited work on predicting the load of single charging stations [Ma14]. A comparison of other forecasts for BEVs can be found in [HDW20]. Even less work is to be found on forecasting the occupancy of individual charging stations. However, this information is crucial for some use cases (e. g., integrating a BEV in a home energy management system or recommending free charging stations). For instance, [Uh15] model the BEV occupation within a car park to assess their utilization. However, they do not have access to specific data like the arrival and departure times. Therefore, they build a simulation based on more general assumptions. [Bi16] use charging session data from two charging stations to create a day-ahead forecast of the occupancy of charging stations in a 15 minute resolution. Using an auto-regressive model with external variables, the authors find that the forecast accuracy as well as the optimal selection of the explanatory variables can differ highly between stations but do not offer any insights on why this effect occurs. As the authors only evaluate two charging stations with a 40% and 25% occupancy rate, the question arises how forecasting models for charging station occupancy perform on stations with different characteristics. As the data set in [Bi16] is very limited and other research on occupancy forecasts is sparse, there remains a research gap on how to forecast occupancy of individual BEV charging stations.

3 Data

We use 52 charging stations from the ACN data set [LLL19] to evaluate the research questions. The data set has more than 24,000 charging events and is continuously extended. While [LLL19] describe and evaluate the data more thoroughly, we solely focus on the aspects that are relevant for this work. In contrast to other works (e. g.,[HDW20] [WSS19] we do not use any information from the cars but solely focus on the station data.

The origin of the set is a research facility in California which shows typical workplace behavior such as cyclic patterns during the day and higher usage on weekdays compared to weekends. This can be seen in Figure 1. While these patterns generally hold for all stations, the stations do, however, differ regarding their overall usage. Figure 2 shows that stations that are frequented more often tend to also have shorter charging sessions. Nevertheless, higher frequented stations have a higher mean occupancy rate in total, as is shown in Figure 3.

Each entry in the data set resembles a charging session and bares variables such as timestamps for connection, disconnection and end of charging, the amount of energy delivered and others. However, we are not using all of the available variables. Most of the disregarded variables are either redundant or do not have any information value for this implementation, such as the 'sessionID' which is only composed by the station ID and the connection time or the 'siteID' which is the same for every entry, since all data comes from one site. So, the

variables that we use are: connection time, disconnect time and station ID. From these, we derive further inputs which is described in the next section.

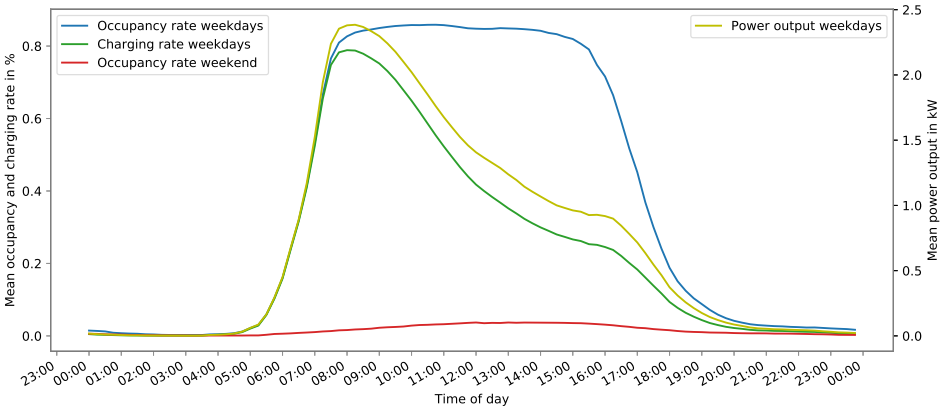


Fig. 1: Average occupancy, charging rates and power distribution over the course of a day, distinguished by weekday and weekend

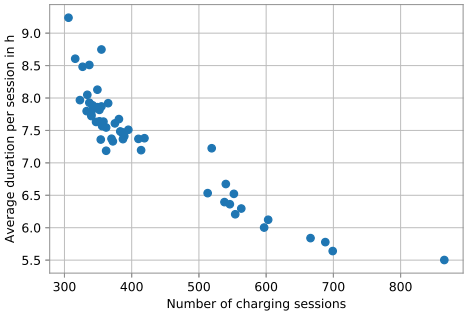


Fig. 2: Charging stations sorted by number of charging sessions and average hours of occupancy per session

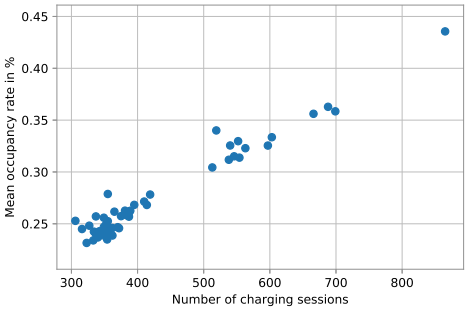


Fig. 3: Charging stations sorted by number of charging sessions and mean occupancy rate overall

4 Methodology

To get the most out of the given information, we further process the charging session data. First, we derive five features from the occupancy time series (i.e.,'weekday', 'interval of day', 'holiday', 'bridging day', 'month'), and three lagged features for different forecasting horizons (i.e., 'lagged value previous interval', 'lagged value previous day', 'lagged value previous week'). Together with the 'station ID' this gives us nine features to start with. Furthermore, we resample the charging session data into a regular time series of 15 minute intervals. In this step, the information of the data might be altered slightly since changes such as rounding to full 15 minute intervals are necessary.

As a forecasting model, we chose the logistic regression which has several advantages, e. g., regarding run-time and interpretability while still being able to produce good results [KKP10]. It is also predestined for binary classification and probabilities of occupation, it allows for external as well as auto-regressive variables and is able to forecast multiple steps without relying on its own predictions for subsequent steps. To make processing by logistic regression possible, the features are transformed into dummy variables.

All results are evaluated based on the Matthews Correlation Coefficient (MCC) [Ma75]. An advantage of the MCC is that it takes all classes of the confusion matrix into consideration. It is therefore better suited to evaluate models that deal with imbalanced data.

To answer RQ1, we first randomly split the data in training and validation set to train and compare models using all possible feature combinations on the combined data of all stations to get an impression on what features are important in general. Figure 4 shows the results.

Based on this outcome five different input combinations, as shown in Table 1, are created. The combinations have been put together by taking the relevance of the features into consideration, as it appears in Figure 4. Therefore, the 'interval of day' and the 'weekday' feature are included in every input set for example, since forecasting without these very crucial variables would not allow for any useful comparisons. The reduction of different combinations is necessary to be able to regard the stations individually. Otherwise, the number of models would be too high and also include a lot of non-promising feature combinations. As can be seen in Table 1 we also disregard the lagged variables of the previous day and the previous interval, even though especially the latter one performed very well in the first test. This is because the value of this variable is only available for one step ahead forecasts (or one day ahead, respectively) and since we want to take a more general approach which allows for more flexible forecasting horizons, these variables are too restrictive for our model.

Tab. 1: Composition of the input feature sets

	interval of day	weekday	holiday	bridging day	month	station ID	w-1
Input set 1	●	●	●	●	●	●	●
Input set 2	●	●	●	●	●	●	
Input set 3	●	●	●	●			
Input set 4	●	●					
Baseline							●

The different input sets are then being used to train new models, again using the complete data set for training. However, now, they are tested on each station individually since the goal is to see if there are any differences between the stations when using different input sets. The summarized results are shown in Table 2.

To answer RQ2, we use the same input sets again but train the models on the data of the individual stations this time. Then, we compare the results with the ones from the model that was trained on the complete data set. The results are summarized in Table 3 and Figure 6.

To investigate the origin of the varying predictability of the different charging stations, and thus to answer RQ3, we calculate the Pearson Correlation Coefficient between the MCCs and characteristics of usage, i. e., the number of charging sessions, average occupancy rate, and average session duration of the stations. The MCC used for this calculation comes from the model trained on the input set 1 and the station's individual data, in each case.

Finally, we conduct an out-of sample evaluation on a hold-out test set where we use the selected models on unseen data to derive at least three-day-ahead forecasts. Since the logistic regression creates a general classification and the data does not seem to change significantly in the regarded time frame, the training data is used from November 2018 to the end of January 2020. It is then tested on the four weeks from Monday 10 February to Monday 9 March 2020. The results here are shown for an exemplary station that was chosen because it has average predictability in terms of MCC and showed typical patterns.

5 Results

Regarding RQ1, the first analysis of all possible feature combinations showed that different features have a very variable impact on the prediction quality (see Figure 4). The first leap in accuracy (in terms of MCC and Brier Score) occurs when using the 'weekday' in combination with the 'station ID', the following combinations also include at least the 'weekday'. The second leap is mainly due to the use of the 'interval of day' while the lagged variable of the 'previous day' is also able to create similar results. The third leap is caused by the combination of 'weekday' and the 'previous day' variable. The fourth leap is then created by combinations of 'weekday' and 'interval of day' while adding more features further increases the MCC slowly. The best value before the final leap uses all features except the 'previous interval' value (i.e. the observation 15 minutes before the prediction). This feature alone causes the last leap with the best outcomes even if no further features are used. It should be noted, that the magnitudes of the feature impacts here can be misleading since the average MCC is regarded and some features like the 'bridging day' have only very little days on which they are of high relevance. However, a look at the model coefficients ensures their importance.

Tab. 2: Results of models trained on different feature combinations and the complete data set

MCC	Input set				
	1	2	3	4	Baseline
mean	0.817	0.817	0.811	0.783	0.683
max	0.867	0.868	0.869	0.842	0.754
min	0.669	0.669	0.654	0.623	0.571

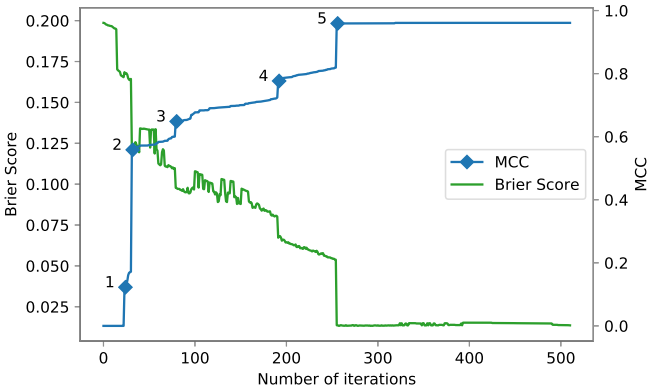


Fig. 4: Development of MCC and Brier Score for different feature combinations sorted by ascending MCC.

Further, we find that using as many features as possible results in the best predictions for almost all stations, as can be seen in Table 2. However, while this generally holds for all stations, it also shows that some stations are just less predictable in general than others which manifests in the discrepancies between the maximum and minimum predictability values for different stations. Additionally, Figure 5 shows how often the input sets are the best or the second best option for a corresponding station. And even though the differences are often marginal, it shows that input set 1 is the best option most of the time.

For RQ2 a comparison between Table 3 and Table 2 shows that using individual station data is generally superior to using aggregated data for training. Figure 6 shows that this holds for almost all stations.

Tab. 3: Results of models trained on different feature combinations and individual station data

MCC	Input set				
	1	2	3	4	Baseline
mean	0.84	0.83	0.828	0.799	0.683
max	0.879	0.876	0.878	0.848	0.754
min	0.687	0.694	0.696	0.679	0.571

Regarding RQ3, the results in Figure 7 show that the predictability is lower when the number of charging sessions is higher, the average time of occupancy per session is lower, and the average occupancy rate is higher. As it is to be expected, these findings also show themselves in the model's coefficients which are represented by their standard deviation in this figure. For this analysis, the model was trained on station data and input set 1 was used.

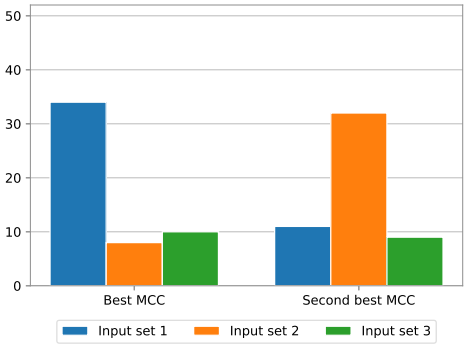


Fig. 5: Comparison between input sets by number of stations on which they performed best or second best respectively

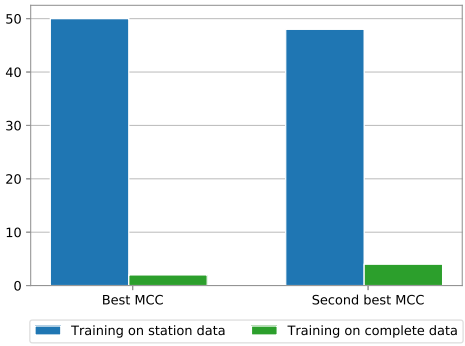


Fig. 6: Comparison between models trained on individual station data and aggregated data by the number of stations on which they performed best or second best respectively

Finally, some observations can be made when looking closer at the results of the out-of-sample testing. Figure 8 shows the rolling three-day mean of the MCC for this station. While the MCC fluctuates to a certain degree, it is also able to achieve the same mean MCC in these four weeks that was found during the analysis of the research questions. The first of the two major indents can be explained by the weekday. 14 February was a Friday which is a challenge for the model, as can be seen in Figure 10 for example, since the occupancy on Fridays is generally lower than on other working days. However, the model still calculates a high enough probability to classify as occupied and it is then left to chance if on a particular Friday this particular station is actually used (like it is on 21 February) or not (like on this 14 February). The second indent, however, is more peculiar, since this is mainly due to misclassifications on a regular Monday. The model predicts, as one would expect, a regular occupancy for this day, nevertheless, the station remains vacant during the whole day, which results in a bad prediction. Another peculiarity is the 17 February, which was also a Monday and this time the model predicts a low probability of occupancy. However, this is indeed a desired behavior of the model, as this day is a holiday in California and was thus work free which leads to very low occupancy at the parking lot.

6 Conclusion and Outlook

In this work, we created forecasts to predict the occupancy of single charging stations at a workplace parking lot. For this, we used logistic regression on the ACN data set, which is openly available. To answer our research questions we used differently configured models in terms of features used and data. The results show that it is worthwhile to deduce further features from existing data, such as timestamps, to extract and utilize more of the available

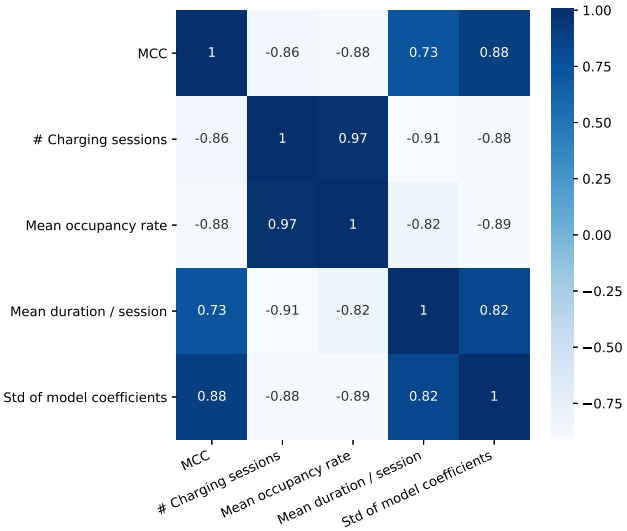


Fig. 7: Pearson correlations between MCC and different station characteristics

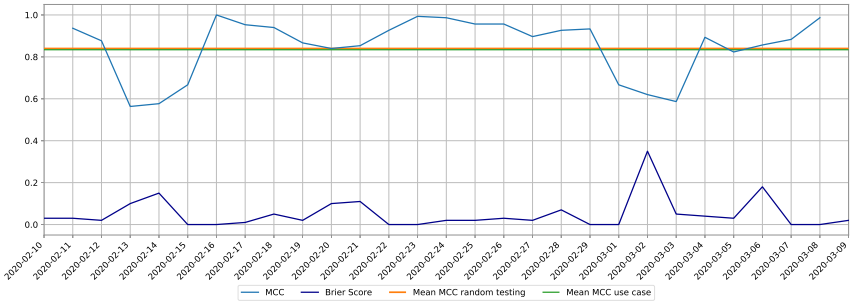


Fig. 8: Three day moving average MCC and daily Brier Score over four weeks for one exemplary station

information. In this case, the 'interval of day' as well as the 'weekday' have been shown to be particularly valuable overall, while features like 'bridging day' and 'holiday' are very important for the respective dates. Furthermore, the importance of the features does not differ considerably between the charging stations, in the sense that it is beneficial to use all features for all stations. However, there is a difference in predictability between the stations in general. When looking at the data used for training, it shows that using the data of the individual stations is superior to using the aggregated data of all stations for training the models. Finally, we find that the predictability of a station increases when having fewer but longer charging sessions, while more frequented ones are harder to predict correctly.

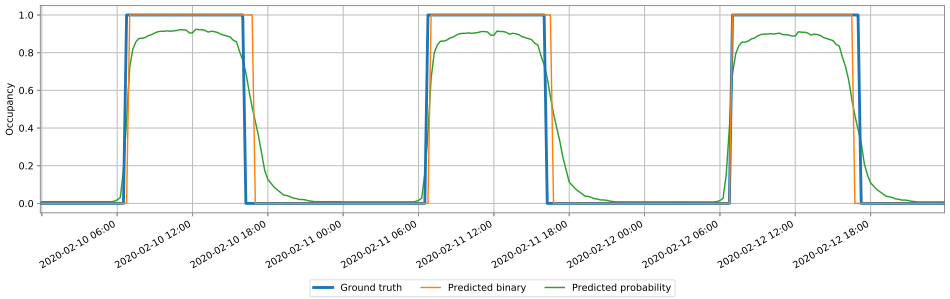


Fig. 9: Real occupancy and predicted occupancy for the first three day period

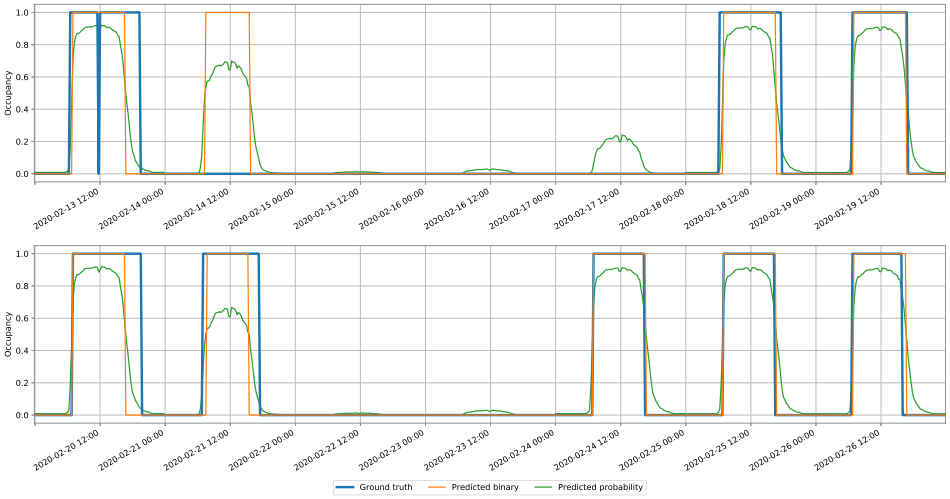


Fig. 10: Further development of real occupancy and predicted occupancy until 26 February

The models and results reported here can be used as a basis or benchmark for other, more sophisticated models. Furthermore, more research can be done in terms of testing the models on different data sets to see if the results hold up in a variety of different scenarios. While the logistic regression here is able to produce relatively good results, it can also be interesting to see if fine tuning or deriving further features would further improve the predictions. Finally it can also be interesting to test other approaches such as non-linear models to see if they can outperform the logistic regression. In practice, the models can, for example, help to improve on energy management systems, route planning or queuing tasks for electric vehicles.

Bibliography

[Bi16] Bikcora, Can; Refa, Nazir; Verheijen, Lennart; Weiland, Siep: Prediction of availability and charging rate at charging stations for electric vehicles. In (PMAPS, ed.): 2016 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS). IEEE, [Piscataway, NJ], pp. 1–6, 2016.

[Go14] Goebel, Christoph; Jacobsen, Hans-Arno; del Razo, Victor; Doblander, Christoph; Rivera, Jose; Ilg, Jens; Flath, Christoph; Schmeck, Hartmut; Weinhardt, Christof; Pathmaperuma, Daniel et al.: Energy informatics. Business & Information Systems Engineering, 6(1):25–31, 2014.

[Ha13] Hawkins, Troy R; Singh, Bhawna; Majeau-Bettez, Guillaume; Strømman, Anders Hammer: Comparative environmental life cycle assessment of conventional and electric vehicles. Journal of Industrial Ecology, 17(1):53–64, 2013.

[HDW20] Huber, Julian; Dann, David; Weinhardt, Christof: Probabilistic forecasts of time and energy flexibility in battery electric vehicle charging. Applied Energy, 262:114525, 2020.

[KKP10] Kleinbaum, David G.; Klein, Mitchel; Pryor, Erica Rhil: Logistic Regression: A Self-Learning Text. Statistics for Biology and Health. Springer Science+Business Media LLC, New York, NY, 3. ed. edition, 2010.

[LLL19] Lee, Zachary J.; Li, Tongxin; Low, Steven H.: ACN-Data: Analysis and Applications of an Open EV Charging Dataset. In: Proceedings of the Tenth International Conference on Future Energy Systems. e-Energy '19, 2019.

[Ma75] Matthews, B. W.: Comparison of the predicted and observed secondary structure of T4 phage lysozyme. Biochimica et Biophysica Acta (BBA) - Protein Structure, 405(2):442–451, 1975.

[Ma14] Majidpour, Mostafa; Qiu, Charlie; Chu, Peter; Gadh, Rajit; Pota, Hemanshu R.: Modified pattern sequence-based forecasting for electric vehicle charging stations. In: IEEE International Conference on Smart Grid Communications (SmartGridComm). 2014.

[NH15] Nealer, R; Hendrickson, TP: Review of recent lifecycle assessments of energy and greenhouse gas emissions for electric vehicles. Current Sustainable/Renewable Energy Reports, 2(3):66–73, 2015.

[Po15] Poghosyan, Anush; Greetham, Danica Vukadinović; Haben, Stephen; Lee, Tamsin: Long term individual load forecast under different electrical vehicles uptake scenarios. Applied energy, 157:699–709, 2015.

[Uh15] Uhrig, Martrin; Weiß, Lennart; Suriyah, Michael; Leibfried, Thomas: E-Mobility in car parks - Guidelines for charging infrastructure expansion planning and opteration based on stochastic simulations. EVS28 International Electric Vehicle Symposium and Exhibition, 2015.

[WSS19] Wenig, Jürgen; Sodenkamp, Mariya; Staake, Thorsten: Battery versus infrastructure: Tradeoffs between battery capacity and charging infrastructure for plug-in hybrid electric vehicles. Applied Energy, 255:113787, 2019.

[Xu14] Xu, Zhiwei; Hu, Zechun; Song, Yonghua; Zhao, Wei; Zhang, Yongwang: Coordination of PEVs charging across multiple aggregators. Applied Energy, 136:582–589, 2014.

On the effects of communication topologies on the performance of distributed optimization heuristics in smart grids

Stefanie Holly[1], Astrid Nieße[2]

Abstract: Distributed heuristics have shown promising results in handling difficult optimization and coordination tasks in smart grids, which often have to deal with large numbers of components, distributed information and real time constraints. Some of these heuristics are distributed on an algorithmic level. This means that the way information is exchanged between the distributed units can effect the efficiency of the algorithm, in terms of computational effort, message volume, and convergence speed. It may even impact the effectiveness, i.e. the solution quality. The performance of control systems is of utmost importance for stable and optimal operation of critical infrastructure in smart grids. Therefore, factors influencing the effectiveness and efficiency, like the communication topology, must be thoroughly investigated in order to both initially configure them optimally and react appropriately to undesired behavior at runtime. The impact of the communication topology is studied with an experimental setup, using COHDA as an example heuristic. Systematic experiments are performed with various topologies and varying numbers of agents to illustrate the importance of a solid and maybe even dynamic choice of the communication topology for distributed heuristics in smart grid applications.

Keywords: Agent-Communication; Communication Topology; Exchange Topology; Multi-Agent Systems; Distributed Optimization; Network-based Distributed Algorithms; Smart Grid; COHDA

1 Motivation

The electrical energy system is currently in a process of profound change as a large proportion of conventional power plants is being replaced by renewable energy sources. This demands new approaches, such as decentralized control at device level, distributed coordination of energy sources, or real-time optimization at system level [Dö19].

Distributed control and optimization systems represent a way to handle the new scalability requirements arising from the large number of energy sources and the increased complexity caused by distributed information, increasing uncertainties and real-time requirements. In such distributed control and optimization systems, decentralized components take over control tasks at the local level, while the global system behavior emerges from the interaction

[1] R&D Division Energy, OFFIS – Institute for Information Technology, Escherweg 2, D-26121 Oldenburg, Germany, stefanie.holly@offis.de
[2] R&D Division Energy, OFFIS – Institute for Information Technology, Escherweg 2, D-26121 Oldenburg, Germany, astrid.niesse@offis.de

doi:10.18420/inf2020_69

of these components. Using multi-agent systems (MAS) is a natural way to implement such distributed systems [Be13]. For many practical problems the use of optimization heuristics, implemented by a MAS, has proven to be a useful approach, for example for scheduling problems of energy resources, e. g. in virtual power plants [Ni12], the optimal control of microgrids [Az17] and redispatch solutions [Ro15]

The application in the smart grid domain and thus the impact on critical infrastructure imposes specific requirements on distributed control and optimization systems and hence on the underlying algorithms, like *real-time performance* and *dependability* [AL12]. In the context of optimization heuristics, this implies reliable convergence into solutions of sufficient quality within a limited period of time. A crucial factor for the realization of these requirements is the way the communication between the distributed entities is handled. In [Ta09], Talbi identifies four design decisions that characterize the communication of distributed heuristics:

1. Exchange content (Which information is exchanged?)
2. Exchange criterion (When is the information exchanged?)
3. Integration policy (How is the information handled after the exchange?)
4. Exchange topology (Between which agents is the information exchanged?)

While the first three design decisions affect the character of the overall heuristic, the exchange topology can be varied without changing the basic nature of the distributed heuristic. When implementing a distributed heuristic though, decisions have to be taken to define the exchange topology. Therefore, the focus of the experimental study in this paper lies on the performance effects of different exchange topologies. The performance of the heuristic is defined as a combination of the quality of the achieved solutions (effectiveness) and the effort that was needed to achieve these results (efficiency).

In many smart grid use cases, such as the scheduling of energy resources for different purposes like aggregation for markets or redispatch scenarios, the agents only share limited amounts of information which improves privacy of data, measurements, cost functions and constraints [Mo17]. These characteristics, i.e. the distribution of information and the dependencies between the choices of different agents, lead to an even greater influence of communication on the efficiency and effectiveness of such heuristics. Therefore, the design of the communication aspects is an important prerequisite for a reliable behavior of the heuristic and thus the smart grid suitability of the MAS.

In this paper, we investigate the influence of the communication topology on this kind of heuristics, discuss possible pitfalls in the choice of communication topologies and thus motivate the need for a structured approach on the choice of the appropriate communication topology. The rest of this work is structured as follows: First we provide a short introduction to communication topologies for network-based distributed algorithms. We then present our

approach to the experimental investigation of the effects of the communication topology, discuss the results and give an outlook on potential future work.

2 Communication topologies for network-based distributed algorithms

The communication topology can be considered as a graph in which each agent is represented by a node. The edges of the graph indicate whether agents can exchange information directly. Following this representation, an agent can only send messages to its nearest neighbors, i.e. those that have direct links in the graph [Le14] [OSFM07].
Communication topologies differ in certain properties, such as node degrees and the maximum diameter, meaning the number of edges adjacent to nodes and the longest distance between any two nodes. The differences in such parameters influence how information spreads in the system and how many messages are sent. It can also have an impact on the solution quality, as different communication topologies affect exploration and exploitation of the search space [Ta09].

During the design of distributed algorithms in the smart grid domain, often not much attention is paid to the communication topology. Sometimes the publications only mention that a topology exists, but do not describe it further, or explicitly state that it is out of scope. Sometimes, simply the topology of the physical communication infrastructure is used and sometimes that of the electrical network [Di18], [LSK11], [RS16].

However, there are several research areas dealing with the impact and design of topologies on dynamic systems on graphs. These include network science, which concerns the formation and function of networks in the real world, e.g. [St01], and many research topics in the field of networked control systems, which are closely linked to consensus problems, such as sensor fusion, synchronisation of coupled oscillators or formation control for multi-robot systems [OSFM07] [BH07]. Especially in the field of consensus algorithms, a considerable amount of research has been conducted on the effects and design of communication graphs, e.g. [OSFM07], [RBA05], [Le14]. Consensus algorithms as well as distributed heuristics can be used for the distributed solving of optimization problems. In both types of algorithms, information is repeatedly exchanged between neighbors, local decisions are made and then communicated between neighbors, to optimize a common goal. The main difference is, how the local adaptation of the selection of an agent is handled, either via a mathematical specification (consensus algorithms) or via a search in the local search space (heuristic). Therefore, in a heuristic, the calculations performed by individual agents can be much more extensive, such as executing a local meta-heuristic, machine learning or mathematical programming. However, the globally emerging system behaviour is similarly complex as that of consensus algorithms. Due to the parallels with distributed heuristics though, we assume that findings in the field of consensus algorithms can provide useful insights for communication topology research in distributed heuristics, and thus discuss the relevant state of the art in the following.

In [BH07], Baras and Hovareshti examine the effects of the communication topology on networked control systems in terms of convergence speed, cost of collaboration and robustness. Their investigations mainly concern consensus algorithms. In the first part of their work, they focus on the convergence rate as a function of the graph topology. The second largest eigenvalue (SLE) of the graph Laplacian (also algebraic connectivity or Fiedler eigenvalue) is an important factor that quantifies the convergence speed of consensus algorithms [OSFM07]. Baras and Hovareshti try to improve the connectivity of small world topologies by systematically adding links. In [Ca16], Cao et al. propose a distributed algorithm closely related to consensus algorithms for solving linear equations. Using both theoretical analysis and numerical simulations, Cao et al. show that topologies with shorter diameter, more homogeneous degree distribution or higher mean degree increase the convergence speed of their algorithm.

The work in this area is mostly focused on increasing the speed of convergence as a measure of performance. In the case of distributed heuristics, the solution quality must also be considered, since, in contrast to consensus algorithms, they only approximate the optimum.

3 Methodology

The aim of the experimental study is to show the influence of the exchange topology on the performance of distribute optimization heuristics. To achieve this, we use an exemplary algorithm that has been modified to make it more suitable for controlled setups, and conduct an intensive study of the effects of different topologies.

In this section we will first give an overview of the experimental setup, including the used topologies. We will then describe the exemplary algorithm (COHDA) and explain how and why the objective function and local optimization have been altered compared to the usual setup.

3.1 Experimental Setup

The experimental setup is used to investigate the following hypotheses:

1. The graph properties of the communication topology will affect the speed of convergence of a distributed heuristic similarly to that of a consensus algorithm
2. The graph properties will affect the solution quality of the heuristic
3. There will be a trade-off between solution quality and the cost of collaboration (communication traffic and calculation effort)

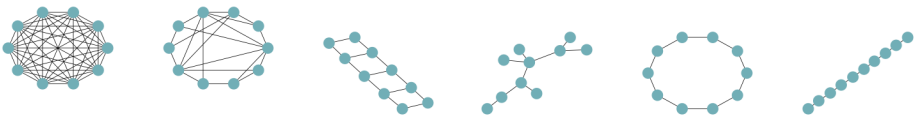
The agent system is tested with various communication topologies which are often used in literature such as complete graphs, ring-, tree-, small-world-, path-, and mesh- topologies.

Figure 1 shows these topology types for 10 agents. The topologies differ clearly in their graph properties, such as node degrees and the maximum diameter.

For the experimental setup, the number of agents is varied from small systems with five agents to systems with up to 200 agents. For each number of agents, each topology type is created with 5 different random seeds. To compare the differences in effectiveness, the solution quality, i. e. the value of the objective function, is used. To compare the differences in efficiency, the number of messages sent, the number of local search space executions performed, and the time required for convergence are considered. All experiments are performed on the same dedicated machine.

In addition, some restrictions have to be made in the experimental setup. No delayed communication is considered. The message volume is not explicitly considered since all messages have the same size for the same number of agents and thus the number of messages is an equivalent indicator.

Fig. 1: Overview of used communication topologies



(a) complete graph (b) small world topology (c) mesh (d) random tree (e) ring (f) path graph

3.2 Algorithm

COHDA is a ‘combinatorial optimization heuristic for distributed agents’ and was developed for the self-organized scheduling of distributed energy resources (DER) in virtual power plants (VPP) (for a detailed discussion see [HS17]). Each agent is responsible for the scheduling of one energy resource and knows the possible schedules and local preferences of its asset. A schedule represents an operational option of the DER, i. e. how much energy is being supplied or withdrawn at what time. The search space contains all possible schedules and can be further limited by operational constraints of the plants [BS13]. COHDA has been proven to always converge at least to a local optimum [HS17].

To investigate the impact of the communication topology, other influences in the experimental setup must be kept under full control [NTS14]. In the application setting of smart grid scheduling problems, the agents’ search spaces consist of a set of feasible schedules, possibly extended by a decoder approach [HS17] [BS13]. As constructing such search spaces with defined properties regarding local minima for large scenarios is very difficult, an abstract and standard problem is chosen for analysis.

A suitable modified version of the algorithm is presented in [BL17]. Bremer et al. adapted COHDA to find the global minimum of a real valued objective function. An agent a_i is

responsible for only one value x_i from a continuous search space. It performs its local optimization to minimize the global objective function, by adapting its own choice of x_i while considering the choices of other agents $x_j, j \neq i$ as temporarily fixed. This approach offers the advantage of simplifying the optimization problem itself, but allowing to choose an objective function that fulfills specific criteria in order to investigate differences in solution quality, i.e. convergence to local minima.

The chosen objective function is a variation of Schwefel's function 2.26 (see Equation 1), since is possesses several essential characteristics [Sc81]. As intended, it is continuous and each agent has to pick one value in the in the range of [-500, 500]. It can be scaled arbitrarily, which allows experiments with any number of agents. Furthermore, it is a multi-modal function which means that it has a high number of ambiguous peaks in the function landscape, i.e. it has many local minima [JY13]. Figure 2 shows a three-dimensional plot of the function. The function is separable, so each variable of the function can be optimized independently from the other variables. This is not the case for most practical problems in the smart grid and also makes the problem easier to solve. To change this property, an additional penalty function has been introduced (see Equation 2).

$$Schwefel_{2.26} : f(\vec{x}) = 418.9829n - \sum_{i=1}^{n} x_i \sin\sqrt{|x_i|} \tag{1}$$

$$Penalty : p(\vec{x}) = \left| \sum_{i=1}^{\frac{n}{2}} x_{2i} - \sum_{i=1}^{\frac{n}{2}} x_{2i-1} \right| \tag{2}$$

This penalty function is zero if the sum of all values selected by agents with even indices and the sum of all values selected by agents with odd indices is equal. The penalty is added to the value computed by the objective function (Equation 1). The goal of the optimization is to minimize the resulting Equation 3.

$$f'(\vec{x}) = f(\vec{x}) + p(\vec{x}) \tag{3}$$

The global optimum remains the same as for the unchanged Schwefel function. $f'(\vec{x})$ is zero if all agents choose $x_i = 420,968746$ for $i = 1, ..., n$. The introduction of the penalty function adds dependencies between the variables, making the objective function inseparable and thus making it harder to find the global optimum. Figure 3 shows the function landscape in a three-dimensional search space. For local optimization of this objective function the Simplicial Homology Global Optimization (SHGO) of [ESF18] is used, since preliminary tests showed fast convergence and excellent performance. Please note that the local optimization function is exchangeable within the limits of the convergence criteria of the algorithm that are fulfilled with choosing SHGO [HS17]. With this setup, we design an examination scenario that can be kept under full control as the properties of the objective function, and thus of the global and local solution spaces, are arbitrarily configurable, and in this case, were designed to be susceptible to convergence into local minima.

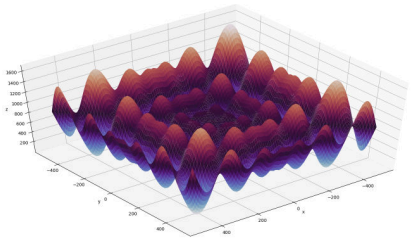


Fig. 2: Schwefel 2.26 for $n = 3$

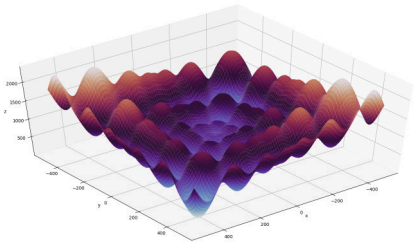


Fig. 3: Schwefel 2.26 with penalty for $n = 3$

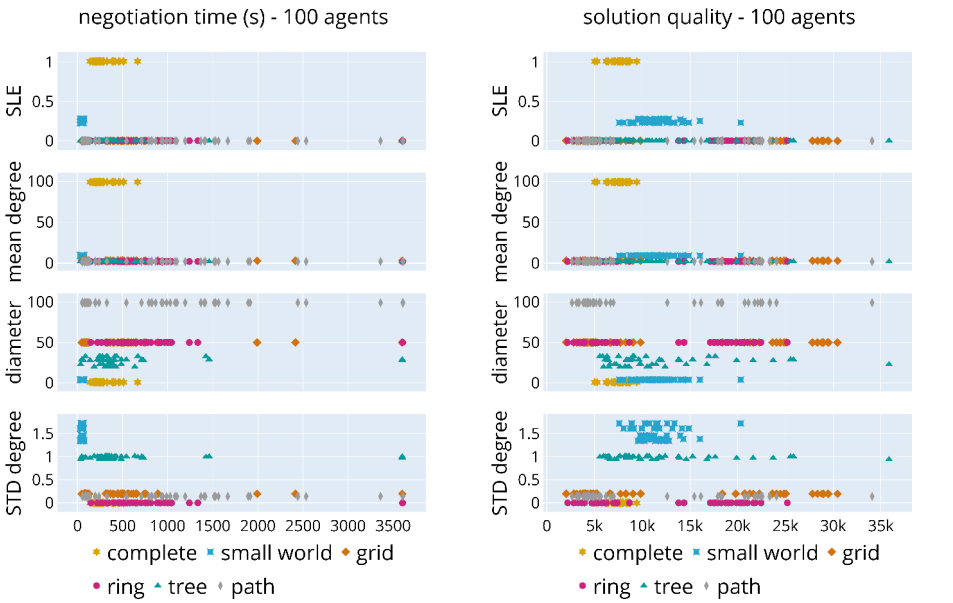
4 Results

To evaluate the previously established hypotheses, we will subsequently review the relevant data for each hypothesis. Thus, we first examine the effects of the relevant graph properties identified in [BH07] and [Ca16], namely SLE, diameter, mean degree and homogeneity of degree distribution, on convergence speed and solution quality (hypothesis 1 and 2). Afterwards we examine how the different topology types perform according to the different performance indicators, i.e. *solution quality*, *costs of collaboration (number of search space executions, number of messages)*, and *negotiation time* and investigate if the expected trade-off between the costs of collaboration and the solution quality exists.

As described in section 2, the communication topology of consensus algorithms has a significant impact on the speed of convergence. To show possible correlations, Figure 4a plots the *negotiation time* against each of the previously discussed graph properties for a MAS with 100 agents. It is apparent that a small diameter, a high mean degree and a large SLE have positive effects on the speed of convergence. The most prominent example for this is the complete graph. The topology with the second fastest convergence is the one with the second best values in these parameters, namely the small world topology. However, no correlation can be found for the STD of degree, since both heterogeneous and homogeneous topologies are associated with all grades of *negotiation time*. Therefore the assumption from hypothesis 1, that the mentioned graph properties also have a positive effect on the convergence speed of distributed algorithms, can be confirmed, with the exception of degree homogeneity.

Figure 4b shows the scatter plot for the four graph properties and the *solution quality*, again for a MAS with 100 agents. Since all graph property variations occur at each *solution quality* level, there appears to be no correlation. According to Talbi [Ta09], the characteristics of the topology make a difference in the degree of exploration and exploitation of the search space. This is reflected in the distribution of *solution quality*. The graph properties, which lead to fast convergence, also reliably result in a relatively good solution quality with low spread. Again, the complete graph and the small world topology are the most vivid examples for this. Ring, grid and tree graph, nearly reach the global optimum. These topologies have a medium diameter, low mean degree and also a rather low SLE. They differ in their degree

Fig. 4: Correlation of negotiation time and solution quality with graph properties - mean degree, standard deviation of the degree (STD degree) and the diameter.



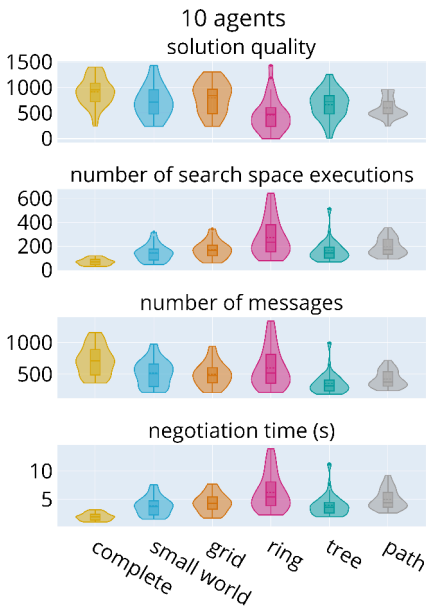
(a) Correlation of negotiation time with graph properties (b) Correlation of solution quality with graph properties

distribution as a ring is perfectly homogeneous and a tree one of the most heterogeneous topologies in the experiment. It can be assumed that this good *solution quality* is due to the fact that the negotiations coincidentally started at a favorable point and thus the topologies led to an extensive exploitation of a promising area of the search space.

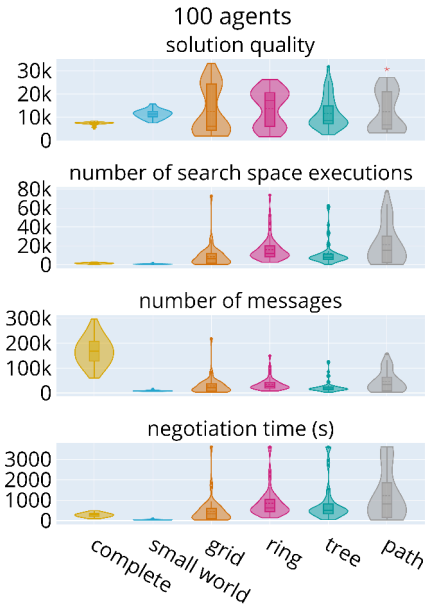
These results indicate that a high level of connectivity can lead to premature convergence and therefore poorer *solution quality*. Less connected topologies do not exhibit this problem and can therefore explore the search space more thoroughly. On the other hand, the exploitation is weaker and therefore they are highly sensitive to the starting point of the search. Hypothesis 2 states that the graph properties also affect the solution quality. This can be affirmed, although the graph properties do not directly impact how good or bad the quality of the solution is, but they do have an effect on how large the variation of the achieved quality and thus the reliability is.

The figures 5a and 5b show the performance of the topology types for a small MAS with 10 agents (5a) and a larger system with 100 agents (5b). These numbers are selected for presentation to show the typical behavior for smaller and larger MAS. Similar results were obtained in simulation runs with different numbers of agents. The figures show how the different topologies perform for all performance indicators and uses violin plots, which

Fig. 5: Extracts from the simulation results for 10 agents(left) and 100 agents (right) with 250 simulation runs per number of agents



(a) Results of the negotiation runs with 10 agents concerning the performance indicators as violin plots [HN98]



(b) Results of the negotiation runs with 100 agents concerning the performance indicators as violin plots [HN98]

show the distribution with an internal box plot surrounded by a density plot [HN98]. The depiction of the solution quality again underlines the results of hypothesis 2 and shows that the topology has an effect on the scattering of quality, especially with increasing system size. In relation to this, the diagram of the negotiation time represents the already explained correlation of the convergence speed with the same graph properties. The two middle rows contain the relevant parameters for collaboration costs (*number of messages*, *number of local search space executions*). The *number of search space executions* is not directly dependent on the communication topology here, since it is not triggered by the event of message arrival. Instead, each agent performs its tasks periodically. Yet it is still indirectly influenced through the total *negotiation time* and reflects the distribution of this indicator.

At first glance, the "*number of messages*"-indicator shows a similar distribution. However, regarding the MAS with 100 agents, it becomes clear that this similarity does not persist for the complete graph. The reason for this is the big difference in the graph properties (SLE, mean degree, degree distribution and diameter), which differentiates the complete graph more and more from the other topologies as the system size increases. It can thus be concluded that there is no clear trade-off between solution quality and collaboration costs

with regard to hypothesis 3. Instead, there is a trade-off between reliable behavior with relatively good solution quality but with potentially high message volumes and unreliable behavior with an increased chance of finding the global optimum. In summary, the small world topology represents the best compromise among the tested topologies, since it achieves relatively good results in a reliable and fast way and with low collaboration costs. If the costs for message transmission are negligible, the complete graph is preferable.

5 Conclusion

Distributed optimization heuristics performed by MAS are a suitable approach to solve various tasks in smart grids, such as the coordination of large numbers of distributed energy resources. Some of these heuristics are distributed at the algorithmic level, making the communication of the agents itself part of the algorithm. With COHDA as a representative of this class of algorithms, the experimental study has shown that the communication topology has an influence on both solution quality and efficiency of such heuristics. The simulation results demonstrated that certain graph properties, such as mean degree, diameter and Fiedler eigenvalue, affect the convergence speed of distributed heuristics as has been shown for consensus algorithms. We presented results that give more insight into the effect of different communication topologies on the degree of exploration and exploitation of the search space. The key findings are:

- Highly meshed topologies converge reliably and quickly into relatively good local optima, though potentially involving greater communication effort. Moreover, such topologies encounter difficulties in escaping local optima and thus finding the global optimum.
- Sparsely meshed topologies have a higher probability of finding the global optimum through higher exploitation. However, they can also diverge completely and only find solutions of low quality. Such cases consume large amounts of resources, in terms of computing power and message volume and can lead to excessive negotiation times.

The use of heuristics operating critical infrastructures leads to the requirement to reliably obtain sufficient results in a limited period of time. The above findings indicate that there is no one-fits-all topology, and thus dynamic topologies might be advisable to weigh exploration and exploitation in an optimal way. In future work, we therefore address the question of how the topology can be intelligently changed during the negotiation period in order to achieve reliable and fast convergence with improved and maybe even guaranteed solution quality, while keeping the costs of cooperation low.

Acknowledgements

This work has been funded by the German Federal Ministry for Economic Affairs and Energy (project number01ME18002B).

Bibliography

[AL12] Alcaraz, Cristina; Lopez, Javier: Analysis of requirements for critical control systems. International journal of critical infrastructure protection, 5(3-4):137–145, 2012.

[Az17] de Azevedo, Ricardo; Cintuglu, Mehmet Hazar; Ma, Tan; Mohammed, Osama A: Multiagent-based optimal microgrid control using fully distributed diffusion strategy. IEEE Transactions on Smart Grid, 8(4):1997–2008, 2017.

[Be13] Beck, Andreas; Derksen, Christian; Lehnhoff, Sebastian; Linnenberg, Tobias; Nieße, Astrid; Rohbogner, Gregor: Energiesysteme und das Paradigma des Agenten. In: Agentensysteme in der Automatisierungstechnik, pp. 21–42. Springer, 2013.

[BH07] Baras, John S; Hovareshti, Pedram: Effects of graph topology on performance of distributed algorithms for networked control and sensing. In: Proceedings of the Workshop on Networked Distributed Systems for Intelligent Sensing and Control, Kalamata, Greece,(http://med. ee. nd. edu/). 2007.

[BL17] Bremer, Joerg; Lehnhoff, Sebastian: An Agent-based Approach to Decentralized Global Optimization-Adapting COHDA to Coordinate Descent. In: International Conference on Agents and Artificial Intelligence. volume 2. SCITEPRESS, pp. 129–136, 2017.

[BS13] Bremer, Jörg; Sonnenschein, Michael: Model-based integration of constrained search spaces into distributed planning of active power provision. Computer Science and Information Systems, 10(4):1823–1854, 2013.

[Ca16] Cao, Hong-Tai; Gibson, Travis E; Mou, Shaoshuai; Liu, Yang-Yu: Impacts of network topology on the performance of a distributed algorithm solving linear equations. In: 2016 IEEE 55th Conference on Decision and Control (CDC). IEEE, pp. 1733–1738, 2016.

[Di18] Ding, Lei; Yin, George Yin; Zheng, Wei Xing; Han, Qing-Long et al.: Distributed energy management for smart grids with an event-triggered communication scheme. IEEE Transactions on Control Systems Technology, 27(5):1950–1961, 2018.

[Dö19] Dörfler, Florian; Bolognani, Saverio; Simpson-Porco, John W; Grammatico, Sergio: Distributed control and optimization for autonomous power grids. In: 2019 18th European Control Conference (ECC). IEEE, pp. 2436–2453, 2019.

[ESF18] Endres, Stefan C; Sandrock, Carl; Focke, Walter W: A simplicial homology algorithm for Lipschitz optimisation. Journal of Global Optimization, 72(2):181–217, 2018.

[HN98] Hintze, Jerry L; Nelson, Ray D: Violin plots: a box plot-density trace synergism. The American Statistician, 52(2):181–184, 1998.

[HS17] Hinrichs, Christian; Sonnenschein, Michael: A distributed combinatorial optimisation heuristic for the scheduling of energy resources represented by self-interested agents. IJBIC, 10(2):69–78, 2017.

[JY13] Jamil, Momin; Yang, Xin-She: A literature survey of benchmark functions for global optimization problems. arXiv preprint arXiv:1308.4008, 2013.

[Le14] Lewis, Frank L; Zhang, Hongwei; Hengster-Movric, Kristian; Das, Abhijit: Cooperative control of multi-agent systems: optimal and adaptive design approaches. Springer, 2014.

[LSK11] Logenthiran, Thillainathan; Srinivasan, Dipti; Khambadkone, Ashwin M: Multi-agent system for energy resource scheduling of integrated microgrids in a distributed system. Electric Power Systems Research, 81(1):138–148, 2011.

[Mo17] Molzahn, Daniel K; Dörfler, Florian; Sandberg, Henrik; Low, Steven H; Chakrabarti, Sambuddha; Baldick, Ross; Lavaei, Javad: A survey of distributed optimization and control algorithms for electric power systems. IEEE Transactions on Smart Grid, 8(6):2941–2962, 2017.

[Ni12] Nieße, Astrid; Lehnhoff, Sebastian; Tröschel, Martin; Uslar, Mathias; Wissing, Carsten; Appelrath, H-Jürgen; Sonnenschein, Michael: Market-based self-organized provision of active power and ancillary services: An agent-based approach for smart distribution grids. In: 2012 Complexity in Engineering (COMPENG). Proceedings. IEEE, pp. 1–5, 2012.

[NTS14] Nieße, Astrid; Tröschel, Martin; Sonnenschein, Michael: Designing dependable and sustainable Smart Grids–How to apply Algorithm Engineering to distributed control in power systems. Environmental Modelling & Software, 56:37–51, 2014.

[OSFM07] Olfati-Saber, Reza; Fax, J Alex; Murray, Richard M: Consensus and cooperation in networked multi-agent systems. Proceedings of the IEEE, 95(1):215–233, 2007.

[RBA05] Ren, Wei; Beard, Randal W; Atkins, Ella M: A survey of consensus problems in multi-agent coordination. In: Proceedings of the 2005, American Control Conference, 2005. IEEE, pp. 1859–1864, 2005.

[Ro15] Robitzky, Lena; Müller, Sven C; Dalhues, Stefan; Häger, Ulf; Rehtanz, Christian: Agent-based redispatch for real-time overload relief in electrical transmission systems. In: 2015 IEEE Power & Energy Society General Meeting. IEEE, pp. 1–5, 2015.

[RS16] Radhakrishnan, Bharat Menon; Srinivasan, Dipti: A multi-agent based distributed energy management scheme for smart grid applications. Energy, 103:192–204, 2016.

[Sc81] Schwefel, Hans-Paul: Numerical optimization of computer models. John Wiley & Sons, Inc., 1981.

[St01] Strogatz, Steven H: Exploring complex networks. nature, 410(6825):268–276, 2001.

[Ta09] Talbi, El-Ghazali: Metaheuristics: From Design to Implementation, volume 74. John Wiley & Sons, 2009.

R. Reussner, A. Koziolek, R. Heinrich (Hrsg.): INFORMATIK 2020,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2021 795

Vehicle Scheduling and refueling of Hydrogen Buses with On-site Electrolysis

Armin Golla[1] Frederik vom Scheidt[2] Nicole Röhrig[3] Philipp Staudt[4] Christof Weinhardt[5]

Abstract: In this paper, we present the first model for hydrogen-fueled bus fleet operation with on-site hydrogen production. To support the planning and operation of municipal hydrogen bus fleets, we transfer and adapt a vehicle scheduling algorithm for conventional buses. We optimize vehicle routes with regard to minimizing operational costs. Based on that, we implement a hydrogen refueling strategy suited to minimize the peak demand for a hydrogen refueling station with on-site electrolysis. To demonstrate the functionality of the approach, it is applied to a real-world scenario for public transport in Karlsruhe, Germany. The findings show that the scheduling strategy for vehicle routing and fueling is suited to reduce peak hydrogen demand and thus the required electrolyser capacity and associated costs. In comparison to a naive benchmark refueling scenario, the peak demand is reduced by 20%, lowering investment costs by 400,000 €.

Keywords: hydrogen mobility; transport scheduling; fuel station operation

1 Introduction

The worldwide efforts to mitigate climate change represent a specific challenge for the mobility sector. Germany has set the target of reducing mobility-related greenhouse gas emissions by at least 40% percent until 2030, compared to 1990 [BM19]. However, while the electricity sector is on the path to integrating renewable power sources, the share of mobility based on renewable resources remains low [Ge20]. A sustainable energy transition therefore would profit largely from using low emission electricity in the mobility sector. One promising avenue to use more of such renewable electricity in the mobility sector is hydrogen. Especially heavy-duty vehicles, including buses, trucks and trains, benefit from hydrogen's high gravimetric energy density and the ability for fast recharging. In first field tests, hydrogen buses for local public transport have proven well equipped to contribute to a sustainable city development [NO18]. A key prerequisite for the successful proliferation of hydrogen buses is an efficient operation of hydrogen supply and bus refueling. On-site hydrogen production from electrolysers can convert electricity, ideally with a high share of renewables, into sustainable hydrogen that can be used directly at the fuelling station. This also reduces the need for fuel transportation. With a large bus fleet and predictable

[1] Karlsruhe Institute of Technology, armin.golla@kit.edu
[2] Karlsruhe Institute of Technology, frederik.scheidt@kit.edu
[3] Karlsruhe Institute of Technology, nicole.roehrig@student.kit.edu
[4] Karlsruhe Institute of Technology, philipp.staudt@kit.edu
[5] Karlsruhe Institute of Technology, weinhardt@kit.edu

doi:10.18420/inf2020_70

operation schedules, public transport companies are especially suited for the implementation of hydrogen buses and an according refueling strategy. It has already been shown that renewable hydrogen can be produced cost competitively in certain locations in Germany [GR19]. However, this finding neglects the demand side. It is assumed that hydrogen will be used at the time and location, where it is produced. To complement this picture, we examine the operation of a hydrogen bus fleet in the municipal transport sector in an integrated approach, thus linking local production and local demand. To this end, we address two research questions:

(1) How can hydrogen-based bus transport with on-site electrolysis be modelled?

(2) How much electrolysis capacity and investment costs can be saved by applying this model in a real world scenario of a municipal fleet with 24 buses?

To answer these questions, we structure this paper as follows: First, we model a vehicle-scheduling problem to optimize the driving distances within each route and to determine the allocation of routes to the hydrogen buses. We solve this problem using flow optimization and set partitioning. The data required for this step includes departure and arrival times as well as the trip lengths, speed limits along the route and the location of the refueling depot. The resulting routes are modelled as time-space-graphs as proposed in [GKS05; KMS06]. In the second stage, we use actual bus schedules of a local public transport provider in the German city of Karlsruhe to demonstrate the applicability of the model. We use our proposed approach to model the scheduling of the bus fleet that comprises all bus lines starting or ending at a distance of three kilometers around the town center. The actual driving distances, as well as the distance to the refueling depot are calculated using the 'OpenStreetMap' software [HW08]. The dimensioning of the electrolyser is one of the peculiarities to consider for the on-site production of hydrogen for a hydrogen bus fleet. With investment costs of 1100 € per kWp [Bö20], electrolysers are one of the cost drivers for hydrogen fuel stations. For a sub-sample of 24 buses on eleven bus lines, we investigate a vehicle scheduling algorithm with regard to a minimized electrolyser capacity and this reduced investment costs. Our results show that an optimal scheduling and refueling of hydrogen buses can reduce the required peak hydrogen production capacity by 20% compared to a naive strategy with uncontrolled refueling. In the example scenario, this reduces the investment costs for the electrolyser by 400,000 €.

2 Related Work

Our work builds on two major strands of literature – the operation of hydrogen refueling stations and the optimization of vehicle scheduling. In this subsection, we briefly review this literature. A hydrogen refueling station generally consists of four system components: The hydrogen production, storage, supply and conditioning. Production of hydrogen can be done either off-site, or on-site. In the case of on-site production the costs for hydrogen delivery can be omitted. If hydrogen is produced via water electrolysis it can be emission-free if

electricity from renewable energy sources is used. The two predominant storage technologies for hydrogen are liquid hydrogen (at cryogenic temperatures of approximately -253 °C and low pressure) or compressed gaseous hydrogen (at ambient temperature and high pressure) [Fi17]. In contrast to liquid fuels, hydrogen cannot simply be pumped into the tank, but is either gradually released into the vehicle tank through pressure compensation or a booster-compressor [SVS13]. The investment costs of the chosen electrolyser, storage technology and filling equipment typically make up a considerable share of total hydrogen costs [Al16; Gr17]. Recently, first studies on hydrogen fuel stations have analysed siting, sizing and operation. Sun et al. [Su19] optimize the location and size of hydrogen refueling stations in Chengdu, China. However, the model does not consider the station operation and the temporal aspects of refueling, and instead just covers the demand of one year. Similarly, Yang and Jiang [YJ20] optimize the location and capacity of hydrogen refueling stations, considering uncertainty of long-run hydrogen demand, but no temporal resolution of refueling events. Focusing on hydrogen-powered freight-trucks, Rose and Neumann [RN20] optimize the location of refueling stations with on-site electrolysis using grid electricity. They analyse a case study for German highways in 2050, utilizing truck driving data at hourly resolution. To our knowledge, all existing hydrogen fuel station studies take hydrogen demand patterns as given and try to adequately design hydrogen supply. Moreover, no study has yet specifically analysed hydrogen refueling stations for public transport. This use case poses particular requirements, as station operation and driving schedules need to be aligned. Vehicle scheduling describes the task of optimally distributing a specified number of given trips to the vehicles of a fleet [BK09]. It plays a central role in operational planning in public transport. The desired result of a vehicle scheduling optimization problem is a schedule that contains all tours of all vehicles. Each tour is a number of trips that start and end at a depot. In the case of public transport, the depot typically includes the infrastructure for refueling. An overview of vehicle scheduling problems is given in [BK09]. An overview of optimal vehicle scheduling for public transport can be found in [KMS06]. Vehicle scheduling problems in public transport can be categorized based on characteristic properties of their problem instances. A first distinction is made with regard to the number of depots within the transport network under consideration, namely between single depot and multi depot setups. While multi depot setups represent NP hard problems, single depot models can be solved in polynomial time [KMS06]. The system in our case study relies on one central single depot. Furthermore, mathematically, vehicle scheduling can include flow models and assignment problems [BK09]. Compared to conventional bus fleets, the operation of hydrogen-fueled fleets with on-site electrolysis requires consideration of additional constraints. In contrast to conventional refueling stations, hydrogen refueling stations can only serve a smaller number of refueling events in a given period. More importantly, conventional fuel stations can rely on regularly filled fuel storage, whereas hydrogen fuel stations with on-site production rely on the production capacity of the electrolyser. Unscheduled refueling requests from buses at the depot cause variable demand, whereas the on-site hydrogen production rate is preferably constant to achieve the highest efficiency of the electrolyser [HK18]. In addition, investment costs are an essential cost factor for hydrogen mobility. Since the electrolyser has to be sized to satisfy maximum demand from buses, the refueling schedule directly impacts the total

		Index	
b	Bus index	B	Set of buses in the fleet
i, j, k	Node indices	V	Set of nodes in the network
e_{ij}	Edge of a network	E	Set of edges in the network
t	Time step index	T	Time horizon
		Variables	
c_{H2}	Hydrogen fuel costs	Q	Cost factor for waiting during a trip
C_{ij}	Weight of edge e_{ij}	T^f	Hydrogen refueling time
D_{ij}	Length of edge e_{ij}	$\overline{V}^e$	Average empty driving velocity
$E^t, E^w,$	Amount of trip, waiting	W_{ij}	Waiting duration on edge e_{ij}
E^d	and depot edges	x_{ij}	Flow on edge e_{ij}
E^s	Number of scheduled trips	y_{ijb}	Assignment of x_{ij} to the trip of bus b
Fij	Trip frequency	z	Electrolyser capacity
h_{bt}	Hydrogen refueling of b in t	α_{bt}	Binary variable, indicating
H^d	H_2 consumption of a bus		if bus b is in the depot at t
	per kilometer	β_{bt}	Binary variable, indicating
H^{max}	Maximum H_2 refuel capacity		if bus b is refueled at t

Tab. 1: Nomenclature

costs of the infrastructure. The refueling of vehicles with hydrogen must therefore already be taken into account in the route allocation and deployment planning of the vehicles. This motivates to schedule vehicles in a way that distributes refueling events as uniformly as possible, while meeting all mobility needs. In the following section, we develop a specific vehicle scheduling model for hydrogen buses that takes these unique aspects into account.

3 Methodology

In this section, we describe the procedure for modelling the vehicle scheduling problem of a hydrogen fueled bus fleet for public transport in combination with on-site electrolysis.

3.1 Time-Space Modeling

To assess the hydrogen demand and refueling frequency of a hydrogen bus fleet, we determine the bus driving distances and trip durations for each bus line. The round trip modeling for the entire bus fleet is based on the time-space modeling methodology presented by Kliewer et al. [KMS06]: First, a time-space network is set up for the trips within the transport network. The network serves as input for a two-step solution of flow optimization and subsequent flow decomposition. The two-step approach allows the determination of optimal routes and optimal assignment of trips along the routes to the vehicles of the fleet. Figure 1 shows a simple example of a time-space graph with one depot ('D') and two

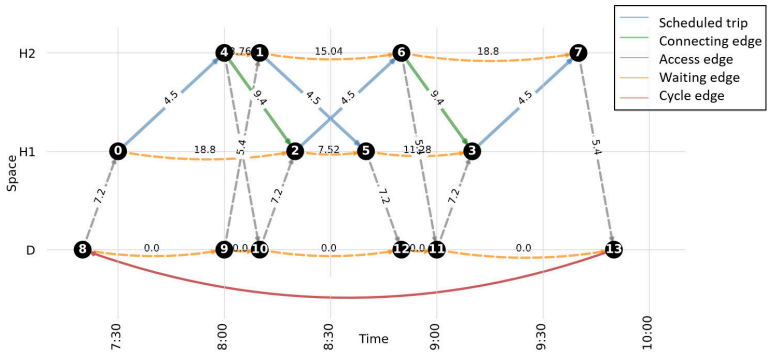


Abb. 1: A simple time-space modeling example with one depot and two stops

stops ('H1' and 'H2'). On the basis of the time-space network, an optimal trip schedule is determined. For this purpose, a flow problem is set up [6]. The goal of flow optimization is to determine the flow x_{ij} for each edge e_{ij} of the network in such a way that all planned timetable trips are completely executed while minimizing the operational costs, consisting of waiting costs and fuel costs. The costs of all edges are given by:

$$C_{ij} = \begin{cases} D_{ij} H^d c_{H2} & \forall e_{ij} \in E^t \\ W_{ij} Q & \forall e_{ij} \in E^w \\ 0 & \forall e_{ij} \in E^d \end{cases} \tag{1}$$

The flow determination is given by:

$$\min \quad \sum_{(i,j)|e_{ij} \in E} C_{ij} x_{ij} \tag{2}$$

$$s.t. \quad \sum_{j|e_{ij} \in E} x_{ij} - \sum_{k|e_{ki} \in E} x_{ki} = 0 \qquad \forall\, i \in V \tag{3}$$

$$x_{ij} = F_{ij} \qquad \forall\, (i,j)|e_{ij} \in E^s \tag{4}$$

$$x_{ij} \in \mathbb{N} \tag{5}$$

The objective function is the minimization of the sum of the costs of all flows. The conservation of flows at each node is ensured by Equation (3). The flow conservation guarantees that in a solution the sum of all incoming flows equals the sum of all outgoing flows at each node. In addition, the determined flow on the trailing edge corresponds to the number of vehicles used in the optimal solution, as all vehicles drive back to the depot at the end of the modelled period and reach the beginning of the next period via a circulation edge [KMS06]. The period in this context is the time horizon over which the scheduling is modeled. Equations (4) and (5) ensure that the flow solution contains all planned timetable trips and has a non-negative integer value.

[6] For basics of flow modeling we refer to [Ta07].

3.2 Operation and Scheduling of a Hydrogen Bus Fleet

For a complete solution of the vehicle scheduling problem, the flow units of the network edges must be assigned to the individual vehicles of the fleet. Additionally, the refueling stops of the hydrogen buses are scheduled. The process is called flow decomposition [KMS06]. In this step, we integrate the specific requirements of hydrogen bus fleets. The hydrogen refueling demand for every bus in the fleet and the production quantity of the electrolyser are added to model the refueling process. To reduce the investment costs for the electrolyser, the objective of the scheduling strategy is to minimize the peak hydrogen consumption while satisfying the result of the flow calculations modeled in Section 3.1:

$$\min z \tag{6}$$

$$s.t. \sum_{j|e_{ij}\in E} y_{ijb} - \sum_{k|e_{ki}\in E} y_{kib} = 0 \qquad \forall\,(i,b) \in V \times B \tag{7}$$

$$\sum_{b\in B} y_{ijb} = x_{ij} \qquad \forall\,(i,j)|e_{ij} \in E \tag{8}$$

The production capacity of the electrolyser within 24 hours for t in one-minute resolution is given by:

$$\sum_{t-1440}^{t} \sum_{b\in B} h_{bt} \leq z \;\forall\, t \in T \tag{9}$$

The minimum and maximum refueling capacity for each bus per time step is given by:

$$H^{min} \leq h_{bt} \leq H^{max} \qquad \forall\,(t,b) \in T \times B|\beta_{bt} = 1 \tag{10}$$

$$h_{bt} = 0 \qquad \forall\,(t,b) \in T \times B|\beta_{bt} = 0 \tag{11}$$

To ensure that a bus can only refuel when it is parked in the depot, α_{bt} indicates whether the bus is in the depot at t (start of refueling) as well as at $t + T^f$ (end of refueling). The variable α_{bt} is set to one if the edge e_{ij} over which bus b moves at time t belongs to the set of depot waiting edges E^d. Otherwise, α_{bt} is set to zero:

$$\alpha_{bt} \geq \beta_{bt} \qquad \forall\,(t,b) \in T \times B \tag{12}$$

$$\alpha_{b(t+T^f)} \geq \beta_{bt} \qquad \forall\,(t,b) \in T \times B \tag{13}$$

$$\alpha_{bt} = y_{ijb} \qquad \forall\, b \in B,\ \forall\,((i,j)|t_i \leq t \leq t_j, e_{ij} \in E^d) \tag{14}$$

$$\alpha_{bt} = 0 \qquad \forall\, b \in B,\ \forall\,((i,j)|t_i \leq t \leq t_j, e_{ij} \in E \setminus E^d) \tag{15}$$

The hydrogen level of each bus is in the interval $[0, H_{max}]$ as described in Equation (16). The filling level is the sum of the initial filling level h_{bt0}, plus the hydrogen filled up to this point in time, minus the hydrogen consumed up to time step t as described in Equation (17):

$$h_{bt_0} + \sum_{t \in T \mid t > t_0} h_{bt} - \sum_{(i,j) \mid t_j > t_0, e_{ij} \in E} y_{ijb} D_{ij} H^d \geq 0 \qquad \forall\, b \in B \qquad (16)$$

$$h_{bt_0} + \sum_{t \in T \mid t > t_0} h_{bt} - \sum_{(i,j) \mid t_j > t_0, e_{ij} \in E} y_{ijb} D_{ij} H^d \leq H^{max} \qquad \forall\, b \in B \qquad (17)$$

$$h_{bt_0}, h_{bt} \geq 0 \qquad (18)$$

The following variables are binary:

$$x_{ij}, y_{ijb}, \beta_{bt}, \alpha_{bt} \in \{0, 1\} \qquad (19)$$

4 A Hydrogen Bus Fleet for Karlsruhe

To demonstrate the functionality of the proposed model, we investigate the operation of a hydrogen bus fleet in Karlsruhe, Germany. For this, we obtain data from three sources: Scheduled bus timetables from the 'General Transit Feed Specifications' archive of the municipal transport authority [Ka20a], distances for bus trips along the scheduled routes from the 'TRIAS' interface for electronic timetable information [Ka20b][7] and bus driving distances as well as the distance to the refueling depot that are calculated using the 'OpenStreetMap' software [HW08]. As the bus schedule is repeated weekly, we investigate the operation over a one-week period.

4.1 Implementation

We examine a section of the Karlsruhe travel authority's network in a radius of three kilometers around the Karlsruhe city center. For this purpose, the trips of the lines starting or ending in the respective radius are selected from the data of all lines. In a second step, we use a subsample of eleven bus lines that are served by 24 buses starting or ending in a 3 km radius around the city center to demonstrate the operation and scheduling of the bus fleet as proposed in 3.2. An overview of the number of lines and timetabled trips within the 3 km radius and the subsample is given in Table 2. For the specification of the hydrogen buses and the refueling process, we assume a maximum fuel capacity H^{max} of 40kg [HK18], a refueling time T^f of 10 minutes [SVS13], a fuel consumption H^d of $0.5 \cdot 10^{-2} kg/km$ [HK18] and an average driving velocity $\overline{V}^e$ of $50 km/h$ for the trips back to the depot.

[7] We would like to thank the Karlsruhe travel authority (KVV) for the provision of the data regarding bus schedules and distances. The KVV is neither responsible, nor liable for the content of this paper.

Radius	Number of bus lines	Bus routes per week	Required number of buses
3 km	58	14,447	74
Sample	11	4,186	24

Tab. 2: Flow optimization for the bus transport network

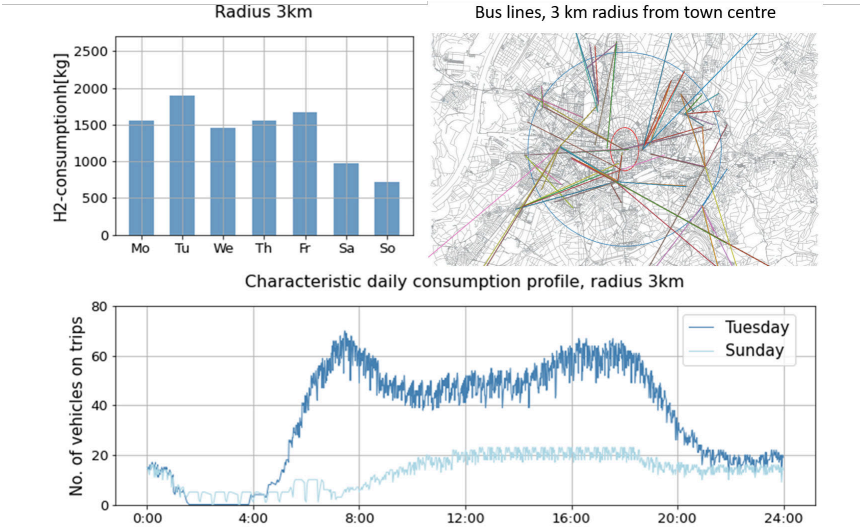


Abb. 2: Trip profiles and hydrogen consumption for a fleet of 24 buses

4.2 Results

In the given schedules of the sample of eleven bus lines, the number of required trips varies both within days and in-between days. For example, there are more bus trips on weekdays than on weekends. Besides, there are more trips during main commuting times and fewer trips at night. Exemplary trip profiles for Tuesday and Sunday are displayed in Figure 2. Uncontrolled refueling, i.e. refueling all buses upon return to the depot at the end of the day, leads to high peaks in hydrogen demand. With no optimized vehicle scheduling, this results in a maximum daily consumption of 639 kg of hydrogen, which causes the need of an electrolyser with a hydrogen production capacity of 1.5 MW. The optimized schedule aims to reduce such peaks. The flow decomposition returns the operation schedule for each of the 24 buses that are used on the eleven bus lines. The resulting trip distance and refueling events of an exemplary bus are depicted in Figure 3. The upper graph shows the distance travelled over time. Horizontal curve sections indicate the time spent waiting on an edge or in the depot. The refueling events for the bus determined in the optimisation are displayed as red vertical lines. The lower diagram in Figure 3 shows the bus' hydrogen storage level over time. For the entire fleet, the individual trip distances and depot waiting times are displayed

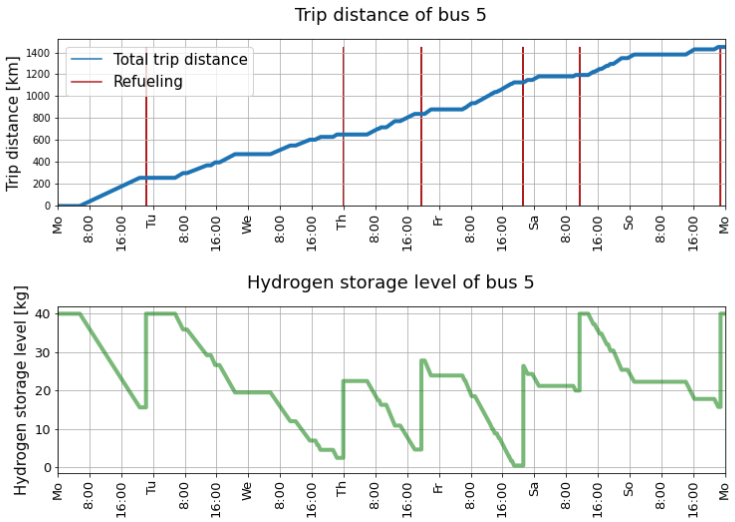


Abb. 3: Trip distance and refueling of a single bus in the fleet

in Figure 4. As the optimization is focused on an even operation of the bus depot fuel station and not on an even workload of the bus fleet, the workload distribution between all buses varies greatly. With the optimization presented in this paper, the maximum consumption within a 24 hour period is 510 kg hydrogen. This is 20% below the benchmark (639 kg). Assuming an electricity consumption of 58 kWh per kg of hydrogen produced by electrolysis [Re17] and 24 hours of production, an electrolyser with a capacity of 1.2 MW is required to produce 510 kg of hydrogen per day. For a maximum daily demand of 639 kg in the benchmark scenario, a 1.5 MW electrolyser is required. With investment costs of 1100 € per kW installed capacity for electrolysis [Bö20], the investment costs are 1,700,000 € in the benchmark scenario and 1,300,000 € in our optimized scenario. For the bus fleet with 24 vehicles, the investment costs for the purchase of an electrolyser can therefore be reduced by 400,000 €.

5 Discussion

The paper describes how a two-step approach of flow optimization and flow decomposition can be used to adapt the scheduling of hydrogen-powered public transport to the requirements of hydrogen supply by on-site electrolysis. In the first step, the flow optimization is described as a procedure that allows an estimation of the daily hydrogen consumption for fleets of different sizes. In the present work, a fleet of 24 buses is considered. The procedure is considered suitable for the application to larger fleets. In the second step, a problem formulation for flow decomposition is set up and the optimization with respect to smoothing

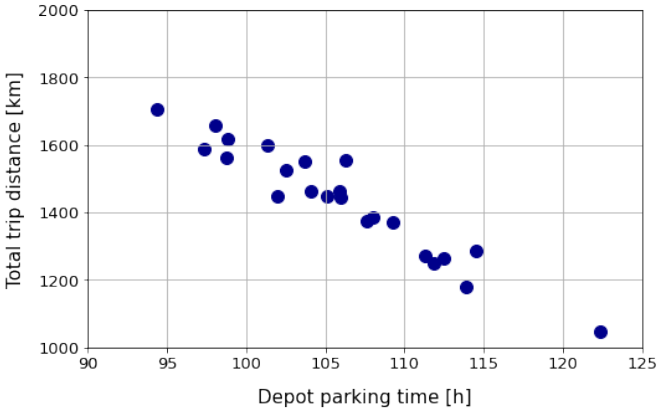


Abb. 4: Depot parking duration and traveling distances for each bus

the load at the filling station is performed. The results show that the required electrolysis capacity and thus the investment costs for the electrolyser can be reduced by 20% compared to a naive strategy. It is possible to analyse, if additional cost reductions can be achieved if the electrolyser is operated dynamically, making use of cheaper electricity at certain times. This could be done via time-of-use tariffs [Gu19] or, in combination with appropriate price forecasting [Sc20], via real-time-pricing tariffs. Further applications of on-site electrolysis could be in the context of local energy communities as a method to utilize locally generated renewable electricity [GHS20]. On-site electrolysis could furthermore be combined with centralized electrolysis and hydrogen delivery to fuel stations [Sc21]. For the decision to transit to hydrogen-based public transport, investment costs for hydrogen buses need to be considered as well. Currently in the range of 500,000 € to 625,000 € per bus, the investment costs are expected to drop to 400,000 € by 2030 [Be15].

6 Conclusion

The vehicle scheduling and operation strategy for a hydrogen bus fleet with on-site electrolysis proposed in this paper can support planning and investment decisions in public transport. We demonstrate how the approach can be used to reduce the required capacity of an electrolyser, lowering investment costs by 20% for an exemplary application in the German city of Karlsruhe. With this paper, we aim to contribute to reducing carbon emissions in the mobility sector by promoting the deployment of hydrogen-fueled public transport bus fleets and to enable locally emission free public transport.

Literature

[Al16] Albrecht, U. e. a.: Kommerzialisierung der Wasserstofftechnologie in Baden-Württemberg - Rahmenbedingungen und Perspektiven, Retrieved from https://www.e-mobilbw.de/fileadmin/media/e-mobilbw/Publikationen/Studien/Studie_H2-Kommerzialisierung_Neu_RZ_WebPDF.pdf, accessed 14.08.2020, 2016.

[Be15] Berger, R.: Fuel Cell Electric Buses – Potential for Sustainable Public Transport in Europe, Munich, 2015.

[BK09] Bunte, S.; Kliewer, N.: An overview on vehicle scheduling models. Public Transp/, S. 299–317, 2009.

[BM19] BMU: Klimaschutzprogramme 2030, Retrieved from https://www.bmu.de/fileadmin/Daten_BMU/Download_PDF/Klimaschutz/klimaschutzprogramm_2030_umsetzung_klimaschutzplan.pdf, accessed 22.03.2020, 2019.

[Bö20] Böhm, H.; Zauner, A.; Rosenfeld, D. C.; Tichler, R.: Projecting cost development for future large-scale power-to-gas implementations by scaling effects. Applied Energy 264/, S. 114780, 2020, ISSN: 03062619.

[Fi17] Fischedick, M. e. a.: Energie der Zukunft? Nachhaltige Mobilität durch Brennstoffzelle und H2, Techn. Ber., Wuppertal: Wuppertal Institut für Klima, Umwelt, Energie, 2017.

[Ge20] German Renewable Energies Agency: Share of renewable energies in electricity, heat and transport in Germany 1990-2018, Retrieved from https://https://www.unendlich-viel-energie.de/media-library/charts-and-data, accessed 14.08.2020, 2020.

[GHS20] Golla, A.; Henni, S.; Staudt, P.: Scaling the Concept of Citizen Energy Communities through a Platform-based Decision Support System. European Conference on Information Systems (ECIS)/, 2020.

[GKS05] Gintner, V.; Kliewer, N.; Suhl, L.: Solving large multiple-depot multiple-vehicle-type bus scheduling problems in practice. OR Spectrum 27/4, S. 507–523, 2005, ISSN: 0171-6468.

[Gr17] Gruger, F.: Elektrolyse an Wasserstofftankstellen - eine geeignete Anwendung von Power to Gas, Rainer Lemoine Institut, Jahreskonferenz Power to Gas, accessed 14.08.2020, 2017.

[GR19] Glenk, G.; Reichelstein, S.: Economics of converting renewable power to hydrogen. Nature Energy 2/, S. 23, 2019.

[Gu19] Guerra, O. J.; Eichman, J.; Kurtz, J.; Hodge, B.-M.: Cost Competitiveness of Electrolytic Hydrogen. Joule 3/10, S. 2425–2443, 2019, ISSN: 2542-4351.

[HK18] Hof, E.; Kupferschmid, S.: Einführung von Wasserstoffbussen im ÖPNV. Fahrzeuge, Infrastruktur und betriebliche Aspekte, 2018, URL: https://www.xn--starterset-elektromobilitt-4hc.de/content/1-Bausteine/5-OEPNV/now_leitfaden_einfuehrung-wasserstoffbussse.pdf.

[HW08] Haklay, M.; Weber, P.: OpenStreetMap: User-Generated Street Maps. IEEE Pervasive Computing 7/4, S. 12–18, 2008, ISSN: 1536-1268.

[Ka20a] Karlsruhe Transport Authority: General Transit Feed Specification, 2020, URL: https://projekte.xn--kvvefa-dg0c.de/GTFS/google_transit.zip.

[Ka20b] Karlsruhe Transport Authority: Trias-Interface, 2020, URL: https://www.vdv.de/oepnv%E2%80%90datenmodell.aspx.

[KMS06] Kliewer, N.; Mellouli, T.; Suhl, L.: A time–space network based exact optimization model for multi-depot bus scheduling. European Journal of Operational Research 175/3, S. 1616–1627, 2006, ISSN: 03772217.

[NO18] NOW GmbH: Einführung von Wasserstoff-bussen im ÖPNV, Retrieved from https://www.starterset-elektromobilität.de/content/1-Bausteine/5-OEPNV/now_leitfaden_einfuehrung-wasserstoffbussse.pdf, accessed 14.08.2020, 2018.

[Re17] Reuter, B.; Faltenbacher, M.; Schuller, O.; Whitehouse, N.; Whitehouse, S.: New Bus ReFuelling for European Hydrogen Bus Depots: Guidance Document on New Bus ReFuelling for European Hydrogen Bus Depots. 2017.

[RN20] Rose, P. K.; Neumann, F.: Hydrogen refueling station networks for heavy-duty vehicles in future power systems. Transportation Research Part D: Transport and Environment 83/, S. 102358, 2020, ISSN: 1361-9209.

[Sc20] vom Scheidt, F.; Medinová, H.; Ludwig, N.; Richter, B.; Staudt, P.; Weinhardt, C.: Data analytics in the electricity sector – A quantitative and qualitative literature review. Energy and AI 1/, S. 100009, 2020, ISSN: 2666-5468.

[Sc21] vom Scheidt, F.; Qu, J.; Staudt, P.; Mallapragada, D.; Weinhardt, C.: The effects of electricity tariffs on cost-minimal hydrogen supply chains and their impact on electricity prices and redispatch costs. Proceedings of HICSS54/, 2021.

[Su19] Sun, H.; He, C.; Yu, X.; Wu, M.; Ling, Y.: Optimal siting and sizing of hydrogen refueling stations considering distributed hydrogen production and cost reduction for regional consumers. International Journal of Energy Research 43/9, S. 4184–4200, 2019.

[SVS13] Smolinka, T.; Vogelstätter, C.; Sicha, E.: Wasserstoff–Infrastruktur für eine nachhaltige Mobilität, hrsg. von e-mobil BW GmbH, Fraunhofer ISE, 2013.

[Ta07] Taha, H. A.: Operations research: An introduction. Pearson, Upper Saddle River, N.J., 2007, ISBN: 0-13-188923-0.

[YJ20] Yang, G.; Jiang, Y.: Siting and sizing of the hydrogen refueling stations with on-site water electrolysis hydrogen production based on robust regret. International Journal of Energy Research 44/11, S. 8340–8361, 2020.

Innovativer Informatikunterricht – Theorie und Praxis

Innovativer Informatikunterricht – Theorie und Praxis

Maximilian Marowsky[1] Anna Fehrenbach[2] Paul Ohm[3]

Im Zuge der Digitalisierung gewinnt insbesondere das Schulfach der Informatik zunehmend an Bedeutung. Um an aktuellen Debatten teilhaben zu können und Deutschland als IT-Wirtschaftsstandort zu stärken, erscheint eine umfassende Informatik-Ausbildung an Schulen obligatorisch. Da der Bedarf an IT-Fachkräften durch die Beschleunigung der Digitalisierung enorm ansteigt, die wenigsten Schülerinnen und Schüler (SuS) jedoch programmieren lernen, blieben im Jahr 2019 in Deutschland rund 124.000 IT-Stellen unbesetzt [BR19]. Vor diesem Hintergrund entstand PearUp, eine digitale Lernplattform, die Lehrkräfte dabei unterstützt, qualitativ hochwertigen Informatikunterricht anbieten zu können und SuS spielerisch die Grundlagen der Informatik zu vermitteln.

Im Workshop werden aktuelle Forschungsfelder wie Digital Game-based Learning, Learning Analytics und Digital Mastery Learning vorgestellt und erlebbar gemacht. Die Teilnehmenden des Workshops bekommen so eine theoretische und praktische Einführung innovativer Ansätze, die durch den Einsatz digitaler Lernmedien möglich werden. Zum Einsatz kommt dabei die Lernplattform PearUp, die Lehrkräfte bei der Unterrichtsdurchführung unterstützen und SuS für das Fach der Informatik begeistern möchte. Die Teilnehmenden schlüpfen im Laufe des Workshops sowohl in die Rolle der Lehrkraft als auch in die der SuS, um eigene Erfahrungen mit DGBL zu sammeln.

Als SuS gründen die Teilnehmenden im Rahmen einer Spielhandlung ein virtuelles IT-Startup, das sich auf das Lösen von Informatikaufgaben spezialisiert hat. Anhand realitätsnaher Prgrammieraufträge lernen sie die Grundlagen des Programmierens kennen und erleben zugleich, wie aufregend und vielseitig die Arbeitswelt der Informatik sein kann. Erledigte Aufträge generieren virtuelle Einnahmen wie Geld, Erfahrungspunkte und Auszeichnungen. Mit steigender Kompetenz und Stufe erhöht sich die Anzahl der verfügbaren Aufträge sowie die Möglichkeiten das eigene Startup auszubauen.

Als Lehrkraft lernen die Teilnehmenden über die Plattform Hilfestellung zu leisten, Aufgaben zu bewerten und den Überblick über alle Klassen und einzelne SuS zu wahren.

[1] Universität Osnabrück, mmarowsky@uos.de
[2] Universität Osnabrück, afehrenbach@uos.de
[3] Universität Osnabrück, pohm@uos.de

doi:10.18420/inf2020_71

Literaturverzeichnis

[BR19] Bitcom Research. (2019, 28. November). Erstmals mehr als 100.000 unbesetzte Stellen für IT-Experten [Pressemeldung]. Abgerufen von https://www.bitkom-research.de/de/pressemitteilung/erstmals-mehr-als-100000-unbesetzte-stellen-fuer-it-experten

[Ma19] Marowsky, M. (2019). Auswirkungen einer spielebasierten E-Learning Anwendung für den Informatikunterricht auf die Motivation und den Lernerfolg der Schüler. Masterarbeit.

Recht und Technik – Datenschutz im Diskurs

Recht und Technik – Datenschutz im Diskurs

Rüdiger Grimm[1], Gerrit Hornung[2], Christoph Sorge[3], Indra Spiecker genannt Döhmann[4]

Vorwort zu den Workshopbeiträgen

Der Workshop „Recht und Technik – Datenschutz im Diskurs“ ist nach sieben Jahren mittlerweile ein fester Bestandteil der Jahrestagung der Gesellschaft für Informatik. Seit dem letzten Jahr haben wir den Kreis der Veranstalter/innen um den Kollegen Gerrit Hornung erweitert, worüber wir uns sehr freuen.

Unverändert bleibt aber die Ausrichtung: Wir bieten auch in diesem Jahr ein Forum für Beiträge von Informatiker/innen und Juristen/innen, die an Fragestellungen des technikbasierten Datenschutzes arbeiten. Wie schon in den Vorjahren werden Themen adressiert, die anwendungsorientiertes Potential für interdisziplinären Diskurs und Zusammenarbeit bieten und die Möglichkeiten aufzeigen, wie Datenschutz durch Technik präzisiert und umgesetzt werden kann.

Der Workshop setzt an den theoretischen und praktischen Aspekten des Schutzes von Daten und Privatheit und der europäischen Datenschutz-Grundverordnung an. Besondere Herausforderungen sehen wir in Fragen zu

- Modellierungen datenschutzkonformer Technikgestaltung
- Entwicklungen im Privacy by Design und Privacy by Default
- Risiken durch und Lösungsansätze mithilfe von Methoden der künstlichen Intelligenz (Machine Learning, Big Data)
- Privacy-Risikoanalysen
- Deepfakes
- Datenschutzgerechte Nutzung von Mobilitätsdaten, z.B. im öffentlichen Nahverkehr

Die Beiträge, die wir aus der Vielzahl an qualitätsvollen Einreichungen in einem *peer-reviewed*-Verfahren mit Unterstützung unserer Gutachter/innen auswählen konnten, zeigen

[1] Fraunhofer SIT Darmstadt und Universität Koblenz-Landau, grimm@uni-koblenz.de
[2] Universität Kassel, gerrit.hornung@uni-kassel.de
[3] Universität des Saarlandes, christoph.sorge@uni-saarland.de
[4] Goethe-Universität Frankfurt am Main, spiecker@jur.uni-frankfurt.de

doi:10.18420/inf2020_72

die thematische und disziplinären Bandbreite der derzeit im Spannungsfeld von Recht und Technik diskutierten Themen; die Einreichungen gingen darüber noch hinaus. Die zunehmende Regulierung der Digitalisierung, die Konkretisierung der DSGVO-Anforderungen und ein insgesamt gestiegenes Bewusstsein dafür, dass nicht alles technisch Machbare auch gesellschaftlich wünschenswert ist, spiegeln sich in den Beiträgen wieder. Neben sehr konkreten Vorschlägen zur Bewältigung von Einzelproblemen sind auch Beiträge mit übergreifenden Einsichten repräsentiert.

Gerade die Entwicklung der Corona-App hat zudem gezeigt, dass gesellschaftliches Bewusstsein für den Datenschutz sehr wohl vorhanden ist und sich in erheblicher Einflussnahme auf den politischen Prozess niederschlägt. Das Vorgehen illustriert zudem, dass aus den Erkenntnissen von Recht und Technik sehr wohl konstruktive technische Lösungsmöglichkeiten für rechtliche Probleme erwachsen können, die gleichzeitig die Privatheit der Nutzer/innen stärken und einen Markt für neue Produkte kreieren.

In diesem Sinne freuen wir uns mit der Veröffentlichung der Beiträge des diesjährigen Workshops darauf, auch im nächsten Jahr wiederum „Recht und Technik – Datenschutz im Diskurs“ anzubieten.

R. Reussner, A. Koziolek, R. Heinrich (Hrsg.): INFORMATIK 2020,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2021 815

IT-Rahmenwerk für den Beschäftigtendatenschutz

Technologieeinführung aus rechtlicher und arbeitswissenschaftlicher Perspektive

Christian K. Bosse[1] Aljoscha Dietrich[2] Hartmut Schmitt[3]

Abstract: Die Digitalisierung der Arbeitswelt führt nicht nur zu mehr Flexibilität und Optimierungsmöglichkeiten, sondern ermöglicht auch tiefgreifende Analyse- und Überwachungsmöglichkeiten bezüglich der Arbeitnehmer. Diese Entwicklung kann daher von diesen als Bedrohung empfunden werden und als Reaktion etwa zu Umgehungs- oder Abwehrstrategien führen. Um dieser Problematik zu begegnen, stellen wir in diesem Beitrag ein Rahmenwerk vor, das im Rahmen eines laufenden Forschungsvorhabens erarbeitet wird und das Unternehmen bei der Entwicklung und Einführung IT-gestützter Lösungen für den betrieblichen Datenschutz, z. B. Privacy Dashboards, unterstützen soll. Ausgehend von juristischen und arbeitswissenschaftlichen Aspekten sind wesentliche Bestandteile des Rahmenwerks ein Qualitätsmodell sowie ein Selbstbewertungsinstrument, welches Unternehmen die Einführung entsprechender IT-Lösungen erleichtert.

Keywords: Beschäftigtendatenschutz; Rahmenwerk; Qualitätsmodell; Privacy Enhancing Technolgies; Privacy Dashboards; Selbstbewertung

1 Einleitung und Motivation

Digitalisierte Unternehmen können heute in umfassender Weise die Daten ihrer Arbeitsprozesse erheben und analysieren. Auf dieser Basis optimieren sie Prozesse, etwa indem sie Produktionsabläufe effizienter und kostensparender gestalten. In Zusammenhang mit der digitalen Transformation der Unternehmen werden allerdings immer mehr personenbezogene Daten der Beschäftigten erhoben und verarbeitet, darunter oft Daten, die Rückschlüsse auf Arbeitsverhalten und -leistung, Konsumverhalten oder persönliche Vorlieben zulassen. Dies kann die informationelle Selbstbestimmung der Beschäftigten gefährden und einen unzulässigen Eingriff in die Privatsphäre darstellen, beispielsweise wenn bei der Erfassung von Bewegungsdaten die Grenze zur unzulässigen Überwachung der Beschäftigten überschritten wird [Bo19].

Zwei aktuelle Entwicklungen verdeutlichen die Problematik: Im Personalbereich analysieren immer mehr Unternehmen personenbezogene Daten, um ihre Entscheidungsprozesse zu unterstützen. Mit dieser Praxis – People Analytics genannt – begeben sich die Unternehmen

[1] Institut für Technologie und Arbeit, Trippstadter Str. 113, 67663 Kaiserslautern, christian.bosse@ita-kl.de
[2] Lehrstuhl für Rechtsinformatik, Universität des Saarlandes, 66123 Saarbrücken, aljoscha.dietrich@legalinf.de
[3] HK Business Solutions GmbH, Mellinweg 20, 66280 Sulzbach, hartmut.schmitt@hk-bs.de

doi:10.18420/inf2020_73

in eine rechtliche Grauzone, von manchen Seiten wird sie sogar als rechtswidrig eingeschätzt [Ha20]. Für Aufsehen sorgte Zalando mit einem selbstentwickelten Bewertungssystem [Ze19], aber auch Hersteller wie Microsoft, IBM oder SAP bieten entsprechende Standardprodukte an. Der Home-Office-Boom im Frühjahr 2020, hervorgerufen durch die Corona-Pandemie, veranlasst immer mehr Unternehmen, ihre Mitarbeiter stärker zu überwachen [Mo20]. Dies geschieht durch organisatorische Maßnahmen – Mitarbeiter müssen ihren Vorgesetzten den Zugriff auf E-Mail-Postfächer und Chats gewähren – oder durch Spezialsoftware, die das Mitarbeitertracking ermöglicht. Der Markt solcher Trackingprogramme hat sich innerhalb weniger Wochen verdreifacht [Mo20].

Seit Mai 2018 sorgt die Datenschutzgrundverordnung (DSGVO) für eine strengere Regulierung auch des betrieblichen Datenschutzes: Unternehmen haben erweiterte Informationspflichten gegenüber den Betroffenen, müssen Verarbeitungsverzeichnisse für personenbezogene Daten erstellen und Datenschutzpannen melden. Das Problem: Viele Regelungen der DSGVO sind offen formuliert und machen keine Vorgaben hinsichtlich Technik und Anwendung. Dadurch wissen Unternehmen oft nicht, wie sie sich genau zu verhalten haben, und empfinden die Formulierungen der DSGVO als schwammig [Ma19]. 74 % der Unternehmen sehen einer aktuellen Bitkom-Studie zufolge Datenschutzanforderungen aktuell als größte Hürde beim Einsatz neuer Technologien [Bi19]. Diskutiert wird zudem, inwieweit durch die DSGVO Mitbestimmungsrechte gemäß Betriebsverfassungsgesetz tangiert werden. Durch Urteile und Konkretisierungen in der Praxis müssen also noch einige Lücken geschlossen werden.

Dem Interesse der Unternehmen, die Potentiale einer umfänglichen Datenanalyse zu nutzen, steht das Recht der Betroffenen auf Privatsphäre und informationelle Selbstbestimmung entgegen. Die Betroffenen wissen oft noch nicht einmal, wer welche personenbezogenen Daten zu welchem Zweck verarbeitet und welche Konsequenzen dies für ihre Privatsphäre hat. Ein erfolgversprechender Ansatz, um mögliche Zielkonflikte zwischen Arbeitgebern und Beschäftigten bzw. Mitarbeitervertretungen aufzulösen und eine datenschutzkonforme Verarbeitung personenbezogener Daten zu ermöglichen, sind dedizierte IT-Lösungen für den Beschäftigtendatenschutz. Diese können beispielsweise in Form von Privacy Dashboards ausgestaltet sein. Privacy Dashboards bündeln sämtliche Datenschutzfunktionen in einer zentralen Oberfläche. Sie stellen zum einen Transparenz her, welche personenbezogenen Daten erhoben und verarbeitet werden, denn nur dann sind die Beschäftigten in der Lage, informierte Entscheidungen zu treffen. Zum anderen geben sie den Beschäftigten ein probates Mittel an die Hand, um eigene Datenschutzpräferenzen effektiv durchzusetzen.

Im Forschungsprojekt TrUSD[4] [Tr20] erforschen wir gemeinsam mit weiteren Partnern generische Modelle und Umsetzungskonzepte, aber auch konkrete Ausgestaltungen und Wirkweisen solcher Privacy Dashboards. Die vorgestellten Zwischenergebnisse geben Einblick in dieses laufende Forschungsvorhaben, können aber noch keine Auswertungen des wissenschaftlichen Erfolgs liefern.

[4] Das Forschungsprojekt «TrUSD – Transparente und selbstbestimmte Ausgestaltung der Datennutzung im Unternehmen» wird gefördert durch das deutsche Bundesministerium für Bildung und Forschung (BMBF).

2 Verwandte Arbeiten

Technische Hilfsmittel, um die Privatsphäre zu schützen bzw. die getätigten Privatsphäre-Einstellungen sicher und zuverlässig durchzusetzen, werden in der Literatur allgemein als *Privacy Enhancing Technologies* (PETs) bezeichnet. Hierbei handelt es sich um eine Vielzahl von Verfahren, wie Anonymisierungsnetze (z. B. TOR) oder Anonymisierungsverfahren in Datenbanken, etwa basierend auf Differential Privacy [Dw06] und k-Anonymität [Sw02]. *Transparency Enhancing Tools* (TETs) hingegen sollen dem Betroffenen die Verarbeitungsprozesse und auch die hieraus entstehenden Konsequenzen erklärbar machen [Fi16]. Privacy Dashboards entsprechen zunächst dem Grundgedanken der TETs indem sie etwa die Verarbeitungsprozesse verdeutlichen, können jedoch auch Funktionalitäten von PETs umfassen. Privacy Insight [BKB16] ist ein Transparenz-Dashboard, das Datenverarbeitungen graphenbasiert anzeigt. Jedoch ist es bisher nicht im industriellen Kontext evaluiert oder für Arbeitnehmer als Nutzer optimiert worden. Es bietet außerdem keine Möglichkeiten zur Selbstbestimmung oder Durchsetzung von Privatheitsbedürfnissen der Nutzer. Die Karlstad-Universität hat mit dem Tool Data Track einen Ansatz vorgestellt, wie die Weitergabe von Daten visualisiert werden kann [Fi16], und die besonderen Anforderungen an Privacy Dashboards in Cloud-Umgebungen erforscht [FAP14]. Allgemeine Anforderungen und eine prototypische Umsetzung wurde von den Telekom Innovation Laboratories in Berlin erforscht [Ra17]. Die TU Berlin arbeitete zusammen mit der Mozilla Corporation an einem benutzerfreundlichen Privacy Dashboard für Firefox OS [Pi15], das Nutzern verschiedene Funktionen von mobilen Endgeräten erklärt, die personenbeziehbare Daten preisgeben, und mit dem die Datennutzung eingeschränkt werden kann. Die Universität Oslo hat ein Identitäts-Dashboard [SJ10] vorgestellt, das Nutzern eine Übersicht über die Verwendung verschiedener digitaler Identitäten und der damit verknüpften Daten gibt. Die Universität Freiburg stellte eine Klassifizierung von Privacy Dashboards [ZAM14] sowie eine empirische Analyse zu deren Akzeptanz [CZM16] vor.

Qualitätsmodelle haben im Software Engineering eine lange Tradition: 1977 nahmen McCall et al. [Mc77] erstmals eine Unterscheidung von Faktoren, Kriterien und Metriken vor, die seitdem als Muster für den Aufbau dieser Modelle dient. Grady & Caswell [GC87] stellten 1987 das Qualitätsmodell FURPS vor, das den Grundstein für den ersten internationalen Standard, ISO 9126, legte. Dessen Revision (ISO 25010) liefert das aktuell umfangreichste Softwarequalitätsmodell. Ein Qualitätsmerkmal Datenschutz fehlt hier allerdings, ebenso relevante Aspekte des strukturellen Umfelds bzw. der Prozesse, in denen ein Softwareprodukt genutzt wird. Von der Konferenz der unabhängigen Datenschutzbehörden des Bundes und der Länder wurde als Teil der Modernisierung des Datenschutzrechts ein Konzept von Schutzzielen verabschiedet. Dieses Konzept sowie eine Methode zur Datenschutzberatung und -prüfung wurden 2015 in Form des Standard-Datenschutzmodells (SDM, [AKT19]) veröffentlicht.

Das Konzept der *Selbstbewertung* entstammt ursprünglich dem Bereich des Qualitätsmanagements und findet sich daher auch als durchgängiger Bestandteil der Normreihe ISO 9000

[Ka11]. Im Zuge der Digitalisierung etablierten sich in den letzten Jahren Selbstbewertungsinstrumente als Online-Check. Diese finden sich in den verschiedensten Anwendungsfeldern, in denen sie mit Hilfe von Checklisten und/oder Fragenkatalogen eine Selbstbewertung vor dem Hintergrund eines Reifegradmodells automatisiert vornehmen. Beispiele sind der Online-Check zum Arbeitsschutz der Berufsgenossenschaft für Gesundheitsdienst und Wohlfahrtspflege [BGW17] sowie der Readiness-Check Digitalisierung des Mittelstand 4.0-Kompetenzzentrums Kaiserslautern [BH18], dessen ausführliche Dokumentation des Entwicklungs- und Umsetzungsprozesses [HSB19] als Grundlage der Eigenentwicklung im TrUSD-Projekt diente. Ein vergleichbares Selbstbewertungsinstrument, das den Beschäftigtendatenschutz in Unternehmen zum Gegenstand hat, existiert aktuell nicht.

3 Rechtliche Perspektive

Spätestens seit der Anwendung der DSGVO am 25.5.2018 verfügen die Mitgliedstaaten der Europäischen Union über ein einheitliches und verschärftes Datenschutzrecht, das aufgrund der drohenden Sanktionen und Geldbußen bei Verstößen Unternehmen zur Handlung drängt. Ziel der Sanktionen und Geldbußen ist es, wirksam, verhältnismäßig und abschreckend zu sein.[5] Ausdruck hiervon sind mögliche Geldbußen von bis zu 20 Millionen (Mio.) EUR bzw. 4 % des gesamten weltweiten Jahresumsatzes eines Unternehmens. Dass abschreckende Strafen nicht nur in der Theorie bestehen, bewiesen die zuständigen Aufsichtsbehörden bereits am Internetprovider 1&1 (9,4 Mio. EUR), der Immobiliengesellschaft Deutsche Wohnen (14,5 Mio. EUR), der Hotelkette Marriott (ca. 110 Mio. EUR) und British Airways (ca. 204 Mio. EUR) [Ba19a]. Daher ist es schon aus rein betriebswirtschaftlicher Sicht den Unternehmen anzuraten, die Datenschutzgesetzgebung zu befolgen. Motivation braucht jedoch nicht nur Angst vor Strafe sein: ein ernstgenommener Datenschutz kann ebenso ein Wettbewerbsvorteil[6] sein und die Einführung kann zum Aufbau eines Risikomanagements genutzt werden [RCH18].

Die Datenschutzanforderungen bestehen nicht nur nach außen, beispielsweise gegenüber den Kunden, sondern auch innerhalb einer Organisation gegenüber den eigenen Mitarbeitern. Die DSGVO erlaubt für den Beschäftigtendatenschutz explizit nationale Regelungen der Mitgliedsstaaten durch die Öffnungsklausel des Art. 88. In Deutschland wurde hierzu § 26 Bundesdatenschutzgesetz (BDSG) geschaffen. Dieser Paragraph ist jedoch sehr unspezifisch, wie schon § 32 BDSG alter Fassung (a. F.) zuvor. In der Vergangenheit hat dies zu sogenanntem Richterrecht geführt, d. h. Detailregelungen wurden aus hochinstanzlicher Rechtsprechung hergeleitet. Neben fehlender Systematik und Ordnung besteht hierbei eine grundsätzliche Problematik bei der Übertragbarkeit auf andere Fälle. Da sich der Kern des

[5] Vgl. Art. 83 Abs. 1 S. 1 DSGVO sowie [Br19].

[6] Insbesondere wenn von der Möglichkeit der Zertifizierung Gebrauch gemacht wird, vgl. Scholz, Rn. 4, DSGVO Art. 42, in [SHS19].

§ 26 BDSG kaum von seinem Vorgänger unterscheidet, ist jedoch von einer weiteren Anwendbarkeit des Richterrechts auszugehen[7], was wohl auch der Gesetzgeber beabsichtigte.[8] Bestrebungen ein eigenes und detaillierter ausgearbeitetes Beschäftigtendatenschutzgesetz zu schaffen, gibt es seit den 1980er Jahren. Der § 32 BDSG a. F. war ursprünglich ein Provisorium und eine direkte Reaktion auf betriebliche Datenschutzskandale. Die Verabschiedung eines bereichsspezifischen Gesetzes scheiterte jedoch und wurde auf einen Zeitpunkt nach Verabschiedung der DSGVO vertagt.[9]

Eine Analyse der Rechtsprechung zum Mitarbeiterdatenschutz lässt drei Themenschwerpunkte erkennen: Überwachung von Mitarbeitern (insbesondere Videoüberwachung), Einsicht- bzw. Auskunftsrechte sowie Verarbeitung von Mitarbeiterdaten. Die Entscheidungen können herangezogen werden, um Umfang und Grenzen zulässiger Verarbeitung von Mitarbeiterdaten zu bestimmen. Ausgangspunkt ist § 26 Abs. 1 S. 1 BDSG, welcher eine Erforderlichkeit der Verarbeitung von Daten im Beschäftigungsverhältnis voraussetzt. Dies führt zu einer Abwägung zwischen Arbeitgeber- und Arbeitnehmerinteresse. Das Richterrecht hat zu einem dreistufigen Prüfungsschema der Verhältnismäßigkeit geführt, bestehend aus (1) Geeignetheit, (2) Erforderlichkeit und (3) Angemessenheit.[10] Diese Systematik wurde hauptsächlich zu Fragen der Videoüberwachung entwickelt, aufgrund ihrer hohen Abstraktion lässt sie sich jedoch auch gut auf andere Fragestellungen und neuere Technologien übertragen, z. B. Analyse von Nutzungsprotokollen, Kommunikation, Keylogger, Positionsbestimmung oder Wearables. Die Verarbeitungsgrundlage kann im Beschäftigungskontext auch auf Betriebsvereinbarungen fußen, die ebenfalls dem digitalen Wandel unterworfen sind. Vorteile sind u. a. die betriebsweite und grundsätzlich beständige Gültigkeit, im Gegensatz zur Einwilligung [SV20].

Neben den Anforderungen aus § 26 BDSG sind auch die allgemeinen Grundsätze der Verarbeitung (Art. 5 DSGVO) und die Betroffenenrechte (Art. 15–22 DSGVO) umzusetzen, wozu auch die Rechte der Mitarbeiter zählen [Bo19]. Art. 25 Abs. 1 DSGVO verlangt vom Verantwortlichen den Einsatz technischer und organisatorischer Maßnahmen, die dafür ausgelegt sind, die Datenschutzgrundsätze wirksam umzusetzen. Diese Aufforderung zum technischen Datenschutz findet sich etwas detailreicher wieder in Art. 32 DSGVO – Sicherheit der Verarbeitung. Ein Dashboard für den Beschäftigtendatenschutz kann als eine technische Maßnahme in diesem Sinne bewertet werden. Eine weitere Anforderung für den Verantwortlichen ergibt sich aus Art. 35 DSGVO, der eine Datenschutz-Folgenabschätzung (DSFA) verlangt, wenn eine Verarbeitung erfolgen soll, welche „[...] voraussichtlich ein hohes Risiko für die Rechte und Freiheiten natürlicher Personen zur Folge [...] [hat]“. Zur Identifikation einer solchen Verarbeitung kann ein Verarbeitungsverzeichnis i. S. d. Art. 30 DSGVO hilfreich sein, das vorgeschrieben ist, sofern ein Unternehmen etwa mehr als 250 Mitarbeiter beschäftigt (Art. 30, Abs. 5 DSGVO). Das in Abschnitt 6 vorgestellte

[7] Vgl. hierzu Zöll, BDSG § 26, Rn. 3 in [TG19].
[8] Vgl. [De17], S. 96f.
[9] Vgl. Riesenhuber, § 26, Rn. 8-10, [BW20].
[10] Vgl. hierzu Zöll, BDSG § 26, Rn. 25 in [TG19].

Selbstbewertungsinstrument kann bei der Erstellung eines Verarbeitungsverzeichnisses und Identifizierung der Notwendigkeit einer DSFA unterstützen.

4 Arbeitswissenschaftliche Perspektive

Die Einführung neuer Technologien und digitaler Lösungen bedeutet für Unternehmen einen wichtigen Schritt zur Sicherung ihrer Wettbewerbsfähigkeit. Die Auswertung von Daten, die in den Arbeitsprozessen manuell oder automatisiert verarbeitet werden, bietet die Möglichkeit, bestehende Prozesse und Arbeitsabläufe zu optimieren. Gleichzeitig birgt die Einführung neuer Technologien diverse Risiken in sich. Denn Digitalisierung im Unternehmen bedeutet nicht nur die Einführung einer neuen Technologie, digitalen Lösung oder Software. Sie ist vielmehr ein tiefgreifender Veränderungsprozess, der das gesamte sozio-technische System umfasst [Ul11, Th15]. Viel zu oft wird der Mitarbeiter bzw. die soziale Sphäre im Unternehmen vernachlässigt, sodass von einer Digitalisierung in zwei Geschwindigkeiten gesprochen wird: Einerseits schreitet die technologische Entwicklung rasant voran, während andererseits die erforderliche Gestaltung der Organisationen deutlich mehr Zeit und Veränderungswillen bedarf [Bo19a]. Dabei ist Technologie nur ein Gestaltungselement im Kontext der digitalen Transformation, die neben der technischen Sphäre eines Unternehmens ebenso die organisationale und soziale Sphäre beeinflusst. Die erfolgreiche Einführung digitaler Lösungen ist daher als eine komplexe Gestaltungsaufgabe zu verstehen, deren Wechselwirkungen zwischen allen drei Sphären (Abbildung 1) zu beachten sind [Bo19a].

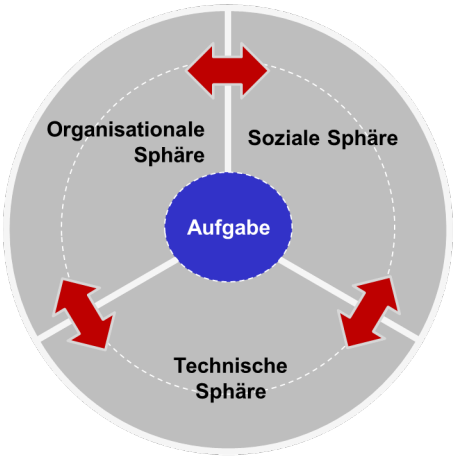


Abb. 1: Wechselwirkung zwischen den drei Sphären der digitalen Transformation als komplexe Gestaltungsaufgabe in Unternehmen

Werden diese komplexen Wirkzusammenhänge nicht ausreichend beachtet, kann es zu Schwierigkeiten bei der Umsetzung kommen. So gilt es bei der Einführung einer neuen Tech-

nologie auf organisationaler Ebene die verschiedenen, kontextspezifischen Regelungen und rechtlichen Vorgaben zu berücksichtigen, wie Gesetze, Normen oder Betriebsvereinbarungen. Gleichzeitig sind die Mitarbeiter einzubeziehen, um eine erfolgreiche Technologieeinführung zu ermöglichen. Die Partizipation bei der Technologieauswahl und -ausgestaltung ist ebenso wichtig wie die Qualifizierung im Hinblick auf die Anwendung. In der Praxis sind viele Szenarien denkbar, in denen ein effizienter Einsatz und das Ausschöpfen des vollen Potenzials einer digitalen Lösung nicht gelingen (können): die ausgewählte Technologie entspricht nicht den Anforderungen am Arbeitsplatz, Mitarbeiter wurden nicht rechtzeitig geschult und sind überfordert oder der Betriebsrat sieht die Mitarbeiter unzulässig überwacht und blockiert den Einsatz der digitalen Lösung [Bo19].

Die komplexen Wirkzusammenhänge zwischen den Sphären sind insbesondere bei der Einführung technischer Lösungen für den Beschäftigtendatenschutz zu beachten. Mit deren Einsatz wird Mitarbeitern die Möglichkeit der Transparenz und Selbstbestimmung bezüglich ihrer personenbezogenen Daten gegeben. Gleichzeitig setzen die Mitarbeiter sich oft erst im Rahmen der Technologieeinführung und -nutzung mit dieser Thematik aktiv auseinander und erfahren bisher unbekannte Details über die Verwendung ihrer personenbezogenen Daten im Unternehmen. Ohne eine vorherige Partizipation der Beschäftigten (soziale Sphäre), eine Anpassung der Prozesse sowie – bei Bedarf – das Abschließen von Betriebsvereinbarungen (organisationale Sphäre) kann dies durch die Mitarbeiter als Überwachung empfunden werden und in nicht intendierte Auswirkungen resultieren. Denkbar ist beispielsweise, dass sich die Beschäftigten ständiger Kontrolle ausgesetzt fühlen und ihr Verhalten derart anpassen, dass es negative Auswirkungen auf sie selbst oder die organisationalen Prozesse und Arbeitsabläufe hat [RS16, DCL15, Pr15]. Des Weiteren kann die Situation in einem gestörten Vertrauensverhältnis zwischen Arbeitgeber und Arbeitnehmer gipfeln, wenn Mitarbeiter versuchen die technische Lösung zu umgehen oder gezielt verfälschte Daten produzieren [Mo15, Pr15]. In der Praxis lassen sich viele Beispiele anführen, z. B. indem Beschäftigte

- bei einer digitalen Zeiterfassung nach dem Ausstechen an den Arbeitsplatz zurückkehren und weiterarbeiten, um hierdurch bestehende Regelungen zu umgehen,
- beim Arbeiten in der Produktion ihre Tätigkeiten bereits als fertig zurückmelden, obwohl nicht alle Tätigkeiten abgeschlossen sind, um so ihre Durchlaufzeiten zu verbessern, oder
- bei einer Videoüberwachung der Eingangsbereiche die Gebäude durch nicht überwachte Notausgänge verlassen, um der Videoüberwachung zu entgehen [Bo19].

Darüber hinaus gibt es Indizien dafür, dass die Überwachung am Arbeitsplatz insgesamt zu einem Verlust der wahrgenommenen Kontrolle und zu einem gesteigerten subjektiven Stresserleben führt [Ba19].

5 Rahmenwerk für den Beschäftigtendatenschutz

Ziel des TrUSD-Projekts ist es, ausgehend von den rechtlichen und arbeitswissenschaftlichen Randbedingungen ein Rahmenwerk zu schaffen, das Unternehmen bei der Entwicklung IT-gestützter Lösungen für den Beschäftigtendatenschutz unterstützt. Da die Unternehmen sich hinsichtlich Größe, Branche, Infrastruktur und Mitarbeiterfähigkeiten unterscheiden, wollen wir möglichst generische, vielseitig einsetzbare Bausteine zur Verfügung zu stellen. Diese sollen den Unternehmen eine geeignete Entscheidungs- und Arbeitsgrundlage bieten, um eine passgenaue Lösung, z. B. ein Privacy Dashboard für ihre Mitarbeiter, zu entwickeln.

In einem Anforderungsmodell haben wir die Bedarfe und Anforderungen der relevanten Stakeholder beschrieben, z. B. Selbstbestimmungs- und Transparenzbedarfe und Benutzer- bzw. Systemanforderungen. Mentale Modelle und Persona-Beschreibungen helfen beim besseren Verständnis der Anwendergruppen, z. B. der Beschäftigten, des Managements oder des Datenschutzbeauftragten. Anwendungsfälle und -szenarien konkretisieren mögliche Einsatzbereiche eines Dashboards, ein Stufenkonzept unterstützt beim schrittweisen Aufbau – vom reinen Informationspanel bis hin zur Transparentmachung von Datenfluss-Manipulationen in Echtzeit [To20]. Ein Architekturkonzept mitsamt integrierten Werkzeugen (z. B. PETs zur Anonymisierung personenbezogener Daten), unterschiedlich gestaltete UI- und Interaktionskonzepte sowie komplementäre Einführungskonzepte helfen außerdem bei der technischen Implementierung und der Einführung im Unternehmen.

5.1 Qualitätsmodell

Ein wesentlicher Bestandteil des Rahmenwerks ist das im Folgenden vorgestellte Qualitätsmodell. Dieses wurde für den Bereich Beschäftigtendatenschutz entwickelt, ist aber auch auf andere Datenschutzbereiche übertragbar. Bei der Entwicklung IT-gestützter Lösungen für den betrieblichen Datenschutz sind unterschiedliche Qualitätseigenschaften von Bedeutung (siehe Abbildung 2). Diese können sich auf die Produktqualität der geplanten Lösung beziehen (z. B. Zuverlässigkeit oder Performanz), aber auch auf deren Nutzungsqualität (z. B. Zufriedenheit der Nutzer), die Prozessqualität (z. B. Prozesskonformität) oder die Strukturqualität (z. B. Kompetenz und Bewusstsein der Mitarbeiter).

Die Grundlage unseres Qualitätsmodells für die Bereiche Produkt- und Nutzungsqualität bildet das Modell der ISO 25010, erweitert um einzelne Teilmerkmale der ISO 9241. Ergänzt wurden außerdem die Gewährleistungsziele des Standard-Datenschutzmodells. Der Bereich Strukturqualität folgt der ISO 9001, der Bereich Prozessqualität dem Modell Gokyo Ri [Kn19]. Bei der Integration des Qualitätsmodells in das Rahmenwerk gilt unser besonderes Augenmerk den Beziehungen zwischen einzelnen Qualitätseigenschaften (vgl. [Wa02]). Es ist ersichtlich, welche Eigenschaften sich verstärken, z. B. Korrektheit und Zuverlässigkeit, und welche Eigenschaften miteinander konkurrieren, d. h., eine Maßnahme zur Verbesserung der einen Eigenschaft führt potentiell zu einer Verschlechterung der anderen Eigenschaft, z. B. Verfügbarkeit und Vertraulichkeit.

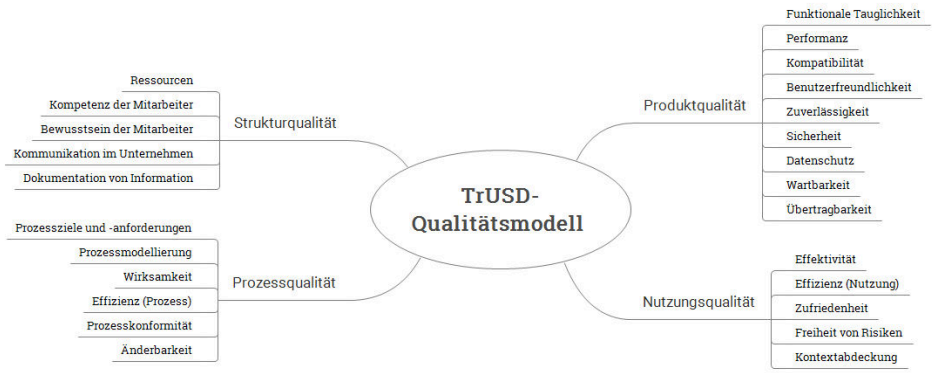


Abb. 2: Übersicht des Qualitätsmodells für den Beschäftigtendatenschutz

5.2 Operationalisierung

Damit Anwender des Rahmenwerks potentielle Konflikte identifizieren und einen geeigneten Trade-off ermitteln können, haben wir die Qualitätseigenschaften jeweils mit typischen Umsetzungsmaßnahmen verknüpft (Beispiel: „Integrität“ ist verknüpft mit „Schutz vor Schadsoftware“). Hierbei haben wir sowohl die verstärkenden als auch die konkurrierenden Beziehungen im Rahmenwerk dokumentiert. Zudem haben wir bei den Qualitätseigenschaften jeweils Checklistenpunkte hinterlegt. Diese helfen zum einen beim Erreichen bestimmter Qualitätseigenschaften, zum anderen ermöglichen sie im Nachgang eine systematische Bewertung der entwickelten Lösung. Auf Grundlage dieses Rahmenwerks entwickeln mehrere Partner des TrUSD-Projekts exemplarische Lösungen, die mit Endanwendern getestet und evaluiert werden. Hierbei wird untersucht, ob die Dashboards mehr Transparenz bei der Erhebung und Nutzung personenbezogener Daten im Unternehmen schaffen, die Durchsetzung eigener Datenschutzpräferenzen ermöglichen und so zu einem fairen Ausgleich zwischen Arbeitgeber- und Arbeitnehmerinteressen beitragen.

6 Betriebsinterne Feststellung des Datenschutzniveaus

Die Umsetzung rechtlicher Vorgaben stellt insbesondere kleine und mittlere Unternehmen vor eine Herausforderung, die oft nur mit externer Unterstützung zu schaffen ist. Denn zumeist verfügen diese Unternehmen nicht über das notwendige Fachwissen, um die in juristischer Fachsprache verfassten rechtlichen Verordnungen zu verstehen und einen rechtskonformen Datenschutz umzusetzen. Gleichzeitig führen die potenziellen Strafen in vielen Unternehmen zu großen Unsicherheiten [Be19]. Selbst wenn einzelne Maßnahmen bereits ergriffen wurden, bestehen oft unbedachte, bis dato unbekannte Lücken im Datenschutz, verdeutlicht an folgendem Beispiel: Um Auswertungen der elektronisch erfassten Arbeitszeiten zu

verhindern, die über die Lohnabrechnung hinausgehen und Rückschlüsse über die Effizienz der einzelnen Mitarbeiter zulassen, werden die Daten und die Auswertungsfunktion der Zeiterfassungssoftware mit einer Zugriffskontrolle versehen. Lediglich die Personalabteilung kann einsehen, welcher Mitarbeiter zu welcher Uhrzeit ein- bzw. ausgestochen hat. Auf den ersten Blick sind somit die personenbezogenen Daten über die individuellen Arbeitszeiten geschützt. Die zentrale Stechuhr befindet sich jedoch im Eingangsbereich des Gebäudes, der mit einem elektronischen Schließsystem gegen Fremdzutritt gesichert und videoüberwacht ist. Somit ist es Dritten (z. B. dem Sicherheitsdienst) möglich, anhand der vom Schließsystem erfassten Daten oder der Auswertung der Videoaufnahmen die Arbeitszeiten der Mitarbeiter zu eruieren. Während die Auswertung von Videomaterial aus Überwachungskameras meist in Betriebsvereinbarungen geregelt ist, werden die Erfassungsdaten eines elektronischen Schließsystems oft nicht bedacht.

Um den Ist-Zustand des Datenschutzes im Unternehmen ohne großen Ressourcenbedarf oder Unterstützung externer Berater zu ermitteln, bietet sich der Einsatz eines Selbstbewertungsinstruments an. Diese Instrumente erfreuen sich in den letzten Jahren großer Beliebtheit und werden von Unternehmen beispielsweise zur Ermittlung des digitalen Reifegrades [BH18] oder zur Einschätzung der Organisation des Arbeitsschutzes [BGW17] eingesetzt. Dabei werden meist nicht nur die Ergebnisse der Selbstbewertung dokumentiert, sondern auch Entwicklungspotenziale aufgezeigt und Lösungen aus Expertensicht vorgeschlagen [HSZ18]. Ein solches Selbstbewertungsinstrument, das den Beschäftigtendatenschutz in Organisationen und Unternehmen zum Gegenstand hat, entwickeln und erproben wir im Rahmen des TrUSD-Projekts. Ziel des Instruments ist es zum einen, Organisationen und Unternehmen für das Thema Datenschutz zu sensibilisieren. Zum anderen wird Geschäftsführern, Führungskräften und weiteren Verantwortliche eine praxisgerechte Unterstützung bei der (Weiter-)Entwicklung eines unternehmensspezifischen und rechtskonformen Datenschutzes zur Verfügung gestellt.

Um die Entwicklung anzugehen, haben wir zunächst Online-Selbstbewertungsinstrumente aus anderen Themenbereichen analysiert. Grundlage für die Entwicklung unseres Instruments bildet ein Kriterienkatalog, der eine systematische Erfassung des Ist-Zustands und eine Bewertung der Umsetzung datenschutzrechtlicher Anforderungen ermöglicht. Dieser Kriterienkatalog basiert zum einen auf den Ergebnissen der Projektarbeit, beispielsweise der Anforderungserhebung bei den Anwendungspartnern, den erhobenen Datennutzungs- und Schutzbedarfen sowie den Selbstbestimmungs- und Transparenzbedarfen. Zum anderen referenzieren wir die Anforderungen der DSGVO sowie die Gewährleistungsziele des Standard-Datenschutzmodells: Datenminimierung, Verfügbarkeit, Integrität, Vertraulichkeit, Nichtverkettung, Transparenz, Intervenierbarkeit und Evaluierbarkeit. Das Selbstbewertungsinstrument ist außerdem eng verbunden mit dem Verarbeitungsverzeichnis des Art. 30 DSGVO, in dem alle Verarbeitungstätigkeiten in der Zuständigkeit des Verantwortlichen zu erfassen sind. Besteht ein Verarbeitungsverzeichnis, kann es genutzt werden, um die Fragen schneller zu beantworten. Besteht es nicht, kann das Selbstbewertungsinstrument bei dessen Erstellung unterstützen. Folgend kann dies bei der Identifizierung von Verarbeitungsvorgän-

gen genutzt werden, welche voraussichtlich ein hohes Risiko für die Rechte und Freiheiten der Beschäftigten darstellen und daher eine Datenschutz-Folgenabschätzung i. S. d. Art. 35 DSGVO bedürfen. Ergänzend werden weiche Faktoren adressiert, die ebenfalls ein wichtiger Bestandteil des Datenschutzes in Organisationen sind [Wa12], zum Beispiel die Information, Sensibilisierung und Qualifizierung der Mitarbeiter. Ziel des Selbstbewertungsinstruments ist es, eine möglichst vollständige und umfassende Abdeckung der verschiedenen Facetten des Beschäftigtendatenschutzes sowie eine allgemeine Verständlichkeit auch für juristische Laien zu erreichen.

Das Selbstbewertungsinstrument haben wir als Online-Befragung mit der Open-Source-Software LimeSurvey umgesetzt. Mit Hilfe eines für Laien gut verständlichen Fragenkatalogs, der auf dem umfassenden Kriterienkatalog basiert, werden sowohl rechtliche als auch arbeitswissenschaftliche Aspekte erhoben. Die Online-Befragung wird dabei in verschiedene Themenbereiche strukturiert, um sie übersichtlicher zu gestalten. Darüber hinaus wird die Anzahl der zu beantwortenden Fragen durch den Einsatz von Filterfragen optimiert, sodass gegebenenfalls im Unternehmenskontext irrelevante Fragen nicht angezeigt werden. Auf diese Weise wird das Verhältnis von Aufwand zu Nutzen bei Einsatz des Selbstbewertungsinstruments unternehmensindividuell optimiert.

Insgesamt bietet das Selbstbewertungsinstrument die Möglichkeit, den Ist-Zustand des Beschäftigtendatenschutzes zu erheben. Dies geschieht durch eine auf den gegebenen Antworten basierende Einschätzungen zur Rechtskonformität und zum Reifegrad im Bereich des Beschäftigtendatenschutzes bzgl. der bisher umgesetzten Maßnahmen. Jeder Teilnehmer erhält unmittelbar die Auswertung seiner individuellen Ergebnisse und daran anknüpfende Handlungsempfehlungen zur Weiterentwicklung des Beschäftigtendatenschutzes. Zudem stehen Checklisten bereit, mit deren Hilfe die rechtskonforme Umsetzung im Unternehmen gefördert wird. Dieses Selbstbewertungsinstrument wird nach Abschluss der Test- und Evaluationsphase online frei zur Verfügung stehen.

7 Fazit

Durch die Digitalisierung und die Anwendbarkeit der DSGVO sind Unternehmen vor neue Herausforderungen im Beschäftigtendatenschutz gestellt. Geeignete IT-Lösungen, beispielsweise in Form von Privacy Dashboards, unterstützen Unternehmen bei einer rechtskonformen Umsetzung des Beschäftigtendatenschutzes. Die verbesserte Transparenz und Mitbestimmung erzeugen bei den Mitarbeitern zudem eine höhere Akzeptanz der Datenverarbeitung. Besonders relevant für eine erfolgreiche Umsetzung sind eine umfassende und möglichst vollständige Anforderungserhebung und das hieraus abgeleitete Qualitätsmodell, beides Bestandteile des vorgestellten Rahmenwerks. Ein komplementäres Selbstbewertungsinstrument unterstützt Unternehmen bei der Analyse des Ist-Zustands sowie bei der Implementierung und Evaluation technischer bzw. organisationaler Lösungen.

Literaturverzeichnis

[AKT19] AK Technik, DSK: Das Standard-Datenschutzmodell, Version 2.0. Beschluss der 98. DSK, Trier, 2019.

[Ba19] Backhaus, N.: Kontextsensitive Assistenzsysteme und Überwachung am Arbeitsplatz. Zeitschrift für Arbeitswissenschaft, Vol. 73 Iss. 1, S. 2-22, 2019.

[Ba19a] Bayrischer Rundfunk: Datenschutzgrundverordnung: Die Schonfrist ist vorbei, https://www.br.de/nachrichten/bayern/datenschutzgrundverordnung-die-schonfrist-ist-vorbei, Stand: 29.4.2020.

[Be19] Becker, W. et.al.: Digitale Arbeitswelten im Mittelstand. Veränderungen und Herausforderungen. Springer, Wiesbaden, 2019.

[BGW17] BGW, Berufsgenossenschaft für Gesundheitsdienst und Wohlfahrtspflege: Online-Check zum Arbeitsschutz. Heilberufe, vol. 69, iss. 7-8, S. 41, 2017.

[BH18] Bosse, C. K.; Hellge, V.: Digitalisierung im Mittelstand. Zeitschrift für Organisationsentwicklung, 01/18, S. 102-103, 2018.

[BKB16] Bier, C. et al.: PrivacyInsight: The Next Generation Privacy Dashboard. 4th Annual Privacy Forum, APF, Frankfurt a.M. September 7-8, 2016.

[Bi19] Bitkom e.V. (2019). Bitkom zieht gemischte Jahresbilanz zur DS-GVO `https://www.bitkom.org/Presse/Presseinformation/Bitkom-zieht-gemischte-Jahresbilanz-zur-DS-GVO`, Stand: 29.4.2020.

[Bo19] Bosse, C. K. et.al.: Beschäftigtendatenschutz: Rechtliche Anforderungen und technische Lösungskonzepte. In (Schweighofer, E.; Kummer, F.; Saarenpää, A., Hrsg.): Tagungsband des 22. Internationalen Rechtsinformatik Symposions (IRIS), 2019.

[Bo19a] Bosse, C. K. et.al.: Digitalisierung im Mittelstand erfolgreich gestalten. In (Bosse, C.K.; Zink, K.J., Hrsg.): Arbeit 4.0 im Mittelstand. Chancen und Herausforderungen des digitalen Wandels für KMU. Springer, Berlin/ Heidelberg, 2019.

[Br19] Brink, S.: Bußgeldrahmen nach der DS-GVO, ZD 2019, 141.

[BW20] Brink, S; Wolff, A.: BeckOK Datenschutzrecht, 32. Edition, 1.5.2020, C.H. Beck, 2020.

[Cr12] Cranor, L. F.: „P3P is dead, long live P3P!, http://lorrie.cranor.org/blog/2012/12/03/p3p-is-dead-long-live-p3p/, Stand: 29.4.2020.

[DCL15] Da Cunha, J. V.; Carugati, A.; Leclercq-Vandelannoitte, A.: The dark side of computer-mediated control. Information Systems Journal, 25, S. 319-354, 2015.

[CZM16] Cabinakova, J.; Zimmermann, C.; Mueller, G.: An Empirical Analysis of Privacy Dashboard Acceptance: The Google Case. ECIS, Istanbul, Turkey, June 12-15, 2016.

[De17] Deutscher Bundestag: Drucksache 18/11325, 2017.

[Dw06] Dwork, C.: Differential Privacy, in: Automata, Languages and Programming: 33rd ICALP 2006, LNCS, Bd. 4052, S. 1-12, Springer, Berlin/Heidelberg, 2006.

[FAP14] Fischer-Hübner, S. et al.: How can Cloud Users be Supported in Deciding on, Tracking and Controlling How their Data are Used?, in Privacy and Identity Management for Emerging Services and Technologies, S. 77–92. Springer Berlin/Heidelberg, 2014.

[Fi16] Fischer-Hübner, S. et al.: Transparency, Privacy and Trust – Technology for Tracking and Controlling My Data Disclosures: Does This Work?. 10th IFIP TM, Jul 2016, Darmstadt, Germany. pp.3-14

[GC87] Grady, R. B.; Caswell, D. L.: Software Metrics: Establishing a Company-Wide Program. Prentice-Hall, Englewood Cliffs, N.J., 1987.

[Ha20] Hagelüken, A.: Arbeitswelt – Personalanalyse von Mitarbeitern oft rechtswidrig. In: Süddeutsche Zeitung vom 02.03.2020. München: Süddeutsche Zeitung.

[HSB19] Hellge, V.; Schröder, D.; Bosse, C.K.: Der Readiness-Check Digitalisierung. Ein Instrument zur Bestimmung der digitalen Reife von KMU. Mittelstand 4.0-Kompetenzzentrum Kaiserslautern. https://kompetenzzentrum-kaiserslautern.digital/wp-content/uploads/2019/01/Broschüre_Readiness_Check_Digitalisierung_Januar_2019_final.pdf

[HSZ18] Hellge, V.; Schröder, D.; Zink, K.J.: Der Readiness-Check „Digitalisierung“ als Instrument im digitalen Transformationsprozess. In (Lingnau, V.; Müller-Seitz, G.; Roth, S., Hrsg.): Management der digitalen Transformation. Vahlen, 2019.

[Ka11] Kamiske, G.F.; Brauer, J.-P.: Qualitätsmanagement von A-Z. Hanser, 2011.

[Kn19] Kneuper, R.: Messung und Bewertung von Prozessqualität mit Gokyo Ri. http://www.kneuper.de/GokyoRi/, Stand: 29.4.2020.

[Ma19] Martin-Jung, H.: Datenschutzgrundverordnung – Das verflixte erste Jahr. In: Süddeutsche Zeitung vom 25.05.2019. München, Süddeutsche Zeitung.

[Mc77] McCall, J. A.; Richards, P. K.; Walters, G. F.: Factors in Software Quality. US Rome Air Development Center Reports I-III. U.S. Department of Commerce, Washington, 1977.

[Mo15] Mohammad, M.: IT Surveillance and Social Implications in the Workplace. Proceedings of the 2015 SIGMIS Conference on Computers and People Research, 2015.

[Mo20] Moorstedt, M.: https://www.sueddeutsche.de/digital/home-office-ueberwachung-tracking-chef-zoom-1.4868739, SZ, 2020, Stand: 29.4.2020.

[Pi15] Piekarska, M. et.al.: Because we care: Privacy Dashboard on Firefox OS. In: Proceedings of the 9th Workshop on Web 2.0 Security and Privacy, 2015.

[Pr15] Pritchard, G. W. et.al.: How to Drive a London Bus: Measuring Performance in a Mobile and Remote Workplace, 33rd ACM CHI ’15, S. 907-916, 2015.

[Ra17] Raschke, P. et.al.: Designing a GDPR-compliant and usable privacy dashboard. In: IFIP International Summer School on Privacy and Identity Management, S. 221-236. Springer, Cham, 2017.

[RCA19] Reiserer, K.; Christ, F.; Heinz, K.: Beschäftigtendatenschutz und EU-Datenschutz-Grundverordnung, DStrR 2018, 1501.

[RS16] Rosenblatt, A.; Stark, L.: Algorithmic Labor and Information Asymmetries: A Case Study of Uber’s Drivers. Int. Journal of Communication, 16, S. 3758–3784, 2016.

[SHS19] Simitis, S.; Hornung, G.; Spiecker gen. Döhmann, I.: Datenschutzrecht, NOMOS, 2019

[SJ10] Scudder, J.; Jøsang, A.: Personal Federation Control with the Identity Dashboard. Policies and Research in Identity Management, IDMAN 2010, Oslo, November 18-19, 2010.

[SV20] Schulze, M.; Volk, T.: Die Anpassung von Betriebsvereinbarungen an die Betriebswirklichkeit, ArbR Aktuell 2020, 60.

[Sw02] Sweeney, L.: k-anonymity: a model for protecting privacy. International Journal on Uncertainty, Fuzziness and Knowledge-based Systems, Vol 10, Issue 5, 2002, S. 557-570.

[TG19] Taeger, J.; Gabel, D.: Kommentar DSGVO – BDSG. Deutscher Fachverlag GmbH, Frankfurt a.M., 3. Auflage, 2019.

[Th15] Thul, M. J.: Der sozio-technische Systemansatz. In (Zink, K. J. et al., Hrsg.): Veränderungsprozesse erfolgreich gestalten, 2. Aufl., Springer Vieweg, S. 278-283, 2015.

[To20] Tolsdorf, J. et.al.: Privatheit am Arbeitsplatz. DuD, Vol. 44, Nr. 3, S. 176-181, 2020.

[Tr20] TrUSD – Transparente und selbstbestimmte Ausgestaltung der Datennutzung im Unternehmen. https://www.trusd-projekt.de, Stand: 29.4.2020.

[Ul11] Ulich, E.: Arbeitspsychologie, 7. Auflage, Schäffel-Poeschel Verlag, 2011.

[W3C06] W3C, „The Platform for Privacy Preferences 1.1 (P3P1.1) Specification“, W3C Working Group Note, Nov. 2006.

[Wa02] Wallmüller, E.: Qualitätsmodelle im Software Engineering. In: MQ - Management und Qualität 2002(9). Galledia Verlag, Berneck, 2002.

[Wa12] Wagner, E.: Datenschutz als Bildungsauftrag. DuD, 02/12, S. 83-87, 2012.

[ZAM14] Zimmermann, C.; Accorsi, R.; Muller, G.: Privacy Dashboards: Reconciling Data-Driven Business Models and Privacy, ARES 2014, 2014, S. 152–157.

[Ze19] Zeit Online (2019): Zonar: Datenschutzbehörde prüft Mitarbeitersoftware von Zalando. URL: https://www.zeit.de/arbeit/2019-11/zonar-zalando-mitarbeiter-scoring-software, Stand: 29.4.2020.

Meldepflicht von IT-Sicherheits- und Datenschutzvorfällen durch Mitarbeitende - Betrachtung möglicher arbeitsrechtlicher Konsequenzen

Dirk Müllmann[1], Melanie Volkamer[2]

Abstract: Die Pflicht zur Meldung von IT-Sicherheits- und Datenschutzvorfällen in Unternehmen ist eine zentrale organisatorische Maßnahme zum Schutz von deren IT-Infrastruktur. Mitarbeiter offenbaren mit der Meldung des Vorfalls jedoch oftmals eigenes Fehlverhalten, das vom Arbeitgeber zur Grundlage arbeitsrechtlicher Konsequenzen gemacht und somit gegen sie verwandt werden kann. Die Angst vor diesen Konsequenzen kann Arbeitnehmer davon abhalten, der Meldepflicht nachzukommen und der Meldemoral im Unternehmen schaden. Das hat wiederum negative Konsequenzen für das Unternehmen selbst, dem es angesichts unterlassener Meldungen nicht möglich ist, schnell auf Vorfälle zu reagieren und sie effektiv einzudämmen. Der Beitrag untersucht vor dem Hintergrund der datenschutzrechtlichen Meldepflichten für Datenschutzverstöße die rechtlichen Grundlagen der arbeitsrechtlichen Mitteilungspflichten von Mitarbeitern. Er geht ferner auf die Frage der Einschlägigkeit des Selbstbelastungsverbots im arbeitsrechtlichen Kontext ein und analysiert die arbeitsrechtlichen Konsequenzen der Offenbarung von eigenem Fehlverhalten durch Arbeitnehmern bei der Erfüllung einer Mitteilungspflicht. Auf dieser Grundlage entwickelt er einen Vorschlag, wie die Verlässlichkeit der Meldung von IT-Sicherheits- oder Datenschutzvorfällen durch Mitarbeiter verbessert werden kann.

Keywords: Meldepflicht; Datenschutzvorfall; IT-Sicherheitsvorfall; Arbeitnehmer; Selbstbelastungsfreiheit; arbeitsrechtliche Konsequenzen

Ein adäquater Schutz der IT-Infrastruktur vor Cyberangriffen eines jeden Unternehmens bzw. einer jeden Organisation ist nur durch eine geeignete Kombination aus organisatorischen und technischen Maßnahmen möglich. Organisatorische Maßnahmen beinhalten in der Regel auch Security Policies[3] und Sensibilisierungsmaßnahmen. Hierbei geht es darum, wie Mitarbeitende Risiken erkennen und sich idealerweise verhalten, um die Risiken zu minimieren. Eine wichtige Komponente der organisatorischen Maßnahme ist das Melden von IT-Sicherheits- und Datenschutzvorfällen[4]. Eine Pflicht zum Melden ergibt sich aus

[1] Karlsruher Institut für Technologie, Kompetenzzentrum für Angewandte Sicherheitstechnologie (KASTEL), Zentrum für Angewandte Rechtswissenschaft (ZAR), Vincenz-Prießnitz-Straße 3, 76131 Karlsruhe, Germany, dirk.muellmann@kit.edu.

[2] Karlsruher Institut für Technologie, Kompetenzzentrum für Angewandte Sicherheitstechnologie (KASTEL), Institut für Angewandte Informatik und Formale Beschreibungsverfahren (AIFB), Kaiserstraße 89, 76133 Karlsruhe, Germany, melanie.volkamer@kit.edu.

[3] *Herath/Rao*, Encouraging information security behaviors in organizations: Role of penalties, pressures and perceived effectiveness, Decision Support Systems 47 (2009), 154, 157, https://dx.doi.org/10.1016/j.dss.2009.02.005.

[4] *Grispos/Glisson/Bourrie/Storer/Miller*, Security incident recognition and reporting (SIRR): an industrial perspective, 2017, arXiv preprint, https://arxiv.org/abs/1706.06818.

doi:10.18420/inf2020_74

dem Pflichtenkanon des Arbeitsverhältnisses selbst[5]. Dennoch haben viele Unternehmen und Organisationen eine Meldepflicht für IT-Sicherheits- bzw. Datenschutzvorfälle explizit eingeführt, z. B. in Form von Dienstanweisungen. Hier wird festgelegt, welche Vorfälle – in der Regel unmittelbar – an welchen Personenkreis zu melden sind. Die Meldepflicht ist aus einer Reihe von Gründen eine wichtige Komponente der organisatorischen Maßnahmen[6]:

So sind Cyberangriffe immer schwerer zu erkennen. Außerdem liegt es in der Natur der Menschen, dass sie Fehler machen und daher einen Angriff übersehen oder unvorsichtig mit personenbezogenen Daten umgehen. Es kann also auch bei einer idealen Kombination von organisatorischen und technischen Maßnahmen[7] und für das Thema Sicherheit und Datenschutz hoch sensibilisierten Mitarbeitenden nicht davon ausgegangen werden, dass es keine IT-Sicherheitsvorfälle gibt[8]. Das zeitnahe Melden von IT-Sicherheitsvorfällen ist daher wichtig, um die Schäden eines erfolgreichen Cyberangriffs sowie die Kosten der Schadensbehebung, aber auch rechtliche Konsequenzen und den Imageschaden so gering wie möglich zu halten. Durch zeitnahes Melden, können Experten die Situation auch zeitnah nach dem Angriff untersuchen und technische Schutzmaßnahmen ergreifen. Außerdem werden die Verantwortlichen so rechtzeitig informiert und können entsprechende organisatorische Maßnahmen umsetzen. Im Fall eines Datenschutzvorfalls gibt es zudem gesetzliche Vorgaben, die von den Unternehmen bzw. den Organisationen eine Meldung innerhalb vorgegebener Fristen verlangen. So sieht Art. 33 Absatz 1 S. 1 DSGVO zum Beispiel in diesen Fällen eine Meldung binnen 72 Stunden an die gemäß Art. 55 DSGVO zuständige Aufsichtsbehörde vor. Eine vergleichbare Pflicht existiert gemäß § 8b Abs. 4 BSIG auch für Betreiber kritischer Infrastrukturen im Fall von Störungen der Verfügbarkeit, Integrität, Authentizität und Vertraulichkeit ihrer informationstechnischen Systeme.

Eine offensichtliche Voraussetzung dafür, dass Mitarbeitende IT-Sicherheits- und Datenschutzvorfälle melden, ist das diese durch die verschiedenen organisatorischen Maßnahmen wissen, was ein IT-Sicherheits- bzw. Datenschutzvorfall ist, dass sie diese melden sollen und dass es wichtig ist diese zu melden, um den Schaden für das Unternehmen bzw. die Organisation so gering wie möglich zu halten[9]. Dieses Wissen wird aber nicht zwangsläufig dazu führen, dass jeder IT Sicherheitsvorfall gemeldet wird: Mitarbeitende müssen sich in der Regel mit dem Melden eines IT-Sicherheitsvorfalls einen Fehler eingestehen und gegebenenfalls zumindest indirekt zugeben, dass sie sich nicht an die Security Policies gehalten haben. Dies kann sie vom Melden eines Vorfalls abhalten. Insbesondere Mitarbeitende, die

[5] Vgl. hierzu Kap. 1.2.

[6] *Jaatun/Albrechtsen/Bartnes et al.*, A Study of Information Security Practice in a Critical Infrastructure Application. In: Rong/Jaatun/Sandnes et al. (Hrsg.) Autonomic and Trusted Computing, ATC 2008 Lecture Notes in Computer Science, vol 5060, 527, 527 ff.

[7] *Werlinger/Hawkey/Beznosov*, An integrated view of human, organizational, and technological challenges of IT security management, Information Management & Computer Security, 17 (2009), 4, 4 ff., https://doi.org/10.1108/09685220910944722.

[8] *Dutta/Roy*, Dynamics of organizational information security. System Dynamics Review: The Journal of the System Dynamics Society, 24 (2008), 349, 349 ff., https://onlinelibrary.wiley.com/doi/abs/10.1002/sdr.405.

[9] *Humphrey*, Identifying the critical success factors to improve information security incident reporting, 2017, https://dspace.lib.cranfield.ac.uk/handle/1826/12739.

Angst vor persönlichen Konsequenzen haben, werden Vorfälle weniger zuverlässig melden. Dabei können sie sowohl Angst vor dem eigenen Imageschaden als auch vor rechtlichen Konsequenzen haben. Eine wesentliche Rolle spielt hierbei die allgemeine Fehlerkultur im Unternehmen bzw. der Organisation[10].

Im Vordergrund der vorliegenden Untersuchung steht die Analyse der persönlichen arbeitsrechtlichen Konsequenzen im Zusammenhang mit der Wahrnehmung von Meldepflichten. Aus Sicht der IT-Sicherheit könnte man denken, dass es am besten wäre, die Meldepflicht mit der eindeutigen Aussage zu verknüpfen, dass das Melden eines Vorfalls keine persönlichen rechtlichen Konsequenzen hat, wohl aber dessen Nichtmeldung. Das wäre aber zu kurz gegriffen, weil eine solche Aussage dazu führen würde, dass Mitarbeitenden sich an keine Security Policies mehr halten müssen, solange sie die Vorfälle melden. Das wiederum würde die organisatorischen und gegebenenfalls auch technischen Schutzmaßnahmen aushebeln.

Ziel dieses Beitrags ist es einen Vorschlag zu erarbeiten, wie arbeitsrechtlich mit Vorfällen umgegangen werden könnte, um möglichst wenige Mitarbeitenden davon abzuhalten Vorfälle zu melden, gleichzeitig Mitarbeiter zu motivieren sich möglichst an die Security Policies zu halten. Dazu wird zunächst die rechtliche Ausgangslage analysiert.

1 Rechtliche Ausganslage

Gesetzliche Meldepflichten für den Fall von IT-Sicherheits- oder Datenschutzverstößen existieren für eine Vielzahl unterschiedlicher Situationen. Dabei unterscheiden sich die gesetzlichen Grundlagen und Ausgestaltungen der Meldepflicht von Institutionen oder Unternehmen an Behörden je nach Anwendungsfall und Regelungsgebiet. Die Meldepflicht in der darunterliegenden Ebene, also dem Verhältnis zwischen Arbeitnehmer und Arbeitgeber, beruht hingegen immer auf denselben arbeitsrechtlichen Normen. Angesichts der Fokussierung des Beitrags auf die arbeitsrechtlichen Konsequenzen von Meldepflichten für Mitarbeitende sollen die Meldepflichten von Unternehmen und Institutionen, die ein Auslöser für Meldepflichten von Mitarbeitern sein können, lediglich exemplarisch anhand Art. 33 DSGVO dargestellt werden. Auf die in anderen Bereichen ebenfalls bestehenden unternehmerischen Meldepflichten, wie zum Beispiel bei IT-Sicherheitsverstößen kritischer Infrastrukturen, sei an dieser Stelle jedoch ausdrücklich verwiesen.

1.1 Die Meldepflicht für Datenschutzvorfälle von Unternehmen und Organisationen als Beispiel gesetzlicher Meldepflichten an Aufsichtsbehörden

Um die Informationslage der Aufsichtsbehörden zu verbessern, geeignete Maßnahmen zum Schutz der Betroffenen einzuleiten sowie die Rechtsdurchsetzung und -befolgung

[10] *Werlinger/Hawkey/Beznosov*, An integrated view of human, organizational, and technological challenges of IT security management, Information Management & Computer Security, 17 (2009), 4, 4 ff., https://doi.org/10.1108/09685220910944722.

datenschutzrechtlicher Normen zu steigern,[11] verlangt die Datenschutzgrundverordnung die Meldung der Verletzung des Schutzes von personenbezogenen Daten an die Aufsichtsbehörde. Art. 4 Nr. 12 DSGVO definiert eine solche Schutzverletzung als "*eine Verletzung der Sicherheit, die zur Vernichtung, zum Verlust oder zur Veränderung, ob unbeabsichtigt oder unrechtmäßig, oder zur unbefugten Offenbarung von beziehungsweise zum unbefugten Zugang zu personenbezogenen Daten führt, die übermittelt, gespeichert oder auf sonstige Weise verarbeitet wurden*". Art. 33 Absatz 1 S. 1 DSGVO sieht in diesen Fällen eine Meldung der Verletzung binnen, in der Regel, 72 Stunden an die gemäß Art. 55 DSGVO zuständige Aufsichtsbehörde vor.

Durch diese gesetzliche Verpflichtung sieht sich der Verarbeiter einer Situation ausgesetzt, in der er die Behörden gegebenenfalls über ein Fehlverhalten seinerseits informieren muss, das in der Folge als sachliche Grundlage für eine Sanktionierung in einem Buß- oder Strafverfahren genutzt werden könnte. Obwohl die Nichtbefolgung der Meldepflicht gemäß Art. 83 Abs. 4 lit. a) DSGVO ebenfalls bußgeldbewehrt ist,[12] könnte die Verpflichtung des Verarbeiters, sich in einem Verfahren selbst zu belasten, daher dennoch im Konflikt mit dem Selbstbezichtigungsverbot des "nemo tentur"-Grundsatzes stehen.[13] Vor diesem Hintergrund hat der deutsche Gesetzgeber in den §§ 42 Abs. 4, 43 Abs. 4 BDSG ein Verwertungsverbot für die Meldungen und Benachrichtigungen von Datenverarbeitern in Ordnungswidrigkeiten- und Strafverfahren vorgesehen, sodass eine Verwendung in einem Verfahren nur mit dessen Zustimmung erfolgen dürfte. Die Europarechtskonformität dieser Regelung ist jedoch umstritten.[14] Damit die Unternehmen der Meldepflicht von Schutzverletzungen personenbezogener Daten an die Aufsichtsbehörden nachkommen können, sind sie dringend auf die Mitwirkung ihrer Mitarbeitenden angewiesen. Durch das Melden eines Datenschutzvorfalls offenbaren sie aber möglicherweise eigenes Fehlverhalten, das zu der Schutzpflichtverletzung geführt hat und arbeitsrechtlich sanktioniert werden könnte. Das könnte sie davon abhalten, Sicherheitsvorfälle, insbesondere solche, die von ihnen verschuldet wurden, zu melden, was wiederum eine frühzeitige Gegenreaktion und Maßnahmen des Arbeitgebers und Datenverarbeiters erschwert. Eine dem Verwertungsverbot in §§ 42 Abs. 4, 43 Abs. 4 BDSG analoge Regelung für Mitarbeitende besteht im BDSG jedoch nicht. Es wird zwar im Rahmen der Straf- und Bußgeldvorschriften eine analoge Erweiterung des Verwertungsverbots auf Personen, die für andere handeln, im Sinne der §§ 9, 14 OWiG erwogen.[15] Die Konsequenzen einer etwaigen Selbstbelastung von Arbeitnehmern

[11] *Martini* in: Paal/Pauly (Hrsg.), DS-GVO / BDSG, 2. Aufl., 2018, Art. 33 DSGVO, Rn. 10; *Dix* in: Simitis/Hornung/Spiecker gen. Döhmann (Hrsg.), DSGVO, 2019, Art. 33, Rn. 1.

[12] *Hladjk* in: Ehmann/Selmayr (Hrsg.), DS-GVO, 2. Aufl., 2018, Art. 33, Rn. 22.

[13] *Spittka*, Si Tacuisses. . . - Nemo Tenetur und die DSGVO, in: Taeger (Hrsg.), Die Macht der Daten und der Algorithmen, 2019, 141, 144; *Reif* in: Gola (Hrsg.), DSGVO, 2. Aufl., 2018, Art. 33, Rn. 44.

[14] *Brink* in: Wolff/Brink (Hrsg.), Beck'scher Onlinekommentar Datenschutzrecht, Art. 33, Rn.15; *Spittka*, Si Tacuisses. . . - Nemo Tenetur und die DSGVO, in: Taeger (Hrsg.), Die Macht der Daten und der Algorithmen, aaO., 141, 150ff.; *Paal*, ZD 2020, 119, 124; *Bergt* in: Kühling/Buchner (Hrsg.), DSGVO / BDSG, 2. Aufl., 2018, § 43 BDSG, Rn. 11 ff.; *Spittka*, RDV 2019, 167, 170 ff.

[15] *Brodowski/Nowak* in: Beck'scher Onlinekommentar Datenschutzrecht, 31. Ed., 2020, § 43 BDSG, Rn. 25, 27; *Boms*, ZD 2019, 536, 539 f.

durch die Meldung von Schutzverletzungen werden momentan aber weder in Straf- oder Bußgeldverfahren noch auf zivil- und arbeitsrechtlicher Ebene gesetzlich adressiert.

1.2 Die arbeitsrechtliche Benachrichtigungspflicht

Arbeitnehmer treffen gegenüber ihren Arbeitgebern sowohl Auskunfts-[16] als auch Benachrichtigungspflichten. Beide unterscheiden sich dadurch, dass einem Auskunftsanspruch nur auf Anforderung nachgekommen werden muss,[17] während bei Benachrichtigungspflichten alle erforderlichen Informationen unaufgefordert mitzuteilen sind[18]. Die rechtliche Grundlage zur Herleitung einer nicht explizit vereinbarten arbeitsrechtlichen Benachrichtigungspflicht des Arbeitnehmers gegenüber seinem Arbeitgeber stellen die arbeitsvertraglichen Nebenpflichten dar.[19] Die Informationsbeschaffung des Arbeitsgebers ist sowohl auf der Basis eines Auskunfts- als auch einer Mitteilungspflicht denkbar. Vorliegend dürfte dem Benachrichtigungsanspruch jedoch größere Bedeutung zukommen. Da es erforderlich ist, von IT-Sicherheits- bzw. Datenschutzvorfällen bereits zu erfahren, wenn sie noch nicht nach außen getreten und für andere sichtbar geworden sind, muss auch die Meldepflicht für IT-Sicherheits- und Datenschutzvorfällen durch Mitarbeitende unaufgefordert wahrgenommen werden. Anders ist das Ziel schnell auf Vorfälle reagieren zu können für Unternehmen und Organisationen nicht zu erreichen. Die Mitteilungspflicht des Arbeitnehmers kann in diesem Fall mit seiner Verantwortung begründet werden, das Integritätsinteresse seines Arbeitgebers zu wahren.[20] Daraus ergibt sich für ihn die Verpflichtung, im Vorfeld einer Schädigung zu handeln und drohende Störungen und Schäden an Betriebsmitteln zur Kenntnis zu bringen.[21] Die Annahme einer Benachrichtigungspflicht stellt in diesen Fällen ferner den einzigen

[16] Die Herleitung eines arbeitsrechtlichen Auskunftsanspruchs ist umstritten. Insbesondere in der Rechtsprechung wird vertreten, dass er als Nebenpflicht zum Arbeitsvertrag § 242 BGB entspringt (so BAG, Urt. v. 07.09.1995, 8 AZR 828/93, BAGE 81, 15; LAG Hamm, Urt. v. 03.03.2009, 14 Sa 1689/08, CCZ 2010, 237, 238; ArbG Saarlouis, Urt. v. 19.10.1983, 1 Ca 493/83, ZIP 1984, 364; wohl auch *Lützler/Müller-Sartori*, CCZ 2011, 19, 19f.), während die Literatur den Anspruch oftmals auf §§ 666, 675 BGB stützt (*Dann/Schmidt*, NJW 2009, 1851, 1852f. Mit Differenzierung, zwischen dem unmittelbaren und mittelbaren Arbeitsbereich des Arbeitnehmers; ebenso *Spehl/Momsen/Grützner*, CCZ 2014, 170, 171).

[17] *Fischer* in: Bamberger/Roth/Hau/Poseck (Hrsg.), Beck'scher Onlinekommentar BGB, 53. Ed., 2020, §666, Rn.5; BGH, Urt. v. 16.6.2016, III ZR 282/14, Rn. 37, NJW-RR 2016, 1391, 1394.

[18] *Fischer* in: Bamberger/Roth/Hau/Poseck (Hrsg.), Beck'scher Onlinekommentar BGB, 53. Ed., 2020, §666, Rn.3; *Schäfer* in: Münchener Kommentar zum BGB, 8. Aufl., 2020, §666, Rn. 22.

[19] *Preis* in: Erfurter Kommentar zum Arbeitsrecht, 20. Aufl., 2020, § 611a BGB, Rn. 736; *Spinner* in: Münchener Kommentar zum BGB, 8. Aufl., 2020, §611a, Rn. 993, 1030; *Joussen* in: Beck'scher Onlinekommentar Arbeitsrecht, 54. Ed., 2019, §611a BGB, Rn.446; *Reichold* in: Münchener Handbuch zum Arbeitsrecht, Band I, 4. Aufl., 2018, §55, Rn. 4, 8.

[20] *Joussen* in: Beck'scher Onlinekommentar Arbeitsrecht, 54. Ed., 2019, §611a BGB, Rn.446.

[21] *Joussen* in: Beck'scher Onlinekommentar Arbeitsrecht, 54. Ed., 2019, §611a BGB, Rn.446; *Preis* in: Erfurter Kommentar zum Arbeitsrecht, 20. Aufl., 2020, § 611a BGB, Rn. 742.

Weg zur Erfüllung der ebenfalls den Arbeitnehmer treffenden Schadensminderungspflicht[22] gegenüber dem Arbeitgeber dar.[23]

1.3 Arbeitsrechtliche Konsequenzen der Meldung von IT-Sicherheits- und Datenschutzvorfällen vor dem Hintergrund der Selbstbelastungsfreiheit

Sofern ein IT-Sicherheitsvorfall auf eine Pflichtverletzung, z.B. die Nicht-Einhaltung einer Security Policy, durch einen Arbeitnehmer zurückgeht, kann ihm ein Fehlverhalten vorgeworfen werden, das arbeitsrechtliche Konsequenzen nach sich ziehen kann. Beispiele hierfür können z.B. in der Interaktion mit Phishing-E-Mails, der unbefugten (ggf. zunächst unbewussten) Weitergabe von Daten an Unbefugte, dem (längerem) Unterlassen von Updates oder der Nicht-Nutzung vorgeschriebener Sicherheitsmaßnahmen gesehen werden. Während besonders schwere Verstöße eine ordentliche oder gar fristlose Kündigung[24] zur Folge haben können, wird gerade bei einmaligen oder nur leicht fahrlässig begangenen Sicherheitsverletzungen durch Mitarbeiter aber nur eine Abmahnung in Betracht kommen.[25]

Indem der Arbeitnehmer in diesen Fällen die verschuldete Verletzung meldet, liefert er dem Arbeitgeber zugleich die Grundlage für arbeitsrechtliche Maßnahmen gegen ihn selbst. Dieser Umstand könnte im Widerspruch zur Selbstbelastungsfreiheit gemäß dem nemotentur-Grundsatz stehen, der verfassungsrechtlich aus dem Grundrecht auf informationelle Selbstbestimmung nach Art. 2 Abs. 1 GG i.V.m. Art. 1 Abs. 1 GG oder ergänzend aus dem Recht auf ein faires und rechtsstaatliches Verfahren nach Art. 20 Abs. 3 GG bzw. Art. 6 Abs. 1 EMRK abgeleitet wird.[26] Er ist zudem sowohl in Art. 48 Abs. 2 der Charta der Grundrechte der Europäischen Union als auch in Art. 6 Abs. 3 EMRK als Verteidigungsrecht vorgesehen.[27]

Die Selbstbelastungsfreiheit schützt nach herrschender Ansicht nur vor staatlich veranlasstem Aussagezwang in staatlichen Verfahren, insbesondere Straf- und Ordnungsverfahren.[28] Die vom Bundesverfassungsgericht aufgestellten Grundsätze zum Schutz gegen Selbstbezichtigung und daraus resultierende strafrechtliche Konsequenzen beschränken sich aber

[22] *Preis* in: Erfurter Kommentar zum Arbeitsrecht, 20. Aufl., 2020, § 611a BGB, Rn. 744ff.; *Spinner* in Münchener Kommentar zum BGB, 8. Aufl., 2020, §611a, Rn. 1001; BAG, Urt.v. 01.06.1995, 6 AZR 912/94, NZA 1996, 135, 136.

[23] Vgl. auch *Joussen* in: Beck'scher Onlinekommentar Arbeitsrecht, 54. Ed., 2019, §611a BGB, Rn.446; *Reichold* in: Münchener Handbuch zum Arbeitsrecht, Band I, 4. Aufl., 2018, §55, Rn. 8.

[24] Vgl. nur ArbG Siegburg, Urt. v. 15.01.2020, 3 Ca 1793/19.

[25] *Fuhlrott*, NZA 2019, 649, 650, 652f.; *Niemann* in: Erfurter Kommentar, 20. Aufl, 2020, §626 BGB, Rn. 29f.

[26] BVerfG, Beschl. v. 13.01.1981, 1 BvR 116/77, Rn. 18, NJW 1981, 1431, 1432; BVerfGE 38, 105, 113; BGHSt 14, 358, 264, BGH, NJW 1989, 1228, 1229; EGMR, NJW 2002, 499, 501.

[27] EuGH, Urt. v. 07.01.2004, C-204/00 (Alborg), Slg. 2004,I-123, Rn.64f.; Urt.v.25.10.2001, Slg. 2002, I-8275, Rn. 273f.;*Jarass*, Charta der Grundrechte der EU, 3. Aufl., 2016, Art. 48, Rn. 31; EGMR, Urt. v. 08.04.2004, 38544/97, Rn. 46, JR 2005, 423.

[28] BVerfG,Beschl. v. 13.01.1981, 1 BvR 116/77, Rn. 18f., NJW 1981, 1431, 1432; *Wessing* in: Hauschka/Moosmeyer/Lösler, Corporate Compliance, 3. Aufl., 2016, §46, Rn. 50; *Schaefer*, NJW-Spezial, 2010, 120.

nicht auf diese Verfahren.[29] Sie entfalten nach Ansicht der herrschenden Meinung jedoch keine Wirkung gegenüber dem Arbeitgeber, da hier kein staatlicher Aussagezwang gegeben sei.[30] Zur Begründung hierfür wird angeführt, dass es dem Arbeitnehmer frei stehe, sich zu äußern, sodass er im Fall einer Äußerung auch mit deren Konsequenzen leben müsse.[31] Dieser Argumentation ist vor dem Hintergrund der Annahme einer arbeitsvertraglichen Benachrichtigungspflicht jedoch nicht zuzustimmen. Wenn eine solche Pflicht besteht und sanktioniert werden kann, steht dem Arbeitnehmer eine Äußerung gerade nicht frei. In Bezug auf etwaige strafrechtliche Konsequenzen einer Mitteilung an den Arbeitgeber ist daher der in der Literatur vertreten Ansicht zuzustimmen, dass die vom Verfassungsgericht in der Gemeinschuldnerentscheidung aufgestellten Grundsätze[32] auch auf Äußerungen gegenüber dem Arbeitgeber übertragen werden müssen[33]. Für sie gilt daher ein strafrechtliches Verwertungsverbot.

Auf die im vorliegenden Beitrag untersuchte Mitteilungspflicht gegenüber dem Arbeitgeber und die aus ihr für den Arbeitnehmer resultierenden arbeitsrechtlichen Konsequenzen hat das jedoch keinen Einfluss. Da der nemo-tenetur-Grundsatz nur in staatlichen Verfahren und gegenüber staatlichen Organen, insbesondere mit Bezug zum Strafrecht gilt, ist er auf das privatrechtliche Verhältnis zwischen Arbeitnehmer und -geber nicht direkt anwendbar.[34] In der Situation muss zwar eine Abwägung zwischen einem berechtigten, billigens- und schützenswerten Interesse des Arbeitgebers auf Information und dem Interesse des Arbeitnehmers vorgenommen werden, ein Fehlverhalten nicht zuzugeben und sich nichts selbst belasten zu müssen.[35] Eine Benachrichtigungspflicht im Zusammenhang mit Datenschutz- und IT-Sicherheitsverstößen wird dabei im Ergebnis jedoch regelmäßig zu bejahen sein. Für sie streiten sowohl die Schäden, die dem Arbeitgeber ohne die Erfüllung der Informationspflicht drohen, als auch die datenschutzrechtliche Pflicht zur Meldung von Sicherheitsverstößen, mit der auch die Interessen der Datenobjekte gewahrt werden. Auch die meist schuldhafte Verursachung eines Verstoßes durch den Arbeitnehmer spricht eher für die Annahme einer Mitteilungspflicht. Gegen sie können lediglich die im Vergleich dazu weniger gravierenden arbeitsrechtlichen Konsequenzen für den Arbeitnehmer ins Feld geführt werden.

[29] BVerfG, Beschl. v. 13.01.1981, 1 BvR 116/77, Rn. 19, NJW 1981, 1431; *Wessing* in: Hauschka/Moosmeyer/Lösler, Corporate Compliance, 3. Aufl., 2016, §46, Rn. 50.

[30] OLG Karlsruhe, Beschl. v. 06.09.1988, 1 Ss 68/88, NStZ 1989, 287, 288; *Wessing* in: Hauschka/Moosmeyer/Lösler, Corporate Compliance, 3. Aufl., 2016, §46, Rn. 50; *Bittmann/Molkenbur*, wistra 2009, 68.

[31] OLG Karlsruhe, Beschl. v. 06.09.1988, 1 Ss 68/88, NStZ 1989, 287, 288. Anders im Fall des Gemeinschuldners, der nach der zum Zeitpunkt der Gemeinschuldnerentscheidung des Verfassungsgerichts geltenden Rechtslage gemäß § 100 KO zur Auskunft verpflichtet war: BVerfG, Beschl. v. 13.01.1981, 1 BvR 116/77, Rn. 26 f., NJW 1981, 1431.

[32] BVerfG, Beschl. v. 13.01.1981, 1 BvR 116/77, Rn. 26 f., NJW 1981, 1431, 1432.

[33] *Wessing* in: Hauschka/Moosmeyer/Lösler, Corporate Compliance, 3. Aufl., 2016, §46, Rn. 50, 56; *Schrader/Thoms/Mahler*, NZA 2018, 965, 969; *Dann/Schmidt*, NJW 2009, 1851; 1855; LAG Hamm, Urt. v. 03.03.2009, 14 Sa 1689/08, CCZ 2010, 237.

[34] *Spehl/Momsen/Grützner*, CCZ 2014, 170, 171; *Dann/Schmidt*, NJW 2009, 1851, 1855; *Lützeler/Müller-Satori*, CCz 2001, 19, 20.

[35] Vgl. *Dann/Schmidt*, NJW 2009, 1851, 1853; *Spehl/Momsen/Grützner*, CCZ 2014, 170, 171.

Nach aktueller Rechtslage ist daher davon auszugehen, dass die arbeitsrechtliche Sanktionierung des Fehlverhaltens eines Arbeitnehmers rechtmäßig wäre, selbst wenn das Fehlverhalten nur aufgrund seiner selbstbelastenden Mitteilung vom Arbeitgeber erkannt werden konnte. Nur einer strafrechtlichen Verwertung der Meldung steht ein von der Rechtsprechung entwickeltes Verbot entgegen. Darüber hinaus kann angesichts des Bestehens der arbeitsrechtlichen Nebenpflicht zur Benachrichtigung und Schadensabwendung auch das Unterlassen der Meldung eines Verstoßes durch den Arbeitnehmer vom Arbeitgeber arbeitsrechtlich sanktioniert werden.[36] Je nach Schwere eines solchen Verstoßes gegen die Mitteilungspflicht kommen zur Sanktionierung der Nichtmeldung durch den Arbeitnehmer ebenfalls gestufte Maßnahmen von der Abmahnung bis zur fristlosen Kündigung in Betracht.[37] Die Beweislast für den Nachweis eines Fehlverhaltens, dessen Schwere und die Rechtmäßigkeit der darauf basierenden Sanktionen des Arbeitnehmers trifft den Arbeitgeber.[38] In Fällen des Unterlassens der Meldung eines IT-Sicherheitsverstoßes hat er nachzuweisen, dass dem Arbeitnehmer ein meldepflichtiger Vorfall bekannt war oder im Rahmen einer ordnungsgemäßen Aufgabenerfüllung hätte bekannt sein müssen und er ihn pflichtwidrig nicht mitgeteilt hat.[39] Sofern ein Arbeitnehmer einen IT-Sicherheitsvorfall meldet und der Arbeitgeber ihn deshalb aufgrund eines vermuteten Fehlverhaltens sanktionieren möchte, muss er demnach aber auch nachweisen können, dass der IT-Sicherheitsvorfall tatsächlich auf einem Fehlverhalten des Arbeitnehmers, z.B. in Bezug auf die Security Policies, beruht. Das Erfüllen dieser Beweispflicht ist angesichts der vielfältigen Quellen für diese Verstöße und der damit uneindeutigen Beweislage nicht immer einfach oder überhaupt möglich. Dies zeigen auch die folgenden Beispiele.

2 Beispielhafte IT-Sicherheits- bzw. Datenschutzvorfälle und die Problematik der Beweislage

Es gibt Datenschutz- und IT-Sicherheitsrechtsvorfälle, die sich eindeutig einem konkreten Verhalten des Arbeitnehmers zuordnen lassen, und andere bei denen das weniger einfach möglich ist. Hier beispielhafte Fälle:

Fall 1: Ein Arbeitnehmer meldet, dass das Passwort seines Accounts für das Unternehmen bzw. die Institution in einer veröffentlichten Datenbank, wie der des Hasso-Plattner-Instituts[40], enthalten ist. Dies kann viele Ursachen haben. So ist die Passwortspeicherung

[36] Vgl. nur *Preis* in: Erfurter Kommentar, 20. Aufl., 2020, §611a, Rn. 748.

[37] *Preis* in: Erfurter Kommentar, 20. Aufl., 2020, §611a, Rn. 748.

[38] BAG, Urt. v. 11.03.1987, 5 AZR 739/85, NZA 1987, 452; Urt. v. 13.03.1987, 7 AZR 601/85, NZA 1987, 518, LAG M-V., Urt. v. 11.02.2020, 2 Sa 133/19, Rn. 36 ff.; LAG Köln, Urt. v. 17.01.2007, 7 Sa 526/06, I. 2. e) bb) aaa); *Schmidt* in: Küttner (Hrsg.), Personalbuch 2020, 27. Aufl., 2020, Stichwort Abmahnung, Rn. 42; *Weizenegger* in: Bredemeier/Neffke, TVöD/TV-L, 5. Aufl., 2017, Vorb. §34 TVöD, Rn. 641f.; *Niemann* in: Erfurter Kommentar, 20. Aufl., 2020, §626, Rn. 234.

[39] Vgl. *Joussen* in: Beck'scher Onlinekommentar Arbeitsrecht, 54. Ed., 2019, §611a BGB, Rn. 446; *Preis* in: Erfurter Kommentar, 20. Aufl., 2020, §611a, Rn. 736.

[40] https://sec.hpi.de/ilc/search?lang=de.

beim Arbeitgeber möglicherweise nicht sicher erfolgt oder der Arbeitnehmer das hat das gleiche Passwort auch für andere Accounts genutzt, bei denen es nicht sicher gespeichert wurde. Auch ist es möglich, dass der Arbeitnehmer auf eine Phishing-Nachricht, die unterschiedlich einfach oder schwer zu erkennen ist, reingefallen ist oder ihm jemand bei der Eingabe des Passworts am Laptop in der Bahn über die Schulter geschaut hat.

Fall 2: Ein Arbeitnehmer meldet, dass ein USB-Stick, auf dem personenbezogene Daten unverschlüsselt gespeichert werden, verloren gegangen ist. Auch hier kommen verschiedene Ursachen in Betracht. Der USB-Stick kann zum Beispiel gestohlen worden sein, während man in ein Gespräch verwickelt wurde (Teil eines Social Engineering Angriffs und dadurch schwer zu verhindern), oder er wurde im Café auf dem Weg zum Vorgesetzen, der um diese Daten gebeten hat, vergessen.

In diesem Zusammenhang ist ebenfalls darauf hinzuweisen, dass einige Ursachen zwar theoretisch als Fehlverhalten des Arbeitnehmers interpretiert werden könnten, sich im Einzelfall aber die Frage stellt, ob der Arbeitgeber seine Arbeitnehmer für das entsprechende Fehlverhalten ausreichend sensibilisiert hat. Hätte der Arbeitnehmer es also besser wissen können und standen ihm die erforderlichen Tools zur Verfügung? Oftmals reicht zum Beispiel die bloße Aufforderung zur Löschung von Phishing-E-Mails nicht aus, da nicht verlangt werden kann, jede dieser Mails zu erkennen (insbesondere wenn entsprechende Schulung nicht angeboten werden). Außerdem kann in der Regel nicht erwartet werden, dass alle Arbeitnehmer in der Lage sind, sich selbst mit einer Verschlüsselungssoftware vertraut zu machen, um Daten auf dem USB-Stick verschlüsselt zu speichern.

3 Vorschlag für Begrenzung des Verantwortungsmaßstabs zur Stärkung der IT-Sicherheit und des Datenschutzes

Die drohenden arbeitsrechtlichen Konsequenzen der Meldung eines IT-Sicherheitsverstoßes können dazu führen, dass der Arbeitnehmer entgegen seiner Verpflichtung einen solchen Vorfall nicht meldet, versucht, ihn zu verschleiern, oder so weit aufzuschieben, dass er seinem Fehlverhalten nicht mehr zugeordnet werden kann. Dies gilt umso mehr als auch die Nichtmeldung eines Vorfalls einen sanktionsfähigen Verstoß gegen seine arbeitsrechtlichen Pflichten darstellt. Die praktische Folge dieses Vorgehens des Arbeitnehmers ist, dass der Sicherheitsvorfall nicht in seinen Anfängen bekämpft und eingegrenzt werden kann und sich immer weiter ausbreitet. Dem Arbeitgeber und dem Datenschutz ist damit nicht geholfen - im Gegenteil. Auch, wenn er den Mitarbeiter bei so einem Fehlverhalten strenger sanktionieren kann, eine Verhinderung oder zumindest Verminderung des Schadens wäre für ihn wünschenswerter.

Um dieses Ziels mittels einer verbesserten Meldemoral von Sicherheitsvorfällen erreichen zu können, erscheinen verschiedene Maßnahmen denkbar. Zunächst könnte in Unternehmen ein Meldesystem etabliert werden, dass eine pseudonymisierte oder anonymisierte Meldung von Schutzverletzungen erlaubt. Arbeitnehmer könnten den zuständigen Stellen im Betrieb

Vorfälle auf diese Weise zur Kenntnis bringen, ohne sich namentlich dazu bekennen zu müssen und direkte arbeitsrechtliche Konsequenzen befürchten zu müssen. Ein solches Meldesystem wird inzwischen auch in vielen Unternehmen zum Schutz von Whistleblowern bei der Bekämpfung von firmeninternen Missständen angewandt.[41] Die Mitarbeiter müssten der Anonymität eines solchen Meldesystem jedoch unbedingt vertrauen, da sie bei Bekanntwerden ihrer Identität auch weiterhin mit arbeitsrechtlichen Konsequenzen rechnen müssten. Insofern wäre eine anonyme Meldung auch nur bei Verstößen sinnvoll, die nicht von ohnehin einer Person zugeordnet werden können. Für ein pseudonymes Meldesystem gelten diese Einschränkungen umso mehr, als dort die Daten des Arbeitnehmers, wenn auch nicht direkt, weiterhin zuordenbar wären. Darüber hinaus erscheint fraglich, ob es überhaupt wünschenswert ist, einen Arbeitnehmer vollständig von der Pflicht zu entbinden, sich zu einem Fehlverhalten bekennen zu müssen. Sofern ein Verstoß, z.B. gegen Security Policies, keine Konsequenzen für ihn hat, weil er ihm nicht zugeordnet werden kann, existiert auch kein Anreiz, sich an die bestehenden Regelungen zu halten. Für das Ziel einer Verbesserung des Datenschutzes und der IT-Sicherheit wäre das sogar kontraproduktiv.

Dasselbe Argument ließe sich auch für die Einführung einer Bagatellgrenze bei Sicherheitsverstößen anführen. Sofern man kleine Verstöße gegen die Schutzvorschriften generell nicht ahnden würde, bestünde kein Anreiz, ein solches Fehlverhalten zu vermeiden. Es kommt in diesem Zusammenhang hinzu, dass die Schwere der Auswirkungen eines Verstoßes gegen IT-Sicherheits- oder Datenschutzregeln sich oftmals erst im Nachhinein offenbaren. Dieselbe Bagatellhandlung, wie ein Klick auf den Anhang einer Mail eines unbekannten Absenders, kann entweder vom Virenschutzprogramm aufgehalten werden oder die Systeme eines ganzen Betriebes lahmlegen. Ob eine Sanktionierung erfolgen würde, wäre beim Vorliegen einer Bagatellgrenze damit vom Zufall abhängig.

Zur besseren Durchsetzung der Meldepflicht, der effektiveren und effizienteren Bekämpfung von Sicherheitsvorfällen sowie zur Stärkung des Datenschutzes und der IT-Sicherheit sollte daher eine teilweise Abkehr vom Gedanken der arbeitsrechtlichen Sanktion erwogen werden. Ein Ansatzpunkt wäre hierbei die Übertragung der Grundsätze der Arbeitnehmerhaftung auf die arbeitsrechtlichen Sanktionen eines vom Arbeitnehmer verursachten und gemeldeten Sicherheitsverstoßes.

Nach den von der Rechtsprechung entwickelten Grundsätzen des innerbetrieblichen Schadensausgleichs haftet der Arbeitnehmer bei betrieblich veranlassten Schäden nur bei Vorsatz und grober Fahrlässigkeit in vollem Umfang, bei mittlerer Fahrlässigkeit nur anteilig und bei leichtester Fahrlässigkeit gar nicht.[42] Dies gilt im Übrigen auch für Schäden, die dem Arbeitgeber oder Dritten durch IT-Sicherheitsvorfälle entstehen, die ein Arbeitnehmer schuldhaft verursacht hat.

Dem Grundgedanken des Schadensausgleichs folgenden sollte in Fällen, in denen Verstöße auf fahrlässiges oder leicht fahrlässiges Verhalten des Arbeitnehmers zurückzuführen sind, auf arbeitsrechtliche Sanktionen gegenüber dem verursachenden Arbeitnehmer in Form von Abmahnungen oder Kündigungen verzichtet werden. Anders als bei der Haftungsverteilung

[41] Vgl. nur: *Steffen/Stöhr*, RdA 2017, 43, 48.

[42] *Koch* in: Schaub/Koch, Arbeitsrecht von A-Z, 24. Auflage, 2020, Stichwort Haftung des Arbeitnehmers; *Wagner*, Münchener Kommentar zum BGB, 7. Aufl., 2017, § 823, Rn. 128.

wäre die Begründung dafür nicht die Kontrolle der innerbetrieblichen Anstrengungen zur Schadensprävention durch den Arbeitgeber.[43] Vielmehr würde er von der Verlässlichkeit rechtzeitiger Meldungen von IT-Sicherheitsverstößen durch Mitarbeiter und die damit einhergehenden Schadensminimierungen profitieren. Außerdem würde der Datenschutz und die IT-Sicherheit insgesamt wesentlich gestärkt, da der teilweise Verzicht auf Sanktionen die „Meldemoral" verbessern würde. Die Meldung würde dann nämlich nicht mehr durch Bedenken vor arbeitsrechtlichen Konsequenzen behindert. Zudem erführe die pflichtgemäße Meldung eines Vorfalls gegenüber einer Nichtmeldung eine Privilegierung, da ein Unterlassen der Meldung von Verstößen fast immer strengere Sanktionen nach sich ziehen würde als eine ordnungsgemäße Erfüllung der Meldepflicht.

Es kommt hinzu, dass dem Arbeitgeber der Nachweis eines Fehlverhaltens des Arbeitnehmers gerade in Fällen nur schwer möglich sein wird, in denen ein Verstoß auf nur leicht fahrlässigem oder fahrlässigem Verhalten beruht. Die Sanktionierung des Arbeitnehmers ist in diesen Situationen somit häufig mit der rechtlichen Unsicherheit einer gerichtsfesten Beweisbarkeit behaftet und somit angreifbar. Das Interesse des Arbeitgebers an einer solchen unsicheren Sanktionierung ist daher auch gering. Zugleich bestünde durch den Sanktionsverzicht in Fällen von leichter und normaler Fahrlässigkeit nicht die Gefahr, dass eine arbeitsrechtliche Sanktionierung von insbesondere wiederholtem oder gleichgelagertem Fehlverhalten ausgeschlossen würde. Auch bei Zugrundelegen eines objektiven zivilrechtlichen Sorgfaltsmaßstabs[44] erfordert der Vorwurf grober Fahrlässigkeit regelmäßig eine subjektive, das normale Maß übersteigende Vorwerfbarkeit des Fehlers.[45] Dies kann gerade bei der Wiederholung einer schon begangenen Pflichtverletzung regelmäßig angenommen werden. Zuletzt spricht auch die Existenz des Verwertungsverbots in den §§ 42 Abs. 4, 43 Abs. 4 BDSG für Meldungen und Benachrichtigungen von Datenverarbeitern in Ordnungswidrigkeiten- und Strafverfahren für eine Bevorzugung von Arbeitnehmern, die Datenschutz- und IT-Sicherheitsverstöße in Betrieben an ihre Arbeitgeber melden. Durch die vorgeschlagene Privilegierung werden nämlich die Folgen der Unanwendbarkeit des Selbstbelastungsverbots auf arbeitsrechtliche Mitteilungspflichten abgemildert. Außerdem käme es zu einer Angleichung der Sanktionen für Arbeitnehmer und Unternehmen. Denn ebenso wie Unternehmen aufgrund des Verwertungsverbots im BDSG keine strafrechtlichen Konsequenzen drohen, wenn sie der ihnen auferlegten Meldepflicht nachkommen, müsste Arbeitnehmer bei Einhaltung der Meldepflicht in Fällen mit geringem Verschuldensvorwurf dann keine arbeitsrechtlichen Folgen befürchten.

Gegenüber den zuvor erwogenen Alternativen eines anonymisierten Meldesystems oder einer Bagatellgrenze bietet der teilweise Sanktionsverzicht einerseits den Vorteil, dass am Verschulden und damit der objektiven Vorwerfbarkeit eines Fehlverhaltens angeknüpft wird. Andererseits setzt er am Grundproblem der Arbeitnehmer an, nämlich deren Angst vor

[43] *Wagner*, Münchener Kommentar zum BGB, 7. Aufl., 2017, § 823, Rn. 128.

[44] Vgl. nur: BVerfG, Beschl. v. 07.05. 1996, 1 BvQ 4/96, NJW-RR 1996, 980; BGH, Urt.v.17.03.1981, VI ZR 191/79, NJW 1981, 1603, 1604; BGH, Urt. v. 13.02.2001, VI ZR 34/00, NJW 2001, 1786,1787; *Lorenz* in: Beck'scher Onlinekommentar BGB, 53. Ed., 2020, §276 BGB, Rn. 20; *Grundmann*, Münchener Kommentar zum BGB, 8. Aufl., 2019, § 276, Rn. 55f.; *Schulze* in: NK-BGB, 10. Aufl., 2019, §276, Rn. 13.

[45] BGH, Urt.v. 11.05.1953, IV ZR 170/52, NJW 1953, 1139; BGH, Urt. v. 17.06.1992, XII ZR 119/91, NJW 1992, 2418; *Grundmann*, Münchener Kommentar zum BGB, 8. Aufl., 2019, § 276, Rn. 55f.

arbeitsrechtlichen Konsequenzen, ohne sie aus ihrer grundsätzlichen Verantwortung für die Einhaltung der Sicherheitsregeln zu entlassen. Insoweit löst der Vorschlag das Problem auf der Basis einer Abwägung zwischen den Interessen und betrieblichen Anforderungen der Arbeitgeberseite, den problemauslösenden Bedenken der Arbeitnehmer und dem Ziel einer Verbesserung des Datenschutzes und der IT-Sicherheit.

4 Fazit

Die Meldepflicht von IT-Sicherheits- oder Datenschutzverstößen ist eine der zentralen organisatorischen Institutionen zum Schutz der IT-Infrastruktur von Unternehmen. Die auf ihr basierenden Maßnahmen, insbesondere in Form früher und effektiver Reaktionen auf Angriffe, können ihre Wirkung jedoch nur entfalten, wenn die Meldepflicht eine hohe Akzeptanz innerhalb eines Unternehmens genießt und in der Praxis auch tatsächlich umgesetzt wird. Drohende arbeitsrechtliche Sanktionen aufgrund eines zusammen mit einer Meldung offengelegten eigenen Fehlverhaltens können der bereitwilligen Wahrnehmung dieser Pflicht entgegenstehen. Mitarbeiter, die der Meldepflicht dennoch nachkommen, erfahren dabei keinen Schutz durch ein Verwertungsverbot oder die Anwendung der Selbstbelastungsfreiheit. Sie erwartet in Form einer Abmahnung oder Kündigung vielmehr theoretisch dieselbe arbeitsrechtliche Sanktion wie ihre Kollegen, die eine Meldung unterlassen.
Zur Stärkung der Durchsetzung einer Meldepflicht und damit einhergehend des Datenschutzes und der IT-Sicherheit in Unternehmen sollte erwogen werden, arbeitsrechtliche Sanktionen gegen Mitarbeiter auszusetzen, die mit einer Meldung gegebenenfalls eigenes leicht fahrlässiges oder fahrlässiges Verhalten offenbaren. Hierdurch könnten Arbeitnehmer die Meldepflicht ohne Sorge vor persönlichen Konsequenzen wahrnehmen, was sich positiv auf die Akzeptanz der Maßnahme auswirken würde und das Unternehmen vor schwerwiegenden Konsequenzen bewahrt, die aus unentdeckten Angriffen auf ihre IT-Infrastruktur entstehen können. Für zukünftige Arbeiten bleibt an dieser Stelle jedoch offen, wie den Arbeitnehmern die Unterscheidung zwischen (leicht) fahrlässigem und grob fahrlässigem Verhalten erklärt werden kann, damit die Akzeptanz und Ausübung der Meldepflicht tatsächlich steigen. Ebenso muss weiter untersucht werden, wie der Arbeitgeber die Arbeitnehmer ausreichend aufklären und ihnen dadurch ein ausreichendes Bewusstsein für entsprechenden Fehlverhaltens ermöglichen kann.

R. Reussner, A. Koziolek, R. Heinrich (Hrsg.): INFORMATIK 2020,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2021 841

Risiken für die Privatheit aufgrund von Maschinellem Lernen*

Verena Battis[1], Lukas Graner[1]

Abstract: Maschinelle Lernverfahren sind aus unserem Alltag fast nicht mehr wegzudenken - selbstlernende Verfahren finden bereits in nahezu allen Bereichen des Lebens Anwendung. In vielen Fällen werden dabei auch private und/oder sensible Informationen verarbeitet. Da selbstlernende Verfahren in der Regel auf sich nicht überschneidenden Datenmengen trainiert und später angewendet werden, ging man lange davon aus, dass es nicht möglich sei, vom finalen Modell Rückschlüsse auf die zum Training verwendeten Daten zu ziehen. Ergebnissen aus der jüngeren Forschung demonstrieren jedoch, dass es sich bei dieser Annahme um einen Trugschluss handelt. Die vorliegende Arbeit erläutert welche Risiken sich für die Privatheit des Einzelnen im Rahmen von maschinellen Lernverfahren ergeben und wie dem unerwünschten Abgreifen von sensiblen Informationen bereits in der Trainingsphase entgegen gesteuert werden kann.

Keywords: Privatheit; Privatsphäre; Risiken; maschinelles Lernen; Angriffe

1 Motivation

Maschinelles Lernen (ML) ist ein Teilgebiet der künstlichen Intelligenz und beschreibt eine Reihe von Lernalgorithmen, die versuchen Strukturen in Daten zu erkennen, um basierend auf diesen Mustern bspw. Klassifizierungs- oder Regressionsaufgaben zu lösen. Der Einsatz von Verfahren des maschinellen Lernens bietet sich immer dann an, wenn die zu lösenden Probleme zu komplex oder zu umfassend sind, um sie analytisch beschreiben zu können [Do18]. Gleichzeitig bedeuten größere Datenmengen auch, dass mehr Informationen zum Trainieren der Lernalgorithmen zur Verfügung stehen, was tendenziell zu besseren Modellen und effizienteren Schätzungen führt [ST17]. Ein maschinelles Lernverfahren welches in den letzten Jahren infolge seiner Flexibilität und guten Generalisierungsfähigkeit besonders an Beliebtheit gewonnen hat, sind sogenannte Neuronale Netze. Diese finden aufgrund ihrer Fähigkeit selbst komplexe, nicht-lineare Funktionen zuverlässig approximieren zu können, in den verschiedensten Bereichen Anwendung – ob im Verarbeiten und Analysieren natürlicher Sprachen, zur Bild- oder Gesichtserkennung oder zum Aufspüren von Anomalien. Diese Fähigkeit, welche häufig mit dem Begriff *Kapazität* beschrieben wird [BV19], ist es aber auch, welche die Privatheit des Individuums bedroht. Modelle mit nicht ausreichender

* Diese Forschungsarbeit wurde vom Bundesministerium für Bildung und Forschung (BMBF) und vom Hessischen Ministerium für Wissenschaft und Kunst (HMWK) im Rahmen ihrer gemeinsamen Förderung für das Nationale Forschungszentrum für angewandte Cybersicherheit ATHENE unterstützt.

[1] Fraunhofer-Institut für Sichere Informationstechnologie SIT, Rheinstraße 75, 64295 Darmstadt
vorname.nachname@sit.fraunhofer.de

doi:10.18420/inf2020_75

Kapazität können den in den Trainingsdaten enthaltenen Zusammenhang nur schwer oder gar nicht abbilden, wohingegen eine zu hohe Kapazität dazu führen kann, dass das Neuronale Netz Eigenschaften der Trainingsdaten auswendig lernt [Go16].
Üblicherweise sind die Datensätze auf denen ML-Algorithmen trainiert und später angewendet werden, disjunkt. In der Konsequenz sollte es nicht möglich sein, vom finalen Modell Rückschlüsse auf die zum Training verwendeten Daten zu ziehen, was einer Anonymisierung der verwendeten Trainingsdaten gleichkommen würde [WBH19]. Dass in großen Datenbeständen - selbst in solchen aus gering strukturierten oder gar unstrukturierten Daten - entscheidende Verknüpfungen gefunden werden können, welche das Herstellen von Personenbezügen ermöglichen, ist bereits hinreichend bekannt [HG16; NS08]. Moderne ML-Angriffe gehen allerdings noch einen Schritt weiter. Hier werden bestimmte Eigenschaften des ML-Algorithmus bewusst ausgenutzt, z.B. die zum Teil extreme Kapazität eines Neuronalen Netzes, um Informationen bezüglich der zum Training verwendeten Daten in Erfahrung zu bringen und somit die Privatheit der Datensubjekte zu kompromittieren.

Im Folgenden werden drei Arten von Rückschlüssen und die korrespondierenden Angriffe vorgestellt: Model Inversion, Membership Inference sowie Model Extraction. Der Fokus der vorliegenden Arbeit liegt auf den Risiken von maschinellen Lernverfahren, weshalb mögliche Gegenmaßnahmen abschließend nur in aller Kürze angesprochen werden.

2 Angriffe auf die Privatheit

2.1 Begriffserklärung

Angenommen ein maschinelles Lernmodell wird auf einem vertraulichen und unveröffentlichten Datensatz trainiert und anschließend zur Nutzung bereitstellt. Allgemein wird zwischen zwei Szenarien unterschieden: ob das Modell vollständig veröffentlicht wird oder ob dem Nutzer lediglich Nutzungszugriff auf das Modell gewährt wird, bspw. über eine API. Wird das Modell an sich veröffentlicht, kann der Nutzer das Modell nach Belieben befragen und besitzt darüber hinaus volles Wissen über den verwendeten Algorithmus, die Architektur und die Parameter des Modells. Man spricht in diesem Kontext von einem *White-Box Zugriff*. Im Gegensatz dazu kann der Nutzer in einem sogenannten *Black-Box Setting* das Modell zwar ebenfalls mit seinen eigenen Datenpunkten befragen, um eine Ausgabe zu erhalten, verfügt darüber hinaus aber über keinerlei Wissen bezüglich des verwendeten Modells, dessen Architektur oder verwendeter Parameter [BR18; KK+12; Sa19]. Je mehr Wissen der Angreifer über das Zielmodell hat, desto größer ist die Chance auf einen erfolgreichen Angriff. Dementsprechend gestaltet sich ein Black-Box Angriff als herausfordernder als ein White-Box Angriff.

Weiter gehen wir davon aus, dass es sich bei dem maschinellen Lernmodell um ein Klassifikationsmodell handelt, dessen Ausgabe aus Wahrscheinlichkeitswerten besteht. Zum einen geben diese Werte an, welcher Klasse bzw. welchem Attribut der eingegebene

Datenpunkt zugeordnet wird. Zum anderen bedeuten Werte nahe 1 auch, dass sich das Modell bezüglich seiner Entscheidung sicherer ist. Resultiert in einem Klassifizierungsproblem mit n Klassen bspw. eine Wahrscheinlichkeit von $1/n$ für eine bestimmte Klasse, so ist sich das Modell wesentlich unsicherer bezüglich seiner Entscheidung, als wenn es einen Datenpunkt mit einer Wahrscheinlichkeit von 0,98 einer der Klassen zuweist.

Die folgenden Beispiele stellen zwar ausschließlich Angriffe auf Neuronale Netze dar, die beschriebenen Ansätze sind jedoch ebenfalls auf eine Reihe anderer maschineller Lernalgorithmen anwendbar.

2.2 Model Inversion

Die Idee der Model Inversion ist es, das Modell selbst zu nutzen, um gezielt Datenpunkte, die zum Training verwendet wurden, zu rekonstruieren. Je nach Intention des Angreifers bedarf es nicht einmal zwangsläufig einer vollständigen Rekonstruktion der Daten, sondern nur bestimmter Eigenschaften. Eine vollständige und perfekte Rekonstruktion des Trainingsdatensatzes würde eine massive Verletzung der Privatheit der Datensubjekte darstellen. Betrachten wir die Software *Faception*, welche von dem gleichnamigen israelischen Konzern vermarktet wird [Fa19]. *Faception* ist ein maschinell lernendes Modell, welches anhand von Gesichtsbildern Rückschlüsse auf die Persönlichkeit der jeweiligen Person schließt. Die Entwickler werben damit, dass ihr Modell anhand eines einfachen Fotos entscheiden kann, ob es sich hierbei um einen Wissenschaftler, einen Bingo-Spieler, einen Pädophilen oder um einen Terroristen handelt [Fa19]. Gerade mit Blick auf die beiden letztgenannten Kategorien kann ein Angreifer ein besonders hohes Interesse daran haben, die zum Training des Modells verwendeten Gesichtsbilder möglichst akkurat zu rekonstruieren.

Da der Angreifer bei dieser Form des Angriffs im Extremfall von einem kleindimensionalen Ergebnisvektor (i.d.R. n Wahrscheinlichkeitswerte gemäß der n zuzuordnenden Klassen) auf einen hochdimensionalen Input zurückschließen muss, steigen die Erfolgschancen des Angriffs je mehr Informationen bezüglich des Modells dem Angreifer vorliegen. Im Optimalfall verfügt der Angreifer über einen *White-Box Zugriff* und kann die Gradienten im Neuronalen Netz direkt berechnen. Diese Gradienten, also alle partiellen Ableitungen einer multivariaten Funktion, symbolisieren den Effekt, den eine marginale Änderung in den Inputwerten (wie etwa einzelner Pixelfarbwerte eines Inputbildes) auf die Ausgabe des Modells hat und stellen die Grundlage vieler Model Inversion Ansätze dar.

Es sei bekannt, dass das Modell Personen anhand von Bildern n unterschiedlichen Klassen zuordnet. Ein Angreifer wäre demnach daran interessiert zu wissen, wie eine Person aussieht, die bspw. zum Training der Klasse 'Terrorist' verwendet wurde. Ein möglicher Ansatz, um einen Model Inversion Angriff durchzuführen ist, ausgehend von einem „leeren“ Startbild (bspw. ein vollständig schwarzes Bild) das Modell wiederholt zu befragen [Sa18]. Verfügt der Angreifer über einen White-Box Zugriff kann dieser nicht nur sämtliche Ausgabewerte beobachten, sondern auch die Gradienten berechnen. Mit Hilfe dieser Gradienteninformation

modifiziert der Angreifer schrittweise die Pixelwerte des Eingabebildes dahingehend, dass die Ausgabekonfidenz für die Zielklasse (hier: 'Terrorist') maximiert wird. Wird das Modell anschließend mit dem finalen, über Pixeloptimierung generierten Bild befragt, wird es dieses der Zielklasse mit einer sehr großen Sicherheit, also einem hohem Wahrscheinlichkeitswert, zuordnen. Hierbei wird naiverweise angenommen, dass das so generierte Bild eine Instanz der Trainingsdaten darstellt oder diese für die Zielklasse zumindest angemessen repräsentiert. Dies ist allerdings nur in einigen wenigen Sonderfällen korrekt, wie im Folgenden demonstriert.

In der akademischen Literatur stößt man häufig auf das Beispiel von Fredrikson et al., die die Trainingsdaten eines Gesichtserkennungsmodells rekonstruiert haben [FJR15]. Allerdings stellt gerade dieser Fall eines Gesichtserkennungsmodells einen Extremfall dar, da jede Ausgabeklasse des Modells eine individuelle Person repräsentiert. D.h. alle Trainingsdatenpunkte einer jeden Klasse sind nur unterschiedliche Aufnahmen der gleichen Person. Wenn der Angreifer also versucht, wie oben beschrieben, einen Input zu generieren, welcher vom Modell mit einer hohen Konfidenz der Zielklasse zugeordnet wird, so bildet er im Grunde einen Mittelwert über alle Trainingsbilder jener Klasse - und nicht wie gewollt, einen tatsächlichen Trainingsdatenpunkt. Im Fall von Fredrikson et al. wurde das anzugreifende Modell auf einem sehr kleinen, standardisierten Datensatz trainiert, welcher ausschließlich aus Frontalaufnahmen besteht. In Kombinationen mit einer naiven Modellarchitektur, die so in der Praxis keine Anwendung findet, resultieren künstlich generierte Bilder, die häufig auch von Menschen wiedererkannt werden können (vgl. Abb. 2, Spalte 2). Sobald die Daten innerhalb der Klassen jedoch eine größere Variation aufweisen oder eine komplexere Modellarchitektur genutzt wird (wie etwa ein Faltendes Netzwerk (engl. Convolutional Network), sind jene Rekonstruktionen, die basierend auf dem Ansatz von [FJR15] gewonnen wurden, nicht mehr als Objekte, geschweige denn spezifische Instanzen des Trainingssets wiederzuerkennen (vgl. Abb. 1, sowie Abb. 2 Spalte 3).

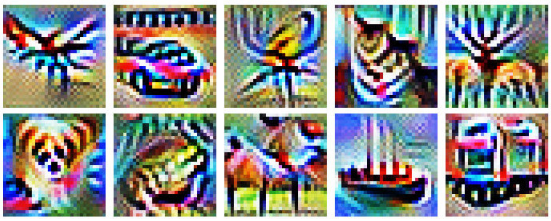


Abb. 1: Model Inversion Angriff auf den CIFAR 10 Datensatz, in Anlehnung an [FJR15]. Die rekonstruierten Klassen sind: Oben - Flugzeug, Auto, Vogel, Katze, Hirsch. Unten - Hund, Frosch, Pferd, Schiff, LKW.

Ergebnisse aus unserer eigenen Forschung (*GenMoIn*) zeigen jedoch, dass erweiterte Angriffsansätze trotz erschwerter Bedingungen - wie z.B. aufgrund hoher Modellkomplexität - dennoch Erfolge erzielen können. Beispielsweise kann der Model-Inversion-Prozess durch das Berücksichtigen von Hintergrundwissen optimiert werden. In den meisten Fällen stammen die für ein zu lösendes Problem verwendeten (Trainings-)Daten nur aus einer

bestimmten Domain. Im Falle einer Gesichtserkennung kann ein Angreifer bspw. davon ausgehen, dass das Trainingsset nur Portraitfotos beinhaltet. Bilder, die etwas anderes als menschliche Gesichter darstellen, wie etwa Objekte oder Rauschen, können von Grund auf ignoriert werden. Unser Ansatz, *GenMoIn*, macht sich genau dieses Vorwissen über die Verteilung der Daten zu Nutze. Dazu bedienen wir uns sogenannten *Generator Netzwerken*, die unter anderem als Teil eines *Generative Adversarial Networks* (GAN) [Go14] trainiert werden können. Ein solcher Generator erhält als Eingabe lediglich einen Vektor und erstellt daraus einen Datenpunkt - etwa ein Bild. Die Werte dieses *Code-Vektors* beschreiben und bestimmen dabei in einer komprimierten Form den resultierenden Datenpunkt. Für den Prozess der Model Inversion wird der Generator vor das anzugreifende Modell geschaltet, sodass die Ausgabe des Generators als Eingabe für das Zielmodell fungiert. Ab hier verläuft *GenMoIn* analog zu dem Angriff nach Fredrikson et al. [FJR15]. Allerdings wird in *GenMoIn* anstatt einzelner Pixelwerte der Code-Vektor selbst schrittweise optimiert. Jeder weitere Schritt verändert somit den Input-Code und damit auch den zu rekonstruierenden Datenpunkt, sodass die Konfidenz des Modells bezüglich der Zielklasse maximiert wird. Der Generator repräsentiert hierbei einen Vorfilter, welcher nur Datenpunkte aus einer bekannten Verteilung generieren kann.

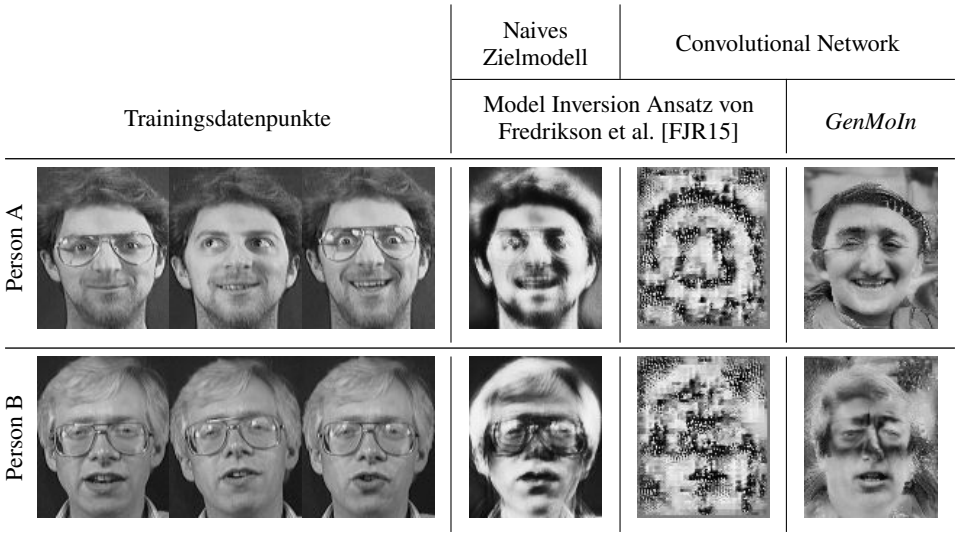


Abb. 2: Model Inversion Angriff auf den ATT Faces Datensatz. Jede Zeile, und somit Person, repräsentiert eine individuelle Klasse. Die linke Spalte stellt jeweils eine Teilmenge der entsprechenden Trainingsbilder dar. Die zweite Spalte zeigt Ergebnisse des Angriffs von Fredrikson et al. [FJR15] bei dem das Zielmodell nur aus einer Eingabeschicht, unmittelbar gefolgt von einer Ausgabeschicht, besteht. Die beiden letzten Spalten beziehen sich auf ein komplexeres Zielmodell mit zwei Convolutional Layern, wobei links Ergebnisse des Ansatzes nach Fredrikson et al. [FJR15] (Spalte 3) und rechts (Spalte 4) die eines unserer eigenen Angriffsansätze (*GenMoIn*) dargestellt sind.

Allerdings ist zu beachten, dass der Generator die Vielfalt der Datenverteilung möglichst umfassend und ohne Lücken gelernt hat - also ein ausreichendes Ausgabespektrum abdeckt. Ist dies nicht der Fall, etwa als Folge eines sogenannten *Mode Collapse*, kann dies zu stark verfälschten Ergebnissen der Model Inversion führen oder eine Rekonstruktion gänzlich unmöglich machen.

Im Fall eines Gesichtserkennungsmodells kann ein potentieller Angreifer davon ausgehen, dass der Trainingsdatensatz ebenfalls aus Aufnahmen von menschlichen Gesichtern bestanden hat. Gemäß dieses Vorwissens verwenden wir für unseren Angriff auf das ATT-Faces-Zielmodell ein Generator Netzwerk, welches Gesichter generiert. Die StyleGAN-Architektur [KLA19] gilt zusammen mit ihrer Weiterentwicklung StyleGAN2 [Ka19] als wegweisend für bildgenerierende Modelle. Deshalb, und weil ebendieses *Generative Adversarial Network* auf einer sehr großen, heterogenen Menge an Gesichtsbildern trainiert wurde [St19], haben wir den Generator des StyleGANs als Vorfilter für unseren Ansatz ausgewählt.

Mittels *GenMoIn* lassen sich auch aus komplexeren Modellen Rekonstruktionen erzielen. Wenngleich nicht vollständig wahrheitsgetreu, können diese für den Angreifer dennoch relevante Informationen enthalten – vor allem im Vergleich zu den stark verrauschten Bildern des Referenzangriffs (vgl. Abb. 2, Spalte 4 und 3). So können selbst in einem nur teilweise rekonstruierten Gesichtsbild trotz fehlender oder inkorrekter Gesichtszüge, andere, potentiell sensible Informationen, wie Hautfarbe oder das Tragen einer Brille, ermittelt werden. Dieser Zusammenhang ist in Abbildung 2 dargestellt. In den Rekonstruktionen nach [FJR15] (Spalte 3) lassen sich zwar Gesichtskonturen erkennen, diese sind jedoch sehr verrauscht und geben keine Auskunft über Details. Im *GenMoIn* -Ansatz (Spalte 4) wurde dagegen das Wissen, dass es sich um Portraitfotos handelt, direkt in den Rekonstruktionsprozess integriert, sodass automatisch ein erkennbares Gesicht generiert wird. Obwohl es sich nicht um exakte Replikationen von Trainingsinstanzen handelt, sind Details, wie die Haarfarbe, prägnante Gesichtszüge oder das Tragen von Accessoires deutlich erkennbar. So sind die charakteristische Nase und dunklen Haare von Person A), wie auch die volle Unterlippe und das Tragen einer Brille von Person B) deutlich in den Rekonstruktionen von *GenMoIn* zu erkennen.

2.3 Membership Inference

Das Ziel des Membership Inference Angriffs ist es, gegeben eines bestimmten Datenpunktes, eine Aussage darüber treffen zu können, ob ebenjener Datenpunkt zum Trainieren des betrachteten Modells verwendet wurde. Die zugrundeliegende Aufgabe des Zielmodells, also ob es sich um eine Klassifikation oder Regression handelt, ist hierbei für den Erfolg des Angriffs unwesentlich. Einem Angreifer geht es rein um das Verknüpfen eines Datenpunktes mit zusätzlich vorhandenen Informationen über den Trainingsdatensatz. Shokri et al. [Sh17] haben nachgewiesen, dass neuronale Netze aufgrund ihrer Kapazität besonders anfällig für Membership Inference Angriffe sind. Dabei trainierten die Autoren ein separates

Angriffsmodell, welches Anhand der Entropie der Ausgabewerte des Zielmodells - also aufgrund der darin enthaltenen Sicherheit bzw. Unsicherheit - beurteilen kann, ob ein Input zum Training des Zielmodells verwendet wurde oder nicht. Hintergrund ist, dass das Zielmodell Informationen, die es bereits während des Trainings „gesehen“ hat, mit einer höheren Konfidenz klassifiziert. Allgemein stellen Angriffe wie die Membership Inference eine Verletzung der Privatheit dar, sind aber besonders dann kritisch, wenn es sich um sensible Informationen handelt, wie bspw. die finanzielle Situation oder medizinische Angaben über eine Person [WBH19].

Ein trainiertes Netz reagiert merkbar anders auf Daten, welche zum Training verwendet wurden, als auf bisher ungesehene Testdaten. Das liegt daran, dass das Training eines Neuronalen Netzes kein einmaliger, sondern ein iterativer Vorgang ist, während dem das Modell den (endlichen) Trainingsdatensatz in unterschiedlichen Konstellationen immer wieder neu bewerten muss. In der Regel ist das Ziel des Trainings eine möglichst gute Anpassung zwischen den Modellentscheidungen und den tatsächlich beobachteten Realisationen zu erreichen. Dafür wird am Ende jeder Trainingsiteration eine Verlustfunktion (engl. loss) berechnet, die die Abweichung zwischen geschätzten und tatsächlichen Werten misst. Während des Trainings werden die Parameter des Modells so verändert, dass die resultierende Verlustfunktion minimiert wird. Genau hier liegt allerdings ein zentrales Problem des überwachten Lernens. Wird ein Modell zu lange auf einem endlichen Datensatz trainiert, beginnt es irgendwann sich Trainingsdatenpunkte zu merken, um die Verlustfunktion weiter zu minimieren. Gleiches gilt für Modelle, deren effektive Kapazität - in Relation zum zu lösenden Problem - zu groß ist[2]. Anstatt Zusammenhänge in den Daten zu erkennen, lernt das Modell die Trainingsdaten und die zugehörigen Ausgaben zu replizieren, was zu einer sinkenden Generalisierungsfähigkeit auf bisher ungesehene Daten führt. Man spricht hier von *Overfitting* - einer Überanpassung des Modells an die gegebenen Daten.
Genau dieses Overfitting ist es, was sich der Membership Inference Angriff zu Nutze macht. Sobald das Modell zu overfitten beginnt, beginnt es auch damit sich Trainingsinstanzen zu merken. Wird das Modell nach abgeschlossenem Training dann wiederrum mit einem Trainingsdatum konfrontiert, wird es die Trainingsinstanz - vereinfacht gesprochen - wiedererkennen und, verglichen mit einem bisher ungesehenen Datenpunkt, mit einer höheren Konfidenz der jeweiligen Klasse zuordnen. Graphisch dargestellt ist dieser Effekt in Abb. 3. Es ist deutlich zu erkennen, dass die Modellprognosen bezüglich der bisher ungesehenen Datenpunkte (orange) mit einer höheren Unsicherheit behaftet sind, als bereits bekannte (Trainings-)Instanzen (blau).

Der gesamte Membership Inference Angriff läuft wie folgt ab: Zunächst befragt der Angreifer das Zielmodell wiederholt, um einen vollständigen Datensatz mit Eingabe und zugehöriger Ausgabe zu generieren. Anschließend wird ein sogenanntes *Schattenmodell* (engl. Shadow Model), welches das Zielmodell in seiner Funktionalität bestmöglich approximieren soll, auf einer Teilmenge ebendieses Datensatzes trainiert. Für das eigentliche *Angriffsmodell*, welches

[2] Die Kapazität eines Neuronalen Netzes lässt sich unter anderem über die Anzahl und Art der verwendeten Schichten (engl. layers) sowie Knoten (engl. nodes) innerhalb einer Schicht steuern. Die genaue Kapazität eines Neuronalen Netzes lässt sich allerdings nur schwer genauer quantifizieren

[3] `https://www.comp.nus.edu.sg/~reza/files/datasets.html`

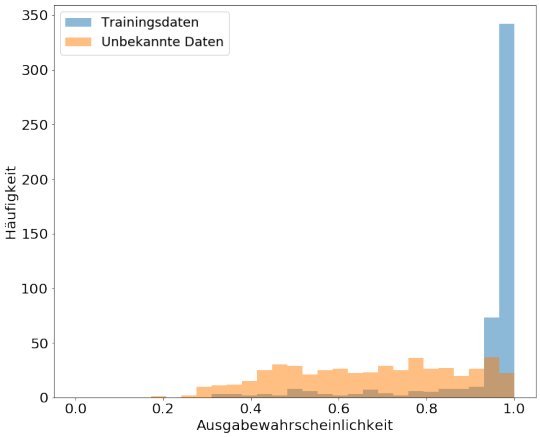


Abb. 3: Verteilungen der Ausgabewahrscheinlichkeiten nach Trainings- und unbekannten Referenzdaten (jeweils 500 Datenpunkte), Purchase10 Datensatz[3].

für die Erkennung der Trainingszugehörigkeit zuständig ist, muss nun zunächst noch das *Schattenmodell* mit den zuvor verwendeten Trainingsdaten und mit einem bisher ungesehenen Referenzdatensatz befragt werden. Die Ausgaben des Schattenmodells bezüglich dieser beiden Datensätze, zusammen mit einem binären Indikator, ob es sich bei dem jeweiligen Datenpunkt um eine Trainingsinstanz handelt oder nicht, dient dem Angriffsmodell als Trainingsdatensatz.

Um obiges Beispiel wieder aufzugreifen - nehmen wir an, das *Faception* Modell sei ausschließlich auf Bildern von Strafgefangenen trainiert worden[4]. Ein Angreifer, der im Besitz eines Portraitfotos ist, kann nun mittels Membership Inference herausfinden, ob die entsprechende Person evtl. bereits eine Haftstrafe verbüßen musste - vorausgesetzt natürlich, dass die fragliche Person tatsächlich Teil des Trainingsdatensatzes war. Dabei ist die Ausgabe des Angriffsmodells, ob Trainingsdatenpunkt oder nicht, unabhängig von der vom Zielmodell prognostizierten Klasse. Dem Angreifer geht es allein um das Verknüpfen der sensiblen Information mit der Person.

2.4 Model Extraction

Allgemein ist das Ziel eines Model Extraction Angriffs das Verhalten und somit die prädiktive Leistung eines Zielmodells auf einen bisher unbekannten Datensatz zu approximieren bzw. im günstigsten Fall zu kopieren. Alternativ kann es für den Angreifer auch Sinn machen die Architektur des Zielmodells zu *stehlen*. Aber warum ein Modell oder dessen Aufbau stehlen,

[4] Hierbei handelt es sich lediglich um eine hypothetische Annahme im Rahmen des Beispiels. Die Entwickler des Modells haben ihre Datengrundlage weder öffentlich kommuniziert noch zugänglich gemacht.

wenn es doch frei verfügbar bzw. nutzbar ist? Neuronale Netze können bspw. sehr komplexe Funktionen approximieren und auch Zusammenhänge in großen Datenmengen finden, die ansonsten vermutlich unbekannt geblieben wären. Die Entwicklung sowie das Training eines solchen Netzes sind häufig sehr zeit- und ressourcenintensiv. Zudem bedarf es einiges an tiefgehenden Verständnisses darüber, wie Neuronale Netze Informationen verarbeiten [Ja19; MDN19; OSF19; Pa19]. Nicht jeder, der solche Techniken anwenden möchte, verfügt über die nötige Rechenleistung, das Fachwissen oder die Menge an - potentiell sensiblen - Daten, die benötigt werden, um ein ML-Modell zuverlässig trainieren zu können. Wir identifizieren demnach drei verschiedene Motivationen, warum ein Angreifer ein Modell stehlen möchte:

- Machine Learning as a Service (MLaaS) Anbieter wie bspw. Google, Amazon oder Microsoft Azure ermöglichen via API-Zugang Zugriff auf hochperformante Modelle, die gegen eine geringe Nutzungsgebühr auf die eigenen Daten angewendet werden können. Je nachdem wie viele Datenpunkte der Angreifer vom Modell verarbeitet haben möchte, kann es sein, dass es für den Angreifer einen monetären Vorteil darstellt das Modell zu stehlen, sodass er es beliebig oft befragen kann, ohne weitere Gebühren für den Dienst zahlen zu müssen.
- Gleichsam kann ein gut performendes Modell auch einen Wettbewerbsvorteil darstellen, sodass ein Konkurrent ebenfalls Interesse an dessen Aufbau und Wirkungsweise hat.
- Die zuvor beschriebenen Angriffe – Model Inversion und Membership Inference – basieren beide darauf, dass der Angreifer entweder volle Einsicht in das Modell hat, und z.B. die Gradienteninformationen unmittelbar abgreifen kann, oder zumindest über eine hinreichend gute Approximation des Zielmodells verfügt. Dementsprechend kann Model Extraction als Vorstufe für die beiden Angriffe genutzt werden, um so deren jeweiligen Erfolgschancen zu erhöhen.

Ein weiterer Angriff auf Neuronale Netze, welcher hier nicht weiter betrachtet wurde, weil er keine direkte Bedrohung für die Privatheit darstellt, sind *Adversarial Examples*. Adversarial Examples sind kleine, für den Menschen in der Regel nicht wahrnehmbare Veränderungen in den Daten, die dazu führen, dass das Modell eine andere als die erwünschte Entscheidung trifft [GSS14]. Anwendungen des autonomen Fahrens verwenden fast ausschließlich Neuronale Netze, die basierend auf den gelieferten Eingaben - Video-, Sensor-, Radardaten – das zu lösende Problem in kleinere Klassifikationsaufgaben aufspalten. Nimmt eine Kamera bspw. ein Stopp-Schild auf, liefert das Neuronale Netz die Ausgabe 'anhalten'. Ein böswilliger Angreifer, dem es gelingt dieses Netz hinreichend mittels Model Extraction zu approximieren, ist nun in der Lage, basierend auf den Gradienteninformationen des gestohlenen Netzes, Adversarial Examples zu kreieren, die dazu führen, dass bspw. ein Stopp-Schild als ein Vorfahrtsschild missinterpretiert wird [Pa17]. Wird dieses Adversarial Example im öffentlichen Raum angebracht, wo es von anderen selbstfahrenden Vehikeln, die

das gleiche Modell wie das Zielmodell nutzen, erfasst werden kann, kann dies verheerende Konsequenzen haben.

Aktuell werden in der Literatur verschiedene Ansätze verfolgt, wie die Approximation eines unbekannten Modells mit möglichst wenig Vorinformationen am erfolgreichsten erfolgen kann. Nicht immer ist eine vollständige Extraktion des Zielmodells samt Architektur und verwendeten Hyperparametern das Ziel. Häufig beschränken sich die Angriffe auch nur darauf einzelne Komponenten zu ermitteln – bspw. ob eine Convolutional-Layer verwendet wurde oder welche Art der Regularisierung [WG18].

Am häufigsten sind sogenannte *Seitenkanalangriffe* (engl. side-channel attack) [Ba18; Du18; HZS18] sowie eine Form der *Knowledge-Distillation* zu finden. Bei letzterem wird versucht das Wissens eines Modells durch wiederholtes Befragen in ein separates, meist kleineres Modell zu überführen. Häufig werden dazu adaptive Verfahren (sog. *learning strategies*) verwendet. Diese identifizieren Trainingsdatenpunkte mit möglichst hohem Informationsgehalt, sodass die Anzahl der Anfragen, die an das Zielmodell gerichtet werden müssen, um ein aussagekräftiges Modell zu destillieren, möglichst klein gehalten werden kann [Co18; Ja19; MDN19; Pa19]. Letzteres ist aus mehreren Gründen relevant für den Angreifer. Zum einen werden durch weniger Anfragen geringere Kosten in einem Pay-per-Use System erzeugt, zum anderen ist die Gefahr, dass der Angriff als solcher identifiziert wird geringer, je weniger Anfragen an das Zielmodell gestellt werden und je weniger Zeit der Angriff an sich benötigt. In diesem Kontext ist auch die Arbeit von Oh et al. [OSF19] zu nennen, die einen Meta-Klassifikator trainiert haben, welcher anhand der Ausgabe des Zielmodells die verwendeten Hyperparameter, wie z.B. die Tiefe des Netzes, bestimmen sollte. In einer Erweiterung ihres Ansatzes gibt das Zielmodell durch Befragung mit einem speziell konstruierten Adversarial Example sogar selbst Informationen über die eigenen Modellspezifikationen preis.

3 Maßnahmen zum Schutz der Privatheit

Basierend auf dem Verständnis der in den vorhergehenden Abschnitten beschriebenen Risiken ist es möglich, Schutzmechanismen gegen diese zu entwerfen. Ziel ist es, aus trainierten Modellen keine oder nur eine tolerierbar geringe Menge an Informationen über die Trainingsdaten extrahieren zu können. Allgemein können solche Schutzmechanismen auf die Datengrundlage selbst, wie auch auf die Trainings- bzw. Entscheidungsphase angewendet werden. Es gilt allerdings zu beachten, dass i.d.R. alle Maßnahmen zum Schutz der Privatheit mit einer Reduktion der Prognosequalität des Modells einhergehen. Bei allen Gegenmaßnahmen gilt es folglich den klassischen Trade-Off zwischen Datennutzen und Datensicherheit genauestens abzuwägen. Die im folgenden beschriebenen Verfahren stellen keine vollständige Auflistung an Schutzmechanismen, sondern nur einige Beispiele dar, wie ein Modell bereits privatheiterhaltend trainiert werden kann.

3.1 Differential Privacy

Differential Privacy ist keine fixe Methode, sondern vielmehr eine Eigenschaft, welche verlangt, dass es für eine beliebige Analyse irrelevant ist, ob ein bestimmtes Datensubjekt im Datensatz enthalten ist oder nicht – in beiden Fällen sollten sich die Ausgabeverteilungen nicht signifikant voneinander unterscheiden. Intuitiv bedeutet dies, dass die Menge an Informationen, die maximal über ein bestimmtes Individuum herausgefunden werden kann, beschränkt wird. Differential Privacy wird in der Regel durch das Hinzufügen von Rauschen erreicht. Der Grad der Perturbation hängt von der Stärke des Einflusses des einzelnen Eintrags auf den Datensatz ab [Dw09]. Eine mögliche Ausprägung ist die ϵ-Differential Privacy. Der Parameter ϵ kontrolliert in diesem Fall wie groß der maximale Effekt eines Individuums auf das Ergebnis einer Analyse ist. Im Umkehrschluss quantifiziert ϵ aber auch in wie weit die Privatheit eines Individuums durch die Analyse kompromittiert werden kann. Kleinere ϵ–Werte gehen daher mit einem höheren Grad an Privatheit, aber auch mit einer stärkeren Perturbation einher, was sich wiederum negativ auf die Qualität der Daten auswirkt [Dw06]. Erschwerend kommt hinzu, dass keine feste Richtlinie existiert, wie der Parameter ϵ optimal gewählt werden soll. Die Wahl reicht von $\epsilon = 0.01$ bis $\epsilon = 1$ in akademischer Forschung bis hin zu Werten zwischen 1 und 10 in industriellen Anwendungsfällen (Google, Apple, US Census Bureau) [PCN18]. Wenngleich Differential Privacy im Vergleich zu vielen anderen privatheiterhaltenden Verfahren nachweislich Anonymität garantiert, so können bereits kleine ϵ- Werte in einem erheblichen Verlust an Klassifizierungsgenauigkeit (engl. accuracy) resultieren, was den Einsatz von Differential Privacy in der Praxis eher schwierig gestaltet [DKM19; DR14].

3.2 Adversarial Regularization

Anstatt wie im herkömmlichen Training eines Neuronalen Netzes nur eine einzelne Zielfunktion zu optimieren, wird bei der *Adversarial Regularization* der Trainingsprozess um einem feindlichen Angreifer erweitert, welcher seinen eigenen Gewinn zu maximieren versucht. Folglich werden bei der *Adversarial Regularization* zwei Modelle - das eigentliche Modell M_{Orig} und ein Angreifermodell M_{Att} - mit sich widersprechenden Zielfunktion trainiert [NSH18]. Dieses Vorgehen ist vergleichbar mit dem Training eines *Generative Adversarial Networks* (GAN). Das Klassifikationsmodell M_{Orig} berechnet die Wahrscheinlichkeit, dass ein Input zu einer beliebigen Klasse gehört und versucht dabei den Vorhersagefehler zu minimieren. Das Ziel des Angriffmodells M_{Att} besteht darin, die Genauigkeit des eigenen Klassifizierungsproblems - ob der betrachtete Datenpunkt bereits zum Training des Zielmodells genutzt wurde oder nicht - zu maximieren (vgl. Membership Inference Angriff, Abschnitt 2.3). Um die Privatheit zu schützen, wird der Gewinn des Angreifermodells als Regularisierungsterm in die zu minimierende Verlustfunktion des Klassifikationsmodells M_{Orig} integriert. Durch die Optimierung der beiden sich widersprechenden Zielfunktionen ist es dem Angreifer nicht möglich ein besseres Angriffsmodell zu entwickeln als jenes,

welches bereits vom Zielmodell antizipiert wurde. Dadurch hilft der Regularisierungsterm dem Trade-off zwischen Privatheit und Klassifizierungsgüte gerecht zu werden. Die Autoren [NSH18] wiesen nach, dass die Optimierung dieses Min-Max-Problems nachweislich vor Membership Inference Angriffen schützt, da der Trainingsalgorithmus im Optimalfall zu einem Gleichgewichtspunkt konvergiert, in welchem der beste Angriff auf die private Information, also ob der Datenpunkt zum Training verwendet wurde, nicht besser als eine zufällige Schätzung ist. Auf Trainingsdatenpunkten basierende Vorhersagen des Modells können folglich nicht von Vorhersagen unterschieden werden, die auf Basis bisher ungesehener Datenpunkte aus derselben Verteilung gemacht wurden. Zusätzlich zu der so erreichten Robustheit gegen Membership Inference Angriffe weist das trainierte Modell im Vergleich zu herkömmlich trainierten Modellen nur einen minimalen Verlust an Klassifikationsgenauigkeit auf.

3.3 Distillation

Distillation wurde ursprünglich von Hinton et al. [HVD15] vorgestellt. Die Idee ist es, wie bereits erwähnt (vgl. Abschnitt 2.4) und wie der Name andeutet, die Essenz des erlernten Wissens eines Modells in ein separates Modell zu überführen. Dieses Vorgehen kommt einer Komprimierung gleich und kann sogar auch - vor allem wenn ein Ensemble an Modellen gegeben ist - genutzt werden, um die Genauigkeit zu erhöhen. Zudem können mittels Distillation höhere Standards in Bezug auf Datenschutz erreicht und somit die Erfolgschancen von Model Inversion und Membership Inference verringert werden. Das bereits trainierte Modell wird hierbei als „Lehrer“ angesehen, von dem ein neues - das destillierte - Student-Modell, lernt. Dazu wird eine nicht zwangsweise annotierte weitere Datenmenge, meist disjunkt von der ursprünglichen Trainingsmenge, durch den Lehrer klassifiziert. Der Student trainiert nun auf ebendieser Datenmenge und erkennt dabei die Ausgabevektoren des Lehrers als zu erlernende Wahrheit (engl. Ground-Truth) an. Einfach ausgedrückt approximiert der Student den Lehrer, wobei er während des Trainingsprozesses nicht in Kontakt mit den ursprünglichen Trainingsdaten kommt. Distillation stellt somit ein datenschutzkonformes Instrument dar, da der Student keinen Zugriff auf potentiell sensible Informationen aus dem Trainingsset hat [Pa18].

4 Diskussion und Ausblick

Maschinelle Lernverfahren werden häufig genutzt, um z.B. unseren Alltag zu vereinfachen oder um Prozessabläufe zu optimieren. Dass sie dabei ein nicht zu vernachlässigendes Risiko für die Privatheit des Einzelnen bergen, ist in den seltensten Fällen bewusst. In dem vorliegenden Artikel wurden drei aktuelle Angriffe auf die Privatheit bzw. geistiges Eigentum vorgestellt sowie einige Gegenmaßnahmen skizziert. Es wurde aufgezeigt, dass gerade jene zentralen Eigenschaften, welche maschinelle Lernverfahren so wertvoll und nützlich machen, auch genau jene sind, die von Angreifern ausgenutzt werden können, um

an sensible Informationen zu gelangen.
Wenngleich die hier vorgestellten Angriffe durchaus eine reelle Gefahr darstellen, so besteht dennoch Optimierungspotential. Vielversprechende Richtungen für zukünftige Forschungen sind demnach privatheiterhaltender Maßnahmen sowie deren Umsetzung in der Praxis. Viele der heute angewendeten Maßnahmen skalieren schlecht oder sind nur für die Anwendung auf einen bestimmten Lernalgorithmus optimiert und auf andere ML-Verfahren nur schwer bis gar nicht anwendbar. Darüber hinaus entstehen durch das Schützen sensibler Informationen immer Kosten - ob in Form von längeren Trainingszeiten oder, dass die Daten signifikant an Nutzen einbüßen. Diese Kosten können unter Umständen so hoch ausfallen, dass in der Praxis aus ökonomischen Gründen von der Verwendung privatheiterhaltender Maßnahmen abgesehen wird [Do18; WBH19].

Literatur

[Ba18] Batina, L. et al.: CSI neural network: Using side-channels to recover your artificial neural network information, 2018, arXiv: 1810.09076.

[BR18] Biggio, B.; Roli, F.: Wild patterns: Ten years after the rise of adversarial machine learning. Pattern Recognition 84/, S. 317–331, 2018.

[BV19] Baldi, P.; Vershynin, R.: The capacity of feedforward neural networks. Neural networks 116/, S. 288–311, 2019.

[Co18] Correia-Silva, J. R. et al.: Copycat CNN: Stealing knowledge by persuading confession with random non-labeled data. In: 2018 International Joint Conference on Neural Networks (IJCNN). IEEE, S. 1–8, 2018.

[DKM19] Dwork, C.; Kohli, N.; Mulligan, D.: Differential Privacy in Practice: Expose your Epsilons! Journal of Privacy and Confidentiality 9/2, 2019.

[Do18] Doebel, I. et al.: Maschinelles Lernen. Eine Analyse zu Kompetenzen, Forschung und Anwendung, Techn. Ber., Fraunhofer-Gesellschaft, Muenchen, 2018.

[DR14] Dwork, C.; Roth, A.: The algorithmic foundations of differential privacy. Foundations and Trends® in Theoretical Computer Science 9/3–4, S. 211–407, 2014.

[Du18] Duddu, V. et al.: Stealing neural networks via timing side channels, 2018, arXiv: 1812.11720.

[Dw06] Dwork, C. et al.: Calibrating noise to sensitivity in private data analysis. In: Theory of cryptography conference. Springer, S. 265–284, 2006.

[Dw09] Dwork, C.: The differential privacy frontier. In: Theory of Cryptography Conference. Springer, S. 496–502, 2009.

[Fa19] Faception: Facial Personality Analysis, https://www.faception.com, Accessed: 2020-04-08, 2019.

[FJR15] Fredrikson, M.; Jha, S.; Ristenpart, T.: Model inversion attacks that exploit confidence information and basic countermeasures. In: Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security. S. 1322–1333, 2015.

[Go14] Goodfellow, I. et al.: Generative adversarial nets. In: Advances in neural information processing systems. S. 2672–2680, 2014.

[Go16] Goodfellow, I. et al.: Deep Learning. MIT press Cambridge, 2016.

[GSS14] Goodfellow, I. J.; Shlens, J.; Szegedy, C.: Explaining and Harnessing Adversarial Examples, 2014, arXiv: 1802.08908.

[HG16] Harmanci, A.; Gerstein, M.: Quantification of private information leakage from phenotype-genotype data: linking attacks. Nature methods 13/3, S. 251, 2016.

[HVD15] Hinton, G.; Vinyals, O.; Dean, J.: Distilling the knowledge in a neural network, 2015, arXiv: 1503.02531.

[HZS18] Hua, W.; Zhang, Z.; Suh, G. E.: Reverse engineering convolutional neural networks through side-channel information leaks. In: 2018 55th ACM/ESDA/IEEE Design Automation Conference (DAC). IEEE, S. 1–6, 2018.

[Ja19] Jagielski, M. et al.: High-fidelity extraction of neural network models, 2019, arXiv: 1909.01838.

[Ka19] Karras, T. et al.: Analyzing and improving the image quality of stylegan, 2019, arXiv: 1912.04958.

[KK+12] Khan, M. E.; Khan, F. et al.: A comparative study of white box, black box and grey box testing techniques. Int. J. Adv. Comput. Sci. Appl 3/6, 2012.

[KLA19] Karras, T.; Laine, S.; Aila, T.: A style-based generator architecture for generative adversarial networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. S. 4401–4410, 2019.

[MDN19] Mosafi, I.; David, E. O.; Netanyahu, N. S.: Stealing knowledge from protected deep neural networks using composite unlabeled data. In: 2019 International Joint Conference on Neural Networks (IJCNN). IEEE, S. 1–8, 2019.

[NS08] Narayanan, A.; Shmatikov, V.: Robust De-anonymization of Large Sparse Datasets, How to break anonymity of the Netflix Prize dataset. In: IEEE S&P 2008. IEEE Computer Society Conference Publishing Services (CPS), 2008.

[NSH18] Nasr, M.; Shokri, R.; Houmansadr, A.: Machine learning with membership privacy using adversarial regularization. In: Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security. S. 634–646, 2018.

[OSF19] Oh, S. J.; Schiele, B.; Fritz, M.: Towards reverse-engineering black-box neural networks. In: Explainable AI: Interpreting, Explaining and Visualizing Deep Learning. Springer, S. 121–144, 2019.

[Pa17] Papernot, N. et al.: Practical black-box attacks against machine learning. In: Proceedings of the 2017 ACM on Asia conference on computer and communications security. S. 506–519, 2017.

[Pa18] Papernot, N. et al.: Scalable private learning with pate, 2018, arXiv: 1802.08908.

[Pa19] Pal, S. et al.: A framework for the extraction of deep neural networks by leveraging public data, 2019, arXiv: 1905.09165.

[PCN18] Page, H.; Cabot, C.; Nissim, K.: Differential privacy an introduction for statistical agencies, Techn. Ber., NSQR. Government Statistical Service, 2018.

[Sa18] Salem, A. et al.: ML-Leaks: Model and Data Independent Membership Inference Attacks and Defenses on Machine Learning Models, 2018, arXiv: 1806.01246.

[Sa19] Sablayrolles, A. et al.: White-box vs black-box: Bayes optimal strategies for membership inference, 2019, arXiv: 1908.11229.

[Sh17] Shokri, R. et al.: Membership inference attacks against machine learning models. In: 2017 IEEE Symposium on Security and Privacy (SP). IEEE, S. 3–18, 2017.

[ST17] Shwartz-Ziv, R.; Tishby, N.: Opening the black box of deep neural networks via information, 2017, arXiv: 1703.00810.

[St19] StyleGAN: Official TensorFlow Implementation, https://github.com/NVlabs/stylegan, Accessed: 2020-04-08, 2019.

[WBH19] Winter, C.; Battis, V.; Halvani, O.: Herausforderungen für die Anonymisierung von Daten. In (David, K. et al., Hrsg.): INFORMATIK 2019. Gesellschaft für Informatik e.V., Bonn, S. 339–352, 2019.

[WG18] Wang, B.; Gong, N. Z.: Stealing hyperparameters in machine learning. In: 2018 IEEE Symposium on Security and Privacy (SP). IEEE, S. 36–52, 2018.

R. Reussner, A. Koziolek, R. Heinrich (Hrsg.): INFORMATIK 2020,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2021 857

Putting Privacy into Perspective – Comparing Technical, Legal, and Users' View of Information Sensitivity

Eva-Maria Schomakers[1] Chantal Lidynia[1] Dirk Müllmann[2] Roman Matzutt[3] Klaus Wehrle[3] Indra Spiecker gen. Döhmann[2] Martina Ziefle[1]

Abstract: Social media, cloud computing, and the Internet of Things connect people around the globe, offering manifold benefits. However, the technological advances and increased user participation generate novel challenges for users' privacy. From the users' perspective, the consequences of data disclosure depend on the perceived sensitivity of that data. But in light of the new technological opportunities to process and combine data, it is questionable whether users can adequately evaluate risks of data disclosures. As mediating authority, data protection laws such as the European General Data Protection Regulation try to protect user data, granting enhanced protection to "special categories" of data. In this paper, we assess the legal, technological, and users' perspectives on information sensitivity and their interplay. Technologically, all data can be referred to as "potentially sensitive." The legal and users' perspective on information sensitivity deviate from this standpoint, as some data types are granted special protection by law but are not perceived as very sensitive by users and vice versa. Our key findings still suggest the GDPR adequately protecting users' privacy but for small adjustments.

Keywords: Information Sensitivity; Privacy; European Data Protection Law

1 Introduction

Technological advances increased user participation online and generated large amounts of user data, which concerns users, who nevertheless disclose a lot of personal information [GGV18]. Users' decisions to disclose data are highly influenced by its perceived sensitivity, i.e., how risky they individually perceive particular information to be [Mo12]. At the same time, data collection and processing has evolved over time so that increasingly more information can be combined, deanonymized, and used to profile individuals – consequently, users may be unaware of novel threats stemming from recent technological advances. As a mediating authority, laws like the European General Data Protection Regulation (GDPR) govern the use of personal data by companies, thereby distinguishing categories of information sensitivity and granting different levels of protection correspondingly.

[1] Chair of Communication Science, Human-Computer Interaction Center, RWTH Aachen University, Campus-Boulevard 57, 52074 Aachen, Germany, {schomakers,lidynia,ziefle}@comm.rwth-aachen.de
[2] Goethe-University Frankfurt/Main, Chair of Public Law, Information Law, Environmental Law and Legal Theory, Theodor-Adorno-Platz 4 60323 Frankfurt/Main, Germany, {muellmann,spiecker}@jur.uni-frankfurt.de
[3] Chair of Communication and Distributed Systems, RWTH Aachen University, Ahornstraße 55, 52074 Aachen, Germany, {matzutt,wehrle}@comm.rwth-aachen.de

doi:10.18420/inf2020_76

However, with the ever-improving potential for data analysis, the question arises whether the regulation (legal perspective) captures what data can potentially become sensitive (technological perspective) and also what data users perceive to be sensitive (users' perspective). In this multidisciplinary paper, we will examine sensitivity of information in a multidisciplinary approach by taking all these three points of view. This assessment forms the basis for comparison between the three perspectives and the discussion of the findings with regard to the interests of online users and implications for future politics.

2 Information Sensitivity from a Technological Perspective

The early 2000's shift of online services towards the **Web 2.0** paradigm constituted a revolution of online services: user participation became an elementary ingredient of modern online services [OR07]. The level of user interaction culminated in the rise of global **social networks** such as Twitter and Facebook, which was enabled by shifting to the **cloud computing** paradigm [Ar10]. In addition to its increased scalability and lowered entry bar for service providers, the cloud's ubiquity also enabled users to outsource their data to simplify sharing or maintaining online backups. Ultimately, cloud computing motivated the advent of **smartphones** and the **Internet of Things (IoT)**. Cloud storage enabled synchronizing numerous devices easily and the cloud's scalable processing power allows service providers to remotely process the data sensed by their users' IoT devices [He13]. As a consequence, systems based on **distributed ledgers**, most notably blockchains, recently gained traction to break up this level of centralization: While initial blockchain systems such as Bitcoin [Na08] or Ethereum [Wo16] focused on achieving decentralized financial services in partially distrusted environments, distributed ledgers are now being explored for, e.g., tamper-proof file storage [Ko17] or managing access control to user data [ZNP15]. Distributed ledgers are experiencing this popularity because their immutability establishes technical accountability among otherwise mutually distrusting parties.

In conclusion, new technologies always simplified the deployment of online services centered around user participation or even enabled novel services over the Internet. However, those opportunities do not come without additional (privacy) challenges, as we detail in the following.

Web 2.0. The shift toward a strong focus on user participation within the Web 2.0 paradigm inherently led to the collection of more user data – and to new insights gained from service personalization, user tracking, and data breaches. Online services are routinely personalized to increase the user experience, e.g., provide better-fitting search results. While **personalized services** can benefit the user, the collected data is potentially highly sensitive. Not only is it possible to deanonymize users solely based on search queries [BZ06], it is further possible to disclose sensitive user data even from properly anonymized data sets [NS08]. **Web tracking** maximizes this form of data collection by monitoring users' browsing behavior across services [MM12], which potentially discloses a much more fine-grained view on users and was oftentimes opaque to the user prior to the GDPR's enactment. Even privacy-aware

users, who actively protect their privacy by deleting cookies or using private browsing, have been shown to be susceptible to web tracking due to their distinct behavior [Ye12]. A major threat also lies in the potential of **data breaches**, which disclose login credentials and other meta data for users' accounts on a regular basis.

Social Media. Especially the rise of social media revolutionized users' online behavior as users can now rapidly share personal moments and thoughts with both their friends and a general audience – leading to unprecedented privacy issues due to sensitive data disclosure. Users can directly share clearly sensitive data with the public (e.g., credit card information) or release it via meta data such as GPS locations stored in uploaded images [Sm12]. These incidents showcase the need for education regarding potential threats of sharing sensitive data on the Internet. Furthermore, the data users share online can be combined and subsequently collectively be exploited as shown by the Cambridge Analytica scandal [RCC18]. Hence, users can be profiled based on their shared data and the increased potential stemming from new analysis methods to exploit such data can cause data to effectively become sensitive.

Cloud Computing. Due to the cloud's multitenancy, single cloud providers could gain access to data of all customers [He14]. Also, a cloud can span multiple data centers. In this case, users lose control over where their data is being stored, which can violate both individual or even legal requirements [HHW13]. Hence, the increased complexity of data management complicates users' risk evaluations.

Smartphones and IoT. The ubiquity of sensing devices also creates new challenges for user privacy [ZMW14]. Third parties can potentially extract very fine-grained information from a user's sensor data via appropriate analysis technologies (e.g., location trajectories [Zi17]). Furthermore, the often-insufficient security of IoT devices for smart homes can potentially leak sensitive information directly from the user's house to the Internet [Se18]. This potential threat is further exemplified by recent advances in **deep learning** [LBH]. Thus, users must be further aware of the potential privacy implications of third parties analyzing their sensed data.

Distributed Ledgers. While privacy-preserving platforms based on distributed ledgers aim to mitigate public data disclosure [ZNP15], sensitive data on such ledgers can have especially devastating consequences due to their oftentimes public nature and immutability by design. For one, the initial promise of blockchains to provide financial privacy has been falsified [Me13]. Secondly, arbitrary data can be stored directly on blockchains, i.e., there is potential for the malicious disclosure of sensitive data of another user [Ma18]. These initial observations indicate that users once again will be facing increasing complexity in the technologies providing their online services in the future.

In conclusion, emerging new technologies can cause intuitively non-sensitive data to become sensitive due to the potentially unwanted impact seizing this data can have. Despite whole research areas being dedicated to protecting sensitive data from being disclosed to

unauthorized parties, there is no general solution to technical data protection. We expect this effect to be further exacerbated in the future as mainstream technology becomes more complex and diverse, and thus harder for users to keep track of regarding potential privacy threats.

3 Information Sensitivity from a Legal Perspective

Since May 2018, the data protection law in the member states of the European Union is almost exclusively determined by the European General Data Protection Regulation (GDPR), leaving only a very limited scope of application to distinct national data protection legislation [WB17]. Depending on the content of the data, the European data protection law distinguishes three different categories: personal, special categories of personal, and non-personal data, all of which are granted different levels of protection.

Art. 4 No. 1 GDPR defines personal data as any information relating to an identified or identifiable natural person. Special categories of personal data consist of personal data referring to particularly sensitive information concerning a natural person. Under Art. 9 sec. 1 GDPR these include data revealing a person's racial or ethnic origin, political opinions, religious or philosophical beliefs, or union membership. Furthermore, genetic, biometric, and health data, as well as data concerning a person's sex life or sexual orientation, are part of this category. Every piece of information not falling under the definition of personal data, however, has to be considered non-personal data under data protection law (cf. Art. 3 No. 1 Proposal for a Regulation of the European Parliament and of the Council on a framework for the free flow of non-personal data in the European Union, COM (2017) 495 final).

The group of personal data summarized under the notion of "special categories" consists of types of personal data which can be described as sensitive personal attributes. They have in common that they concern very personal beliefs or states which bear a special risk of being a leverage point for discrimination [AV18; KB18] and are closely connected to the exercise of fundamental rights (Recital 51, cl. 1 GDPR, [AV18; Fr18; We17]). The processing of this data can result in a severe violation of a person's privacy as well as significant risks to the fundamental rights and freedoms (Recital 51 cl. 1 GDPR, [Fr18]). The legal protection for special categories of personal data, therefore, has to be even stronger compared to common personal data. Following the principle of "ban with reservation to permit" [AV18; Ve18], Art. 9 GDPR restricts the processing of special categories of personal data to less, more specific, and more essential situations compared to mere personal data. Furthermore, only under very strict conditions may personal data of the special categories even be used for decision making based solely on automated processing, including profiling (Art. 22 GDPR). If special categories of personal data are processed, this always leads to the necessity for the processor to keep records of his processing activities (Art. 30 GDPR). If handling this kind of data in a larger scale, the processor has to conduct data protection impact assessment (Art. 35 sec. 3 lit. b) GDPR) and is obliged to appoint a data protection officer (Art. 37 sec. 1 lit. c) GDPR). The rigidity of the separation of these different levels of protection was already

the subject of discussion when it was introduced by the Data Protection Directive [AV18; Si97].

As non-personal data does not fall under the scope of the fundamental rights which establish data protection, it is, hence, neither protected under the GDPR nor national data protection laws. Non-personal data, however, may be protected by other laws under different legal means, e.g., business secrets which are protected by the national civil law [Mü18]. The GDPR as well as the national data protection legislations differentiate between data protection and data security. While data protection aims at the defense of personal data against the dangers of their processing, data security embraces all measures to preserve data from misuse and interference of risks from outside of the process of processing [He03; WB17]. A legal use of personal data always requires adequate safety and security measures. An appropriate security level considers the technical state-of-the-art, the costs of the implementation of the security measures, the probability of occurrence of security risks, nature, scope, context, and purposes of the processing as well as the risks for the rights and freedoms of the natural persons which might especially arise from the accidental or unlawful destruction, loss, or unauthorized disclosure (cf. Art. 32 sec. 1, 2 GDPR). The processing of particularly sensitive data may, hence, only lead to more data security. However, the processing of common personal data does not mean the absence of security measures.

4 Information Sensitivity from the User Perspective

From the user perspective, the perception of how sensitive a type of information is, is influenced by how concerned users are about the data provision and influences how willingly this information is provided [Mo12]. But what causes users to perceive certain information types as more sensitive than others? The perception of sensitivity is related to the perceived risks when disclosing information and, thus, related to the vulnerability and potential losses that are anticipated [Mo12]. Users are concerned about unauthorized use, misuse (e.g., fraud, identity theft, hackers), and improper access [Eu18]. But they also feel that the collection of information itself, targeted advertising, and profiling are violations of their privacy [SLZ18]. Thus, they seem not to differentiate between data privacy and data security. Moreover, more personally identifying information is perceived as more sensitive [MPS13], which goes in line with the GDPR covering personally identifying information. Improving data analysis technologies enable ever-deeper insights about users. Most users may not be aware of what is legally and technically possible and how, presumably non-identifiable or insensitive, data can be linked and used.

Besides limited knowledge about IT and law, there is another aspect that complicates the sensitivity evaluation for the users: Privacy perceptions depend on context and audience [Ni10]. We learn from early childhood on how to manage our privacy in an offline world. But online, data is persistently available over space and time, confusing the context in which we disclose information and those in which it can be accessed and by whom [Ta14]. Thus, users do not only have to include the present audience and context to evaluate the risk of disclosure and

sensitivity of information but also potential access of information in the future by different entities and in different contexts. And the technological possibilities to combine data across services also need to be considered.

The Empirical Approach. To provide a user perspective on sensitivity of information, we conducted an empirical online study in which N = 601 German internet users evaluated 40 data types regarding their felt sensitivity. In the questionnaire, the perceived sensitivity is assessed without contextual frame, as such context-free perceptions of sensitivity mirror assessments users need to make in the digital world. The participants evaluated the 40 data types (cf. Fig. 1) on a 6-point scale from "not sensitive at all" (1) to "very sensitive" (6) in randomized order to prevent sequence effects.

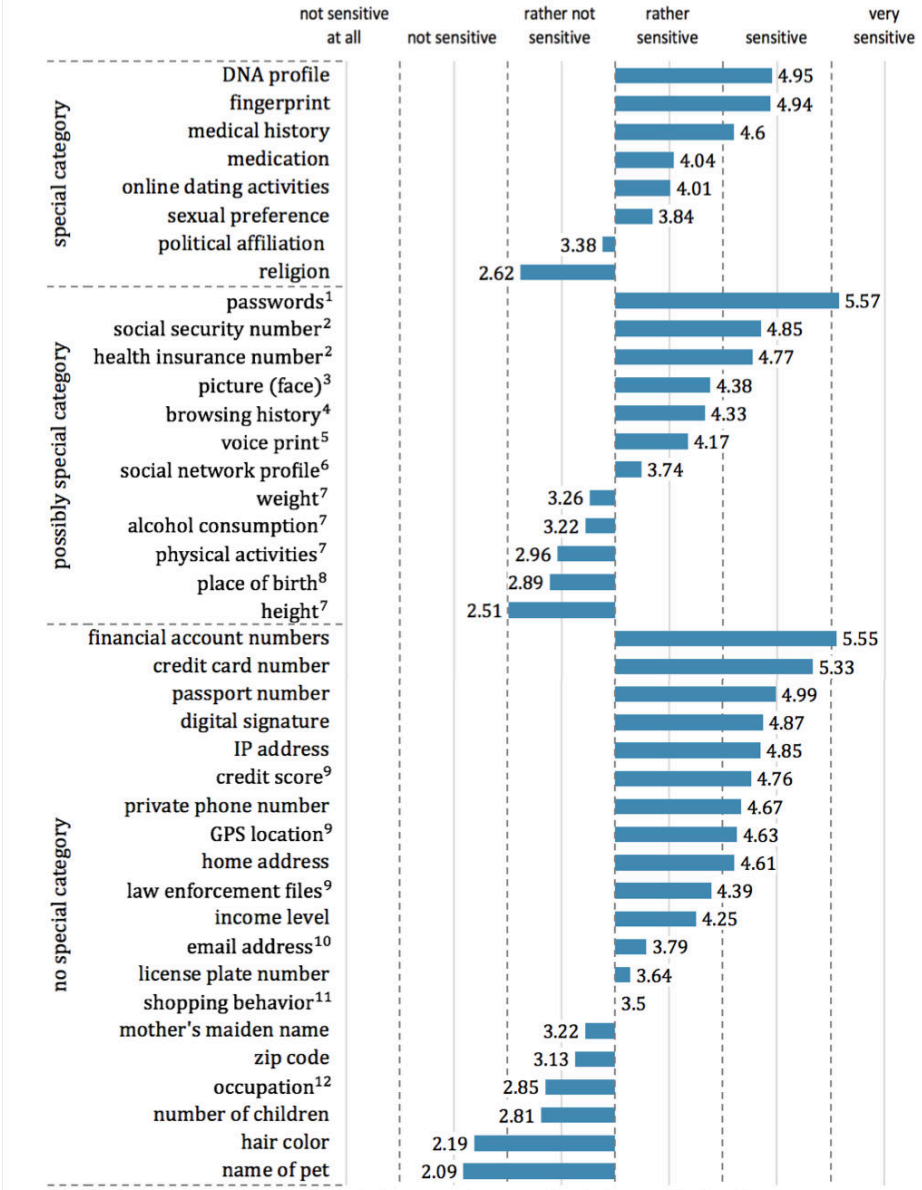
The sample includes 601 participants aged between 15 and 69 years (M = 38.8, SD = 20.2). 59.1% were women. The questionnaire was distributed online via an independent market research company. The education level is quite heterogeneously distributed, showing a good cross-section of German internet users.

Results. The perceived sensitivity for all 40 data types is depicted in Fig. 1. Passwords are perceived as most delicate followed by financial account numbers, with both being rated "very sensitive" (M > 5.5). Personal identifiers like passport number and fingerprint, location and medical history are perceived as "sensitive." Browsing history, medication, and sexual preferences are evaluated as "rather sensitive." "Rather not sensitive" are, e.g., political affiliation, weight, and zip code. The only two information types from this list that are perceived as "not sensitive" are hair color and name of pet. Nothing was, on average, felt to be "not sensitive at all" (M < 1.5).

5 Comparison between Legal, Technical, and User Perspective

Fig. 1 depicts the legal and user perspective on sensitivity. It shows that users' perception of sensitivity is for some data types in line with the legal categorization but also deviates strongly for others. The data types that are perceived as most sensitive by the users (passwords, financial account numbers) are legally classified as possibly special category and no special category. This assessment by users indicates they might not differentiate between data privacy and data security. Rather, the sensitivity evaluation is based on a risk assessment, and users are concerned about unauthorized access and illicit data misuse as well as about data collection, targeted advertising, and profiling. The legal use of personal data, however, under Art. 32 GDPR always requires adequate data security measures proportional to the risks of data processing. Hence, service operators implement established technical protection measures. However, even despite huge efforts to technically protect user data, we experience frequent data breaches [Hu13].

Data protection starts earlier, though, and already tries to minimize occasions and purposes in which personal data is collected and processed. While users' estimation of information



[1] Possible, if fingerprints or face recognition techniques are used as password equivalents, e.g. Apple products.
[2] There is no comprehensive social security number in Germany. Depending on the precise circumstances, an insurance number can be considered a special category of personal data.
[3] Possible, if used to identify biometric data.
[4] Possible, if browsing history allows conclusions about political opinions, religious or philosophical beliefs, union memberships, health data or data about the sexual orientation.
[5] Possible in health related contexts.
[6] Highly depends on the content of the profile.
[7] Possible, if linked to health contexts.
[8] Possible, if linked to racial or ethnical contexts.
[9] Subject to special regulation prohibiting its misuse but no special category of personal data.
[10] Except for the rare occasions in which a person uses an email address referring to religious institutions or unions.
[11] Except for the rare occasions in which they allow conclusions about health or sex life.
[12] Except for the rare occasions in which a person works for a religious or philosophical institution or a union.

Fig. 1: Users' evaluation of the sensitivity of 40 data types (n = 601) categorized into the legal classification.

sensitivity might anticipate the uncontrolled release and accept the necessity of the processing in other contexts, the data protection law works with wider categories and has internalized context dependency. Hence, from a legal point of view, the additional protection of special categories of personal data aims at categories whose general acquisition, irrespective of its legality, might bear severe risks and consequences. Thus, the legislator even restricted the contexts of legal uses compared to regular personal data.

Political affiliation and religion are classified as special category and particularly deserving of protection by the GDPR but are assessed as 'rather not sensitive' on average by the participants of the survey. The German view on data protection is, among other factors, highly influenced by the country's historical experience of two dictatorships cementing their power through surveillance and control and the potential as well as the risks of modern electronic data processing [Bu11; Ma12]. This affected the development of the European data protection law [Re12]. The legal point of view on special categories of personal data, as the most sensitive pieces of information in data protection law, mainly concern issues that can be used as **leverage points for discrimination**, such as religious believes, political opinions, or sexual orientation, and are closely connected to the exercise of fundamental rights, e.g., union memberships. Therefore, they need the particular protection of the democratic society and its laws. This aspect does not have the same importance for common users and the deviation between law and user evaluation can be explained by the methodological approach to report the mean user evaluation: For many users who have mainstream political views or a religion that is not discriminated, these information types may not seem sensitive. The minority of users who may be discriminated on these grounds do not have much weight within the average evaluation but still need protection from discrimination. Short-term financial losses and other acute consequences of released data are more relevant to the user, while the legislator has to consider long-term implications for the individual and the democratic society as a whole.

We have seen that the voluntary sharing of personal information in social media can make users vulnerable and create possibilities for harm. Users see these risks to some extend and state that they are concerned but still disclose this information [GGV18]. One explanation for this privacy paradoxical user behavior is given by the theory of the **privacy calculus** which assumes that users weigh perceived benefits and perceived privacy risks against each other [DH06]. Thus, they disclose information when the benefits outweigh the risks. Correspondingly, the evaluation of risks is only one side of the coin. Self-disclosure on social network sites bring many benefits to the individual including self-representation, relationship development, and social control [LPK13]. These may outweigh the perceived concerns. Granting consent allows users this self-determination with regard to their data. At the same time, however, it poses considerable practical problems as to the aspects of being voluntary and informed [Er17; LL18].

Additionally, users do not make purely rational decisions. Rather, decision making is affected by **cognitive biases and heuristics**. For example, optimism bias leads individuals to perceive themselves as less vulnerable than others [CLC10] and affect heuristics influences

the risk assessment in a way that users tend to underestimate risks when it is associated with things they like [Wa13]. These psychological means will always influence users' decision making to some extent. Here, the aim of data legislation and privacy preserving technologies should be to guarantee users an online environment in which they can freely decide what to share, following the principle of informational self-determination. This also includes data protection via technical and legal means so that users are protected to the largest extent possible.

Another deviation between users' and legal evaluation is the categorization of **location information**. GPS data can comprise distinct locations or even whole trajectories and is felt to be sensitive by users, but location is not part of the group of special categories of data under Art. 9 GDPR. Nevertheless, there is European legislation providing special rules for its processing. Directive 2002/58/EC, which was enacted to complement the former European Data Protection Directive 95/46/EC, defines it as any data processed in an electronic communications network, indicating the geographic position of the terminal equipment of a user of a publicly available electronic communication service. (Art. 2 lit. c) dir. 2002/58/EC) Art. 9 sec. 1 dir. 2002/58/EC allows the processing of this data only after its anonymization or with the consent of the users to the extent and for the duration necessary to provide an additional service. The scope of this provision, however, is limited to the regulation of data processing in the context of providing publicly available electronic communications networks (cf. Art. 3 sec. 1 dir. 2002/58/EC), e.g., by phone companies or closed user groups [Bü18]. For every other purpose and processing, the rules of the GDPR apply subsidiarly, treating location data connected to person as regular personal data. Hence, only in a very limited number of use cases relevant today, is location data further protected by the law. Considering the possibility of creating movement profiles of users through the analysis of location data and the threats to a person's freedom and rights such profiles bear, the protection granted by law seems insufficient. One reason for this situation is the ongoing reform process of the European data protection law. The directive 2002/58/EC relevant at hand will be renewed and transferred into a regulation in the future (cf. Proposal for as Regulation concerning the respect for private life and the protection of personal data in electronic communications and repealing Directive 2002/58/EG, COM 2017/010final-2017/03(COD)). Hence, the current legal state does not, yet, meet today's technical challenges. Despite the oncoming reform, the classification of location data as common personal data within the GDPR should be reconsidered and a higher level of protection for location data should be created.

The contrast between the European law and examples like the Chinese Social Credit System [Me17] shows that, due to sufficient legal regulations in Europe, users are protected from harmful aggregation of data, although this would be technically possible. The collection of license plate numbers to create a governmental "obedience score" is not conceivable, as the European **fundamental rights** exclude, e.g., the comprehensive assessment of a person's behavior and actions. Therefore, European users do not need to be concerned about possible consequences which could result from the collection of such data. The low

sensitivity evaluation of license plate number by the German sample shows that users indeed do not see many risks connected to that information. Reliance on the current data protection law can be one reason for the low sensitivity evaluation of the license plate number by the German sample compared to other cultures [Sc19]. This would indicate that users are aware of the legal protection granted to their data. The other hypothesis – users not being aware of potential risks – could also hold true and is important to consider regarding users' perception of privacy risks in general. Looking at data disclosure decisions through a privacy calculus lens, users need to evaluate the risks of data disclosure. To do that adequately, they need to be aware of these risks. However, they are not always privy to the legal protection and technical means [Eu18].

The authors argue that the main objective should not be informational heteronomy that is imposed by law over the users but **informational self-determination**. But this requires that users are aware and able to evaluate the risks of data disclosures. To empower users here, they need to be well informed. But educational measures at school are limited, especially because of the ever-evolving technical means. Thus, education and information must be available from a trusted source for all citizens, e.g., from a governmental website. Also, qualified media coverage and easy to understand consent forms are required. Finally, technological means to prevent unauthorized access to user data or the derivation of additional information from such data need to be further improved.

In **summary**, the comparison of the different perspectives on information sensitivity shows that users', technical, and legal views deviate to some degree. For example, the law grants protection to data categories as "special" categories that not all users perceive as especially sensitive. This can be seen as unproblematic as users are still able to freely disclose data, thereby giving their explicit consent to process this data. Other data categories, e.g., GPS data, are not given special privacy protection by the GDPR, but they are perceived as sensitive by the users and are, from a technological perspective, very revealing about the individual user. Here, the law nevertheless requires adequate data security measures for data processing, thus still providing protection. Rather, it is a problem that users, by allowing the processing of their data on the basis of consent without being fully aware of possible consequences, often thwart the safeguards of data protection laws which generally tries to limit the amount of processed data and the admissible purposes of its processing. For the premise of informational self-determination, it is of utmost importance to raise users' awareness about possible risks of data disclosure and the legal protection they are entitled to. As long as users decide to give their free, specific, informed, and unambiguous consent to the processing of their data as demanded by Art. 7 sec. 1, 4 No. 11 GDPR, the processing is in accordance with the law. It is, therefore, a manifestation of the users' informational self-determination which would, otherwise, turn into informational heteronomy. Finally, technological means must seek to unburden the users, i.e., provide the best data protection possible while not overly restricting the users' freedom of educated self-expression. In the current state, admittedly, perception of the importance between the three perspectives differs. However, the categories of data seen as sensitive by users and computer science

are adequately protected by law, although it grants other categories, which are not rated as particularly sensitive by users, more protection for the aforementioned reasons. The comparison of the three perspectives has shown how advanced the GDPR is in protecting users' privacy but for small adjustments. The effort of the GDPR to support the privacy interests of customers – structurally inferior compared to many companies and their economic interests – balances inequalities of social forces and strengthens the pluralism and democracy in digital societies.

Acknowledgments

This work has been funded by the German Federal Ministry of Education and Research (BMBF) under funding reference numbers 16KIS0443, 16KIS0444, and 16KIS0446. The responsibility for the content of this publication lies with the authors.

References

[Ar10] Armbrust, M. et al.: A View of Cloud Computing. Commun. ACM 53/4, pp. 50–58, 2010.

[AV18] Albers, M.; Veit, R.-D. In (Wolff, H. A.; Brink, S., eds.): Beck'scher Onlinekommentar Datenschutzrecht. 25th, 2018.

[Bu11] Bull, H. P.: Informationelle Selbstbestimmung – Vision oder Illusion? 2011.

[Bü18] Büttgen, P. In (Scheuerle, K.-D.; Mayen, T., eds.): Telekommunikationsgesetz. 3rd, 2018.

[BZ06] Barbaro, M.; Zeller Jr., T.: A Face Is Exposed for AOL Searcher No. 4417749, Accessed 2020-06-29, 2006.

[CLC10] Cho, H.; Lee, J.-S.; Chung, S.: Optimistic bias about online privacy risks: Testing the moderating effects of perceived controllability and prior experience. Computers in Human Behavior 26/5, pp. 987–995, 2010.

[DH06] Dinev, T.; Hart, P.: An Extended Privacy Calculus Model for E-Commerce Transactions. Information Systems Research 17/1, pp. 61–80, 2006.

[Er17] Ernst, S.: Die Einwilligung nach der Datenschutzgrundverordnung. Zeitschrift für Datenschutz 2017/3, ed. by Hoeren, T.; Schneider, J.; Selmayr, M.; Spies, A.; Wybitul, T., 2017.

[Eu18] European Commission: The Data Protection Act, `https://www.gov.uk/data-protection`, Accessed 2020-06-29, 2018.

[Fr18] Franzen, M. In (Franzen, M.; Gallner, I.; Oetker, H., eds.): Kommentar zum europäischen Arbeitsrecht. 2nd, 2018.

[GGV18] Gerber, N.; Gerber, P.; Volkamer, M.: Explaining the privacy paradox: A systematic review of literature investigating privacy attitude and behavior. Computers & Security 77/, pp. 226–261, 2018.

[He03] Heibey, H.-W. In (Roßnagel, A., ed.): Handbuch Datenschutzrecht. 1st, 2003.

[He13] Henze, M.; Hummen, R.; Matzutt, R.; Catrein, D.; Wehrle, K.: Maintaining User Control While Storing and Processing Sensor Data in the Cloud. International Journal of Grid and High Performance Computing (IJGHPC) 5/4, Dec. 2013.

[He14] Henze, M.; Hummen, R.; Matzutt, R.; Wehrle, K.: A Trust Point-based Security Architecture for Sensor Data in the Cloud. In (Krcmar, H.; Reussner, R.; Rumpe, B., eds.): Trusted Cloud Computing. Springer, Dec. 2014.

[HHW13] Henze, M.; Hummen, R.; Wehrle, K.: The Cloud Needs Cross-Layer Data Handling Annotations. In: 2013 IEEE Security and Privacy Workshops. May 2013.

[Hu13] Hunt, T.: Have I Been Pwned?, `https://haveibeenpwned.com`, Accessed 2020-06-29, 2013.

[KB18] Kühling, J.; Buchner, B. In (Kühling, J.; Buchner, B., eds.): Datenschutz-Grundverordnung, Bundesdatenschutzgesetz: DS-GVO/BDSG. 1st, 2018.

[Ko17] Kopp, H.; Mödinger, D.; Hauck, F.; Kargl, F.; Bösch, C.: Design of a Privacy-Preserving Decentralized File Storage with Financial Incentives. In: 2017 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW). Pp. 14–22, 2017.

[LBH] LeCun, Y.; Bengio, Y.; Hinton, G.: Deep Learning. Nature 521/, pp. 436–444.

[LL18] Leeb, C. M.; Liebhaber, J.: Grundlagen des Datenschutzrechts. Juristische Schulung 2018/6, pp. 534–537, 2018.

[LPK13] Lee, H.; Park, H.; Kim, J.: Why do people share their context information on Social Network Services? A qualitative study and an experimental study on users' behavior of balancing perceived benefit and risk. International Journal of Human-Computer Studies 71/9, pp. 862–877, 2013.

[Ma12] Masing, J.: Herausforderungen des Datenschutzes. Neue Juristische Wochenschrift 65/32, pp. 2305–2311, 2012.

[Ma18] Matzutt, R.; Hiller, J.; Henze, M.; Ziegeldorf, J. H.; Müllmann, D.; Hohlfeld, O.; Wehrle, K.: A Quantitative Analysis of the Impact of Arbitrary Blockchain Content on Bitcoin. In: Proceedings of the 22nd International Conference on Financial Cryptography and Data Security (FC). 2018.

[Me13] Meiklejohn, S. et al.: A Fistful of Bitcoins: Characterizing Payments among Men with No Names. In: Proceedings of the 2013 ACM Conference on Internet Measurement Conference. Pp. 127–140, 2013.

[Me17] Meissner, M.: China's Social Credit System: A big-data enabled approach to market regulation with broad implications for doing business in China, tech. rep., Mercator Institute for China Studies (Merics), 2017.

[MM12] Mayer, J. R.; Mitchell, J. C.: Third-Party Web Tracking: Policy and Technology. In: 2012 IEEE Symposium on Security and Privacy (S&P). Pp. 413–427, 2012.

[Mo12] Mothersbaugh, D. L.; "Foxx II", W. K.; Beatty, S. E.; Wang, S.: Disclosure Antecedents in an Online Service Context: The Role of Sensitivity of Information. Journal of Service Research 15/1, pp. 76–98, 2012.

[MPS13] Malheiros, M.; Preibusch, S.; Sasse, M. A.: "Fairly Truthful": The Impact of Perceived Effort, Fairness, Relevance, and Sensitivity on Personal Data Disclosure. In (Huth, M.; Asokan, N.; Čapkun, S.; Flechais, I.; Coles-Kemp, L., eds.): Trust and Trustworthy Computing. Pp. 250–266, 2013.

[Mü18] Müllmann, D.: Auswirkungen der Industrie 4.0 auf den Schutz von Betriebs- und Geschäftsgeheimnissen. Wettbewerb in Recht und Praxis 64/10, pp. 1177–1183, 2018.

[Na08] Nakamoto, S.: Bitcoin: A Peer-to-Peer Electronic Cash System, White paper, 2008.

[Ni10] Nissenbaum, H.: Privacy in Context: Technology, Policy, and the Integrity of Social Life. Stanford University Press, 2010.

[NS08] Narayanan, A.; Shmatikov, V.: Robust De-anonymization of Large Sparse Datasets. In: 2008 IEEE Symposium on Security and Privacy (S&P). Pp. 111–125, 2008.

[OR07] O'Reilly, T.: What is Web 2.0: Design Patterns and Business Models for the Next Generation of Software. Communication & Strategies 65/First Quarter, pp. 17–37, 2007.

[RCC18] Rosenberg, M.; Confessore, N.; Cadwalladr, C.: How Trump Consultats Exploited the Facebook Data of Millions, https://www.nytimes.com/2018/03/17/us/politics/cambridge-analytica-trump-campaign.html, Accessed 2020-06-29, 2018.

[Re12] Reding, V.: Sieben Grundbausteine der europäischen Datenschutzreform. Zeitschrift für Datenschutz 2012/5, ed. by Schneider, J.; Hoeren, T.; Selmayr, M.; Spies, A.; Wybitul, T., 2012.

[Sc19] Schomakers, E.-M.; Lidynia, C.; Müllmann, D.; Ziefle, M.: Internet Users' Perceptions of Information Sensitivity – Insights from Germany. International Journal of Information Management 46/, pp. 142–150, 2019.

[Se18] Serror, M.; Henze, M.; Hack, S.; Schuba, M.; Wehrle, K.: Towards In-Network Security for Smart Homes. In: Proceedings of the 2nd International Workshop on Security and Forensics of IoT (IoT-SECFOR). 2018.

[Si97] Simitis, S.: Die EU-Datenschutzrichtlinie – Stillstand oder Anreiz. Neue Juristische Wochenschrift 50/5, pp. 281–288, 1997.

[SLZ18] Schomakers, E.-M.; Lidynia, C.; Ziefle, M.: Hidden within a Group of People – Mental Models of Privacy Protection. In: Proceedings of the 3rd International Conference on Internet of Things, Big Data and Security (IoTBDS). Pp. 85–94, 2018.

[Sm12] Smith, M.; Szongott, C.; Henne, B.; von Voigt, G.: Big data privacy issues in public social media. In: 2012 6th IEEE International Conference on Digital Ecosystems and Technologies (DEST). 2012.

[Ta14] Taddicken, M.: The 'Privacy Paradox' in the Social Web: The Impact of Privacy Concerns, Individual Characteristics, and the Perceived Social Relevance on Different Forms of Self-Disclosure. Journal of Computer-Mediated Communication 19/2, pp. 248–273, Jan. 2014.

[Ve18] Veil, W.: Die Datenschutz-Grundverordnung: Des Kaisers neue Kleider. Neue Zeitschrift für Verwaltungsrecht 37/10, pp. 686–696, 2018.

[Wa13] Wakefield, R.: The influence of user affect in online information disclosure. The Journal of Strategic Information Systems 22/2, pp. 157–174, 2013.

[WB17] Wolff, H. A.; Brink, S. In (Wolff, H. A.; Brink, S., eds.): Beck'scher Onlinekommentar Datenschutzrecht. 25th, 2017.

[We17] Weichert, T.: "Sensitive Daten" revisited. Datenschutz und Datensicherheit 41/, pp. 538–543, 2017.

[Wo16] Wood, G.: Ethereum: A Secure Decentralised Generalised Transaction Ledger, White paper, 2016.

[Ye12] Yen, T.-F. et al.: Host Fingerprinting and Tracking on the Web:Privacy and Security Implications. In: The 19th Annual Network and Distributed System Security Symposium (NDSS). Internet Society, 2012.

[Zi17] Ziegeldorf, J. H.; Henze, M.; Bavendiek, J.; Wehrle, K.: TraceMixer: Privacy-Preserving Crowd-Sensing sans Trusted Third Party. In: 2017 Wireless On-demand Network Systems and Services Conference (WONS). Feb. 2017.

[ZMW14] Ziegeldorf, J. H.; Morchon, O. G.; Wehrle, K.: Privacy in the Internet of Things: threats and challenges. Security and Communication Networks 7/12, pp. 2728–2742, 2014.

[ZNP15] Zyskind, G.; Nathan, O.; Pentland, A.: Decentralizing Privacy: Using Blockchain to Protect Personal Data. In: 2015 IEEE Security and Privacy Workshops (S&PW). Pp. 180–184, 2015.

3rd Workshop on Smart Systems for Better Living Environments

SENSYBLE 2020: The 3rd Workshop on Smart Systems for Better Living Environments

Robert Kaiser[1], Ralf Dörner[1]

Keywords: Smart systems; embedded; mobile computing; artificial intelligence; computer graphics; computer vision

Message from the Chairs

Welcome to SENSYBLE 2020, the 3rd workshop on Smart Systems for Better Living Environments. We invite you to join us in participating in a workshop of lively discussions, exchanging ideas about a broad range of topics from smart embedded systems, systems engineering, telecommunication, mobile computing, computational theory, cryptography, to computer graphics, computer vision and artificial intelligence. For the first time, the workshop will be held in a virtual format, inviting participants from anywhere in the world. We do hope that this will enhance participation. A total of 16 papers have been selected for publication. All submissions received three or four reviews from members of the program committee, whom we would like to take this opportunity to thank for their prompt and thorough work. Also, we would like to thank the authors who submitted their work to this workshop. Last but not least, we would like to thank **you** for your interest and your participation! See you at the workshop, stay well!

The Workshop Chairs,

Robert Kaiser and Ralf Dörner
RheinMain University of Applied Sciences
Wiesbaden, Germany

[1] RheinMain University of Applied Sciences, firstname.lastname@hs-rm.de

doi:10.18420/inf2020_77

Program Committee

Uwe Brinkschulte *Goethe University of Frankfurt am Main*
Martin Gergeleit *RheinMain University of Applied Sciences*
Bodo Igler *RheinMain University of Applied Sciences*
Reinhold Kroeger *RheinMain University of Applied Sciences*
Detlef Krömker *Goethe University of Frankfurt am Main*
Matthias Pacher *Goethe University of Frankfurt am Main*
Sebastian Pape *Goethe University of Frankfurt am Main*
Kai Rannenberg *Goethe University of Frankfurt am Main*
Steffen Reith *RheinMain University of Applied Sciences*
Ulrich Schwanecke *RheinMain University of Applied Sciences*
Ulrike Spierling *RheinMain University of Applied Sciences*
Marcus Thoss *RheinMain University of Applied Sciences*
Adrian Ulges *RheinMain University of Applied Sciences*

Assessment of Current Intrusion Detection System Concepts for Intra-Vehicle Communication

Oleg Schell[1], Jan Peter Reinhard[2], Marcel Kneib[3], Martin Ring[4]

Abstract: Nowadays, vehicles incorporate a lot of electronics, which offer both advanced functionalities but also a great attack surface. Once having access to the communication network, an attacker can control critical functions like accelerating or steering. One possibility to detect these malicious intentions consists in the implementation of Intrusion Detection Systems (IDSs), which will even become mandatory via UN regulations in the future. Therefore, it is important for manufacturers and engineers to understand the opportunities and challenges of IDSs in the automotive environment. Giving an overview on these detection mechanisms is the primary goal of this elaboration. After the current vehicular communication architectures and protocols are outlined, potential attacks on the communication network are addressed. Afterwards, existing IDS concepts are presented, while the general requirements on these systems from an automotive perspective are stated and described next. Following the discussion on how to react to a detection, the elaboration is concluded with an outlook on what has still to be achieved to successfully integrate present IDSs into a vehicle.

Keywords: Automotive Security; Intrusion Detection System; Intra-Vehicle Communication

1 Introduction

With increasing functionality of the vehicular ecosystem, the number of electronic components and interfaces that are indispensable for safety realizations and provided services also increases [LOA19]. As a consequence, these advances widen the surface for the execution of cyber-physical attacks that no longer require physical access to the vehicle due to wireless interfaces like Bluetooth, WiFi or the Global System for Mobile Communication (GSM) [Lu14]. By exploiting these interfaces and the security vulnerabilities in the software of Electronic Control Units (ECUs), an adversary can get access to the internal communication network and remotely control crucial functionalities like steering, accelerating or braking [MV15]. It is evident that these possibilities can have severe consequences for both the driver and its environment. At this point, it must be mentioned that this threat does not only affect a single but several vehicle models, including Jeep [MV15], Tesla [NLD17] and BMW [Ca19], among others.

These circumstances made it quickly apparent that malicious activities on the intra-vehicle communication networks had to be detected and prevented. The latest efforts to realize this

[1] Bosch Engineering GmbH, Robert-Bosch-Allee 1, 74232 Abstatt, Germany, oleg.schell@de.bosch.com
[2] Hochschule RheinMain, Kurt-Schumacher-Ring 18, 65197 Wiesbaden, Germany, janpeterreinhard@gmail.com
[3] Robert Bosch GmbH, Mittlerer Pfad 9, 70499 Stuttgart, Germany, marcel.kneib@de.bosch.com
[4] Bosch Engineering GmbH, Robert-Bosch-Allee 1, 74232 Abstatt, Germany, martin.ring@de.bosch.com

doi:10.18420/inf2020_78

intent include UN regulations [UN20], which propose to implement countermeasures on a mandatory basis. One of the possibilities that they suggest, is the utilization of IDSs to provide a security measure on network basis. Since the demands placed on such systems in the automotive domain are different from those in a classic IT environment, this elaboration will outline different IDS approaches and their requirements for the implementation in vehicles. The presented concepts should serve as a guideline for engineers, while the subsequently addressed inadequacies and open questions regarding the realization should give researchers a direction to advance the topic of automotive IDSs.

Protocol	Rates	Description	Use Case
Linear Interconnected Network (LIN)	11.2 or 19.6 KBit/s	Cheap and simple protocol using a linear bus architecture for small intra-vehicle services.	Battery Monitoring, Window Lifter Control, Temperature Sensors
Media Oriented Systems Transport (MOST)	25, 50 or 150 MBit/s	Relatively expensive protocol, which provides high data rates for infotainment applications.	Audio Module, Navigation System, Infotainment
Controller Area Network (CAN)	125 or 500 KBit/s	Most common protocol for vehicular networks, with new CAN FD and CAN XL standards providing higher data rates.	Engine Control, Electrical Stability Control, Transmission Unit
Ethernet	100 MBit/s	Protocol, which is relatively new in the automotive domain and becomes more popular due to high data rates and cost.	ECU Flash Interface, Cameras, Radar, Network Backbones
FlexRay	5 or 10 MBit/s	Fault tolerant protocol with high bandwidths, which is not often used because of its complexity and high cost.	Steering Angle Sensor, Throttle Control, All-Wheel Drive

Tab. 1: Wired communication protocols for intra-vehicle data exchange based on [Al19; Hu19].

2 Automotive Network Architectures and Protocols

Every vehicle is a distributed system, which is made of ECUs communicating over different protocols. The ECUs represent the computing units of a vehicle and differ in performance, memory capacity and robustness, depending on the intended use. In this context, robustness means how well an ECU is protected against temperature, pressure and humidity changes, as well as its level of failure safety. A single ECU can have multiple network interfaces and thus send data over different media. Wired networks are more common in this regard, for which different protocols exist depending on the area of application as stated in Tab. 1. Besides these, there are also wireless standards like Bluetooth Low Energy or ZigBee which can also be deployed in the vehicle, but have not been widely used to date [Hu19].

The potential topologies, which can be used for these communication protocols, vary widely. Besides the star topology, where each device is connected to a central gateway, repeater or hub to route the data to its destination, each ECU can also be linked to its neighbors to form a ring topology. While furthermore a point-to-point connection is the easiest of

all topologies, communication protocols like LIN, CAN and FlexRay are usually realized in a bus topology, where every device is connected to a single communication line and transfers data in a broadcasting manner. Apart from the aforementioned interconnection methods for individual ECUs, the entire communication architecture is undergoing a transformation [He19]. The trend is shifting from an application-specific communication architecture towards a domain-specific one, where the ECUs in each domain communicate with protocols stated in Tab. 1, while Ethernet is used across domains for fast data exchanges. In the future, it is intended to create a centralized architecture that consists of a few high performance controller, which are connected to most of the ECUs or domains with Ethernet. Although Ethernet may replace protocols like FlexRay and MOST [Rö17], there is still the need to secure the individual communication sections and utilized protocols like CAN.

Requirement	Attack	Description
Authenticity	Spoof & Replay	Impersonating network participants without being noticed or replaying prerecorded messages on their behalf.
Confidentiality	Eavesdrop	Unauthorized access to data and information which is transmitted over the vehicular network.
Availability	Flood & Drop	Preventing operation of network participants by either withholding data or flooding the network with irrelevant messages impeding the transmission of relevant data.
Integrity	Manipulate	Manipulating content of transmitted messages in such a way that it remains hidden from the other network participants.

Tab. 2: Security requirements and potential attacks on intra-vehicle networks.

3 Attacks on Intra-Vehicle Communication

The motivation of an intrusion into the communication network of a vehicle is manifold and includes, among others, altering vehicular characteristics like engine performance or mileage, intruding into the driver's privacy or interfering into the control to cause harm. In order to achieve these goals, access to the communication networks is required first. Before vehicles were equipped with wireless interfaces, access could only be gained in a physical way, either by connecting directly to the communication wires or over the On-board diagnostics (OBD)-II port, which is used by workshops for diagnostic purposes. Nowadays, these wireless interfaces like Bluetooth or Wi-Fi of the telematic control unit represent an additional risk through which unauthorized access is possible. By exploiting security breaches and rewriting the software on this ECU [MV13], data can usually be both read and written by the adversary on the network to which this unit is connected to.

Once having access to the communication network, an adversary can perform different malicious actions due to the lack of security mechanisms. As exemplarily stated in Tab. 2, these attacks can be classified according to the security requirement they violate. For this

reason, appropriate security mechanisms have to be considered already during vehicle design or integrated afterwards to prevent such actions.

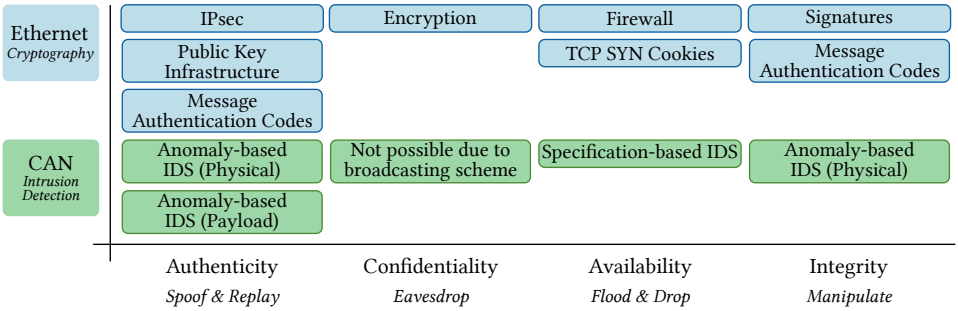


Fig. 1: Possible security mechanisms for vehicular Ethernet and CAN.

4 Intrusion Detection System Concepts

Compared to the remaining communication protocols from Tab. 1 which have been in use for several years, Ethernet is currently finding more and more its way into the vehicle, enhancing communication networks with a high bandwidth and low cost [Rö17]. Here, Ethernet does not provide security mechanisms by itself, they are mainly obtained by using higher level protocols like Transmission Control Protocol/Internet Protocol (TCP/IP). Since these have long been used in classical IT systems such as home computers, server applications or corporate networks, usual security mechanisms, some of which are listed in Fig. 1, can theoretically be applied in the automotive sector [HKD09]. Although the bandwidth allows the transmission of additional data for the proper operation of these security mechanisms, the real-time requirements must still be considered for their implementation [Rö17].

Taking CAN into consideration, the limited hardware resources and communication bandwidths commonly available, exacerbate the implementation of cryptography-based approaches [GM13]. Furthermore, the fact that CAN is already used in almost every vehicle renders a subsequent security provision for existing networks more difficult. This is especially critical, since CAN does not deploy any security mechanisms and can therefore be attacked successfully with ease [Al19]. To remedy this issue, the utilization of IDSs can be considered, which can retrofit basic security aspects into existing and future CAN networks. Since IDSs can be realized in various ways, the most prominent concepts are discussed in more detail, mentioning their merits and shortcomings. At this point it should be noted that the mentioned IDS concepts represent general methodologies and can therefore also be utilized for different communication protocols like Ethernet.

Signature-based IDS Using signature-based detection, the ongoing communication is continuously compared to known attack patterns in the IDS database like the sequence of

transmitted data or malicious instructions. Although usually used in anti-virus software for IT systems, the pattern matching procedure is a resource demanding process for vehicular ECUs. Further, the attack database has to be kept updated and distributed to the individual vehicles, while there is no possibility to recognize novel attacks. This is particularly critical as such patterns have not yet been extensively established for attacks on vehicles, which however are constantly increasing in number. At last, unlike classical IT systems where patterns can be shared, a manufacturer-specific signature cannot be used by another vendor, as other digital platforms are usually implemented. On the other hand, once these patterns are developed, this approach reliably detects known attacks with a low false alarm rate.

Specification-based IDS Communication properties like transmission schedules, communication partner or which ECU is eligible to transmit which messages on the respective network segment, are mostly specified during the design phase of a vehicle and usually do not change after deployment. The same applies to the static communication architectures which are not altered after establishment [Hu17]. An IDS can take advantage of this by considering the specification and establishing rules which are checked during the ongoing communication. For example, in this way it is possible to implement a firewall in units which interconnect several networks and which then are able to block transmissions not complying to the rules. Be it the deviation of data values from a predefined range or the propagation of unauthorized messages in a network, the specification-based approach is able to detect these easily and efficiently. Big disadvantages are that these rules have to be manually created by experts, which is error-prone and thus can lead to a high number of false alarms, while an adversary can circumvent the rules if he acts within acceptable limits.

Anomaly-based IDS Anomaly-based approaches work similarly to specification-based procedures in that they detect deviations from predefined behavior. The difference here is that the predefined behavior is learned by the system itself in the case of anomaly detection. This not only eliminates the need to set up rules but also enables the detection of unknown and novel attacks. Generally, the normal behavior can be learned based on different communication characteristics, which are briefly described in the following.

1. **Payload**: Mainly utilizing machine learning algorithms, IDSs of this category strive to establish a model of the message content and attempt to detect unusual data sequences that can be traced back to attacks. This approach would represent a promising option to detect different types of intrusions, if only the interrelationships were not so complex, the need for data not so high and the computational power not so demanding.
2. **Physical**: ECUs and their electronic components are subject to manufacturing imperfections, leading to small differences in the physical properties like clock timings and voltages. IDSs can utilize these small differences and implement sender identification mechanisms, with which unauthorized transmissions can be detected and the malicious ECU pinpointed. However, since the properties refer to the respective ECUs, an adversary can send unnoticed authorized messages with malicious content

from the compromised ECU. Further, in most cases, high performance analog-to-digital converters (ADCs) or timers are required to record even small differences.

Regardless of which characteristic is selected, anomaly-based approaches necessarily require trustworthy data to learn the normal behavior, whereby the time and data amounts required for this learning should not be neglected. Furthermore, one of the main reasons why this type of IDS has not yet been widely used is that it has a high false alarm rate [Al19], which is especially important when the driver is not to be distracted unnecessarily and when intrusions are not only to be detected but also actively prevented.

As shown in Fig. 1, different IDS concepts provide varying security measures for protocols like CAN. To establish a holistic security system, it is therefore essential to implement a combination of these concepts. For instance, specification-based methods could provide the first line of defense against rudimentary attacks, while anomaly-based procedures detect the presence of more sophisticated attackers. These *hybrid* IDSs allow the incorporation of both digital and physical characteristics making the resulting system more robust and reliable.

5 Requirements

For the design of an intrusion detection approach, different requirements play a major role in the automotive domain. While standards like ISO/SAE DIS 21434 [IS20] and UN regulations [UN20] mandate the realization of security management processes and name best practices, the criteria mentioned here represent a non-formal set of requirements. These are mainly derived from the challenges of implementing IDSs in the automotive sector stated by literature such as [LOA19] or [Al19] and make no claim to completeness.

The most significant difference of automotive systems compared to classical IT is the high importance of *Safety*. Therefore, an IDS is not allowed to affect data by delaying or removing it which may lead to the loss of safety relevant information. This also includes the fact that a restriction of the information availability by an IDS must not take place. Safety comes along with the requirement on *Performance*. In order to not delay data and to be able to analyze all exchanges messages even at high bandwidths, the IDS must have a certain computing performance. Furthermore, it requires a high detection performance to detect all attacks, while not generating false alarms. Although an IDS is a security measure by itself, it also has to meet different *Security* requirements. In this context, it is important that the system is not reducing the functionality and effectiveness of other security concepts such as firewalls or encryption, while not creating new exploits and critical security gaps. Other important requirements relate to the *Privacy* of the driver and other passengers. Because of the high connectivity of modern automotive systems, it is important that an IDS does not leak private data without permission to other systems. Especially, if the IDS uses a cloud-based back-end for incident analysis and transmits sensitive data. Finally, the *Update* of an IDS plays a crucial part for the requirements. In contrast to IDSs for classical IT systems which can be

updated almost at any time via the Internet, with vehicular IDSs it must be ensured that, for example, the rule update procedure of a specification-based approach does not open new security breaches and is not corrupted. If the update takes place over-the-air, short connections to road side infrastructures and disconnections must further be expected.

6 Post Detection and Outlook

An important question that has not yet been clarified in this elaboration is how to deal with intrusion detections in the automotive environment. In the literature a distinction is made between passive and active measures [LOA19], under which logging, notifying and preventing fall. Each of these three methods can be more or less beneficial in individual aspects. Logging, for example, stores information about the potential intrusion, which can then be read out in case of an incident to patch the security gap in the remaining fleet. Although logging does not distract or hinder the driver, the large amount of information must first be stored and subsequently analyzed in time-consuming manual work. Considering to notify the driver in case of an intrusion, the danger is immediately apparent and actions can be carried out. False alarms play a crucial part here, because even with transmission rates of several tens of milliseconds and a low false alarm rate, the driver could be mistakenly warned several times a minute. These false alarms become even more serious if active prevention is taken into account. If safety-relevant data is incorrectly recognized as an attack and on this basis prevented, it becomes apparent that such actions can have far-reaching consequences for the passengers and their environment.

Only after it is clarified how to deal with these detections, IDSs can be effectively put into utilization, whereby further challenges have to be taken into account. Up to now, the throughout implementation of security in a vehicle is regarded as a matter of course by the customer [Hu17]. Therefore, manufacturers keep the available resources for IDSs as low as possible, which stands in contradiction to increasing data amounts and complexity. Achieving lower latencies and real-time capability, which are particularly relevant for safety-critical tasks, ECUs require more hardware resources and computing power. These requirements are especially true for anomaly-based IDSs, which currently receive the greatest focus, as they are most promising to detect sophisticated attacks [LOA19]. For the evaluation of anomaly-based approaches, individually recorded data from test vehicles are often used. Yet, to ensure a better comparability of existing approaches, a publicly available data set with respective communication characteristics is required. In the end, knowledge of different disciplines like artificial intelligence, automotive systems and electrical engineering are crucial for the design and consolidation of different security concepts. Only if this knowledge comes together, a holistic security system can be developed which is able to recognize or prevent not only rudimentary but also the presence of advanced attackers. For this purpose, different concepts have to be employed in a joint approach to realize the essential security requirements. In doing so, the non-formal criteria like safety, performance or privacy must be considered for both current and future communication architectures.

References

[Al19] Al-Jarrah, O. Y.; Maple, C.; Dianati, M.; Oxtoby, D.; Mouzakitis, A.: Intrusion Detection Systems for Intra-Vehicle Networks: A Review. IEEE Access 7/, pp. 21266–21289, 2019, ISSN: 2169-3536.

[Ca19] Cai, Z.; Wang, A.; Zhang, W.; Gruffke, M.; Schweppe, H.: 0-days & Mitigations: Roadways to Exploit and Secure Connected BMW Cars. Black Hat USA/, 2019.

[GM13] Groza, B.; Murvay, S.: Efficient Protocols for Secure Broadcast in Controller Area Networks. IEEE Transactions on Industrial Informatics 9/4, 2013.

[He19] Helge Zinner Julian Brand, D. H.: Automotive E/E Architecture evolution and the impact on the network. IEEE802 Plenary, March 2019, 802.1 TSN/, 2019.

[HKD09] Hoppe, T.; Kiltz, S.; Dittmann, J.: Applying intrusion detection to automotive IT-early insights and remaining challenges. Journal of Information Assurance and Security (JIAS) 4/, pp. 226–235, Jan. 2009.

[Hu17] Humayed, A.; Lin, J.; Li, F.; Luo, B.: Cyber-physical systems security—A survey. IEEE Internet of Things Journal 4/6, pp. 1802–1831, 2017.

[Hu19] Huang, J.; Zhao, M.; Zhou, Y.; Xing, C.: In-Vehicle Networking: Protocols, Challenges, and Solutions. IEEE Network 33/1, pp. 92–98, Jan. 2019.

[IS20] ISO/SAE DIS 21434: Road Vehicles – Cybersecurity engineering, Standard, Geneva, CH: International Organization for Standardization, 2020.

[LOA19] Lokman, S.-F.; Othman, A. T.; Abu-Bakar, M.-H.: Intrusion detection system for automotive Controller Area Network (CAN) bus system: a review. EURASIP Journal on Wireless Communications and Networking 2019/1, p. 184, 2019.

[Lu14] Lu, N.; Cheng, N.; Zhang, N.; Shen, X.; Mark, J. W.: Connected Vehicles: Solutions and Challenges. IEEE Internet of Things Journal 1/4, 2014.

[MV13] Miller, C.; Valasek, C.: Adventures in automotive networks and control units. Def Con 21/, pp. 260–264, 2013.

[MV15] Miller, C.; Valasek, C.: Remote exploitation of an unaltered passenger vehicle. Black Hat USA 2015/, p. 91, 2015.

[NLD17] Nie, S.; Liu, L.; Du, Y.: Free-fall: Hacking tesla from wireless to can bus. Briefing, Black Hat USA/, pp. 1–16, 2017.

[Rö17] Röder, J.: Automotive Ethernet - Die Zukunft der vernetzten Fahrzeugarchitektur - The future of in-vehicle data Management./, July 2017, URL: https://www.vdi-wissensforum.de/news/automotive-ethernet/.

[UN20] UN Task Force on Cyber Security and Over-The-Air issues: Proposal for the 01 series of amendments to the new UN Regulation on uniform provisions concerning the approval of vehicles with regard to cyber security and of cybersecurity management systems. 2020.

R. Reussner, A. Koziolek, R. Heinrich (Hrsg.): INFORMATIK 2020,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2021 883

Development of a Vehicle Simulator for the Evaluation of a Novel Organic Control Unit Concept

Melanie Brinkschulte[1]

Abstract: New challenges in the field of automotive systems (e.g. autonomous driving) require innovative and highly robust vehicle architectures. These are intended to increase the reliability and fault tolerance of the system and therefore realize the transition from Fail-Save to Fail-Operational behavior. Organic computing is a possible approach to achieve these goals. Based on an artificial hormone system and an artificial DNA, a novel organic control unit concept exists.
In this paper we introduce an evaluation tool for this novel concept. Therefore, a simulator physically models the longitudinal and lateral dynamics of a vehicle. For an easy handling, visualization and reproducibility of experiments, an user interface and a scripting language are designed. In extensive evaluation runs the usability of the vehicle simulator is tested. Hereby, real vehicle data is used.

Keywords: Vehicle Simulator; Organic Computing; Fail-Operational

1 Introduction

In this paper we introduce an evaluation tool for a novel organic control unit concept based on an artificial hormone system (AHS) [vBP11] and an artifical DNA (ADNA) [Br15] . Therefore, a simulator physically models the longitudinal and lateral dynamics of a vehicle. The parameters of the vehicle (weight and measures, engine and gear parameters, brake parameters, air and roll resistance, . . .) can be individually chosen. For an easy handling, visualization and reproducibility of experiments, an extensive user interface and a scripting language is designed. Also, this simulator allows the evaluation of various automotive control components (ECUs) like ABS, ASR, power steering and cruise control. All input data like brake, throttle and steering positions can be given by the user interface or the scripting language. In addition, both input options can be used simultaneously. The output values like brake force, wheel speed, vehicle speed, steering angles, etc. are visualized in the user interface, timestamped and written to a logfile for detailed examination. Thereby, the user can choose which physical value is logged as well as the time resolution of the logging process. By fault injection, ECU failures at run-time can be induced at arbitrary times during a simulation run.
This paper is structured as follows: Section 2 gives a short overview of the designed physical models. Following, Section 3 describes the architecture and the interaction of the

[1] University of Mannheim, Chair of Information Systems II, Schloss, 68131 Mannheim, Germany & Goethe University Frankfurt am Main, Computer Science Department, Robert-Maier-Str. 11-15, 60325 Frankfurt am Main, Germany, brinkschulte@uni-mannheim.de

doi:10.18420/inf2020_79

components of the simulator. Section 4 presents an extract of the extensive evaluation and Section 5 discusses related work. Finally, Section 6 concludes this paper.

2 Models

In this work multiple physical models (vehicle dynamics, steering, brake and engine) are designed and used. For reason of space, we can only shortly enumerate these models here. For more details, please refer to [Br19].
The **vehicle dynamics model** is based on the linear single-track model. However, this is not sufficient for the desired purpose and is therefore extended (by adding of longitudinal dynamics, frictional conditions and accuracy by removing the small angle approximation, extension to an rudimentary two-track model) to an efficient nonlinear two-track model without small angle approximation. The **steering model** is a speed-dependent steering system with optional steering assistance. The **brake model** includes characteristic curve mappings and brake cylinder delay. The **engine model** includes an optional adjustable four-wheel drive, as well as drive delay and dead times. Furthermore, a simple **gearbox model** was realized.

3 Simulator

In this section, the architecture as well as the communication and interaction of the components (user interface, physical models, simulator-sensor/actuator interface) of the vehicle simulator is shown. The vehicle simulator is implemented in C++ while Qt 5.11.1 is used to implement the graphical user interfaces.

3.1 Architecture

The simulator consists of three parts: the physical models of the vehicle, the user-interface and the simulator sensor/actuator interface (Figure 1). The physical models have internal state data (e.g. speeds, distances travelled, angles, etc.) and receive vehicle data (e.g. the vehicle mass, vehicle dimensions, etc.), environmental data (e.g. the static/sliding friction value between road and tires) and input data (e.g. a steering angle, an accelerator pedal position, etc.). They then use these to calculate the output data (e.g. forces and accelerations). Through the input and output data, the physical models are connected via the simulator sensor/actuator

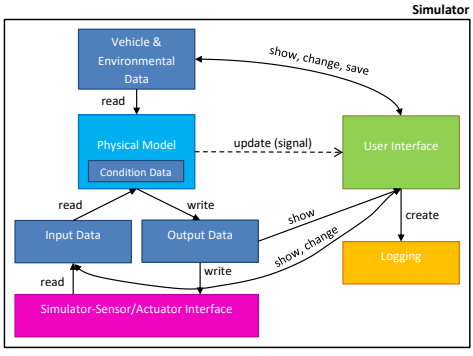


Fig. 1: Architecture of the simulator

interface to the AHS and the ADNA that realize the vehicle's control units. The user interface can display, change and save vehicle condition and environment data. Furthermore, the settings for the creation of a parameterizable log files can be defined in the interface. The visualization consists of a top view with optional fade-in of different force, track and speed vectors as well as a side view, which shows the wheel speeds and the adhesion conditions. Furthermore the simulator is real-time capable. The two-layer real-time architecture is shown in Figure 2. The outer layer uses a $10ms$ period to ensure smooth and jitter-free

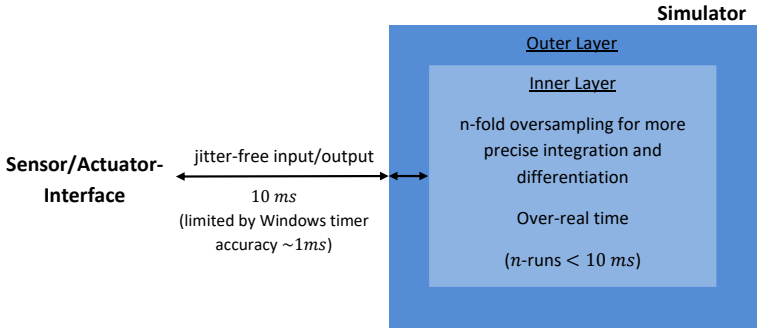


Fig. 2: Real-time architecture of the vehicle simulator

real-time input/output. This cycle time was chosen because the Windows timers used in the implementation have an accuracy of about $1ms$. In the inner layer, a n-fold oversampling (n-fold execution of a simulation run) is performed to achieve a more precise integration and differentiation. The number of passes of the inner layer can be individually adjusted by the user. The only condition is that the time required for this number of runs is less than $10ms$ (run time of outer layer). This means that over-real time is present there.

3.2 Interaction

In Figure 3 the interaction and data transfer between the user interface, the physical models and the simulator-sensor/actuator interface is shown. The physical models are divided into the submodels vehicle dynamics, brake, engine and steering. The sensors of the Simulator Sensor/Actuator Interface receive their inputs from the physical submodels and from the user (e.g. brake pedal sensor, accelerator pedal sensor, etc.). The user has two possibilities to create his inputs. On the one hand, he can make entries via the user interface to directly control the vehicle and influence its environment. On the other hand, in order to enable precisely repeatable experiments, it is possible to specify input in the form of a script file. The input data is then converted by the control units into corresponding actuator values (e.g. brake cylinder control, drive control, steering angle control, etc.) according to the artificial DNA running on them. These values then enter the physical models of the steering, the brake and the engine. The outputs of these models are then passed on to the vehicle dynamics model, which in turn generates new sensor signals in a closed control loop.

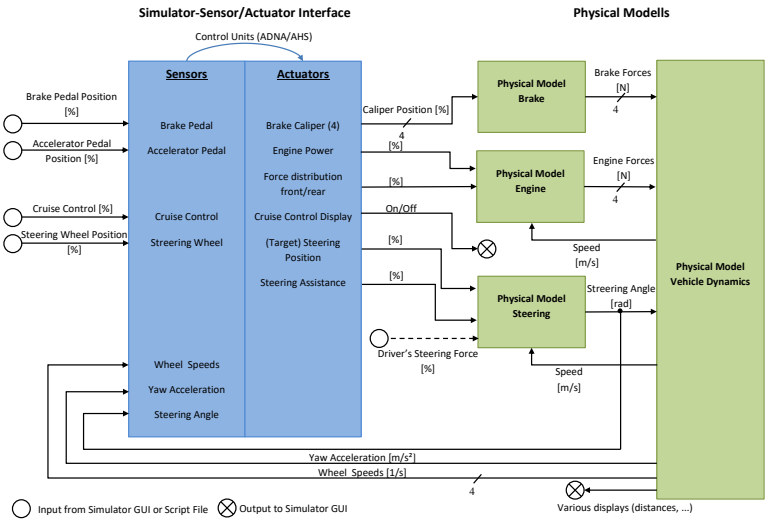
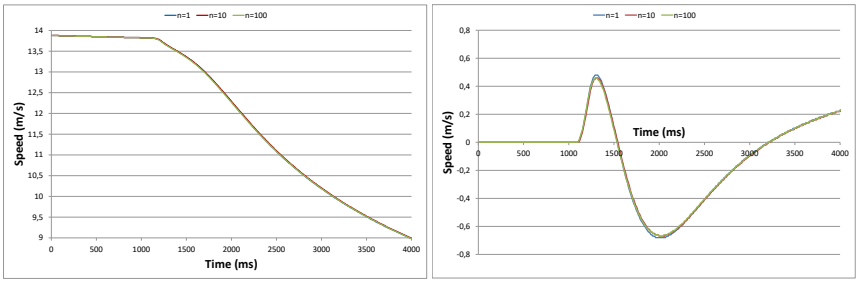


Fig. 3: Interaction and data transfer between the user interface, the physical models and the simulator-sensor/actuator interface

4 Evaluation

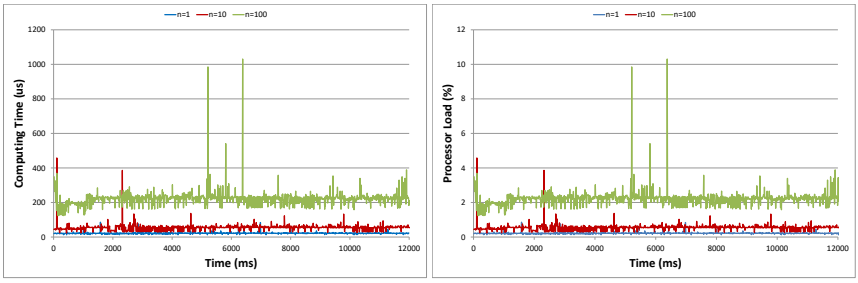
Extensive evaluations of the presented work are made. Unfortunately in the scope of this paper only the most important evaluation results can be presented in detail.
At first, the trade-off between simulator processor load and accuracy is evaluated. Therefore, a simple reproducible experiment (vehicle initiates a curve at a given starting speed of $50\frac{km}{h}$, the static friction is set to a value so that the vehicle doesn't slide) with different numbers of steps ($n = [1, 10, 100]$) per simulation period in the inner layer of the simulator is used. To determine the gain in accuracy, the speed in x- and y-direction are compared exemplary. In the diagrams, minimal deviations of the curves from each other can be seen.



(a) Comparison of the speed in x-direction (b) Comparison of the speed in y-direction

Fig. 4: Comparison of speed in x- and y-direction at $n = 1$, $n = 10$ and $n = 100$ steps per simulation period

The speed in x-direction (Figure 4a) shows the smallest deviations. These never exceed 0.1%, hence we have nearly one line in the figure. For the speed in y-direction (Figure 4b), slightly larger deviations of a maximum of 4% are visible. Nevertheless, the curves are almost identical at $n = 10$ and $n = 100$ steps per simulation period. Only the curves resulting from only one step per simulation period differ slightly more from the others at two points ($1280ms - 1340ms$ and $1810ms - 2140ms$). The resulting computing time for a simulation period and the resulting processor load are shown in Figure 5. As expected, the required computation time per simulation period and thus the processor load increases with growing number of simulation steps n. This results in a maximum calculation time of $1.031ms$ with $n = 100$ which corresponds to a maximum processor load of 10%. It turns out that the



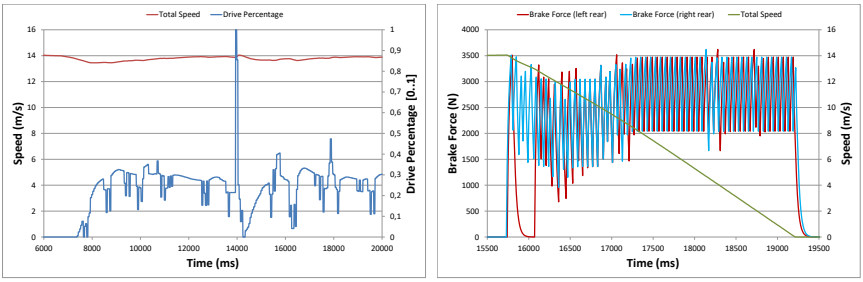
(a) Computing time (b) Processor load

Fig. 5: Required computing time and resulting processor load at $n = 1$, $n = 10$ and $n = 100$ steps per simulation period

model works with high accuracy at low computational effort. Especially $n = 10$ turns out to be a good choice with low overhead and high accuracy.
Next, the longitudinal and lateral dynamics have been evaluated with different experiments[2]. The vehicle is also operated in the non-linear range (skidding). It turns out the simulator

[2] Longitudinal: reality comparison of acceleration time from 0 to 100 $\frac{km}{h}$, maximum final speed and braking distance from 100 $\frac{km}{h}$ to standstill. Lateral: pure steering and steering in combination with braking experiments

behaves as expected and the results are very close to those of a real car. In this paper, we focus on the evaluation of the closed control loop between simulator and control units by means of a cruise control and ABS. The aim of the simulator is to create an evaluation tool for the AHS together with the ADNA as an organic control unit concept. This ECU concept should keep the system operational even in case of component failure and thus show the desired *Fail-Operational* behaviour.
Therefore, in further evaluation the behaviour of the closed control loop in case of failure of individual DNA processors is examined. Failures of processors (ECUs) were added during ABS braking in a corner or speed regulation by cruise control. In Figure 6a the failure of the processor is clearly shown by a peak in the drive power percentage. The processor failure causes the wheel speed sensors of the front wheels to be temporarily lost. The cruise control implemented uses this data to determine the speed of the vehicle. If it now receives no more rotation speed, it assumes that the vehicle has a speed of $0\frac{m}{s}$ and thus accelerates at maximum to reach the set target speed. As soon as the AHS/ADNA has distributed the wheel speed detection to the remaining processors in one of the next hormone cycles by self-healing, the failure is eliminated and the cruise control works correctly again. Here the failure lasts about $100ms$. In Figure 6b the failure of a processor is also clearly visible.



(a) Simulation results of the speed experiment in case of processor (ECU) failure during cruise control

(b) Simulation results of the ABS corner brake experiment in case of processor (ECU) failure

Fig. 6: Simulation results of processor (or control unit) failure experiments

The braking force at the rear left wheel drops to zero shortly after the start of braking at an approximate simulation time of about $15860ms$. After a simulation time of about $16060ms$, the braking force is again controlled by the ABS. The failure was therefore corrected after about $200ms$. In Figure 7 the comparison between the speeds and the distances covered with and without failure is shown. The speed curve during failure shows a slight bend, which is caused by the short-term reduced braking effect during failure. As a result, the vehicle comes to a halt about $250ms$ later than without failure. At the time the vehicle comes to a halt without failure, the vehicle still has a speed of $1.2\frac{m}{s}$ during the failure. This results in an extension of the braking distance of approximately $0.15m$. These evaluations show the suitability of the developed vehicle simulator for the intended application. The simulator is able to simulate failures of processors or control units, which allows to investigate and analyze such failure scenarios.

5 Related Work

Vehicle modeling has always been of big importance for the automotive industry [WSK11]. A general overview of vehicle models can be found in [SHB13] and [MW14]. Here a suitable compromise between model complexity and the number of model parameters or computing time should be found for the respective application, as is also described in [Un13]. This enables an optimal use of the respective model within the scope of the intended application. Highly complex multi-mass models are used, for example, to evaluate the driver's driving experience [Un13] or in comfort simulation [Am13]. If, on the other hand, the lane control of vehicles [Ar15], [He09] or the evaluation of vehicle measurement data during road tests [Se05] are concerned, simple models such as the linear single lane model are usually sufficient.

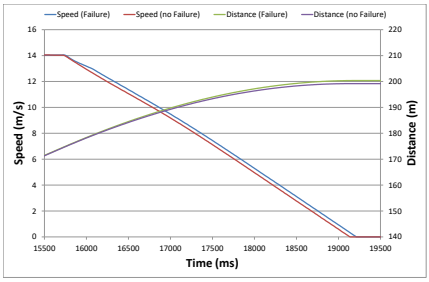


Fig. 7: Comparison between speeds and distances in the ABS-experiments with and without failure

In the presented work, the focus is on vehicle dynamics in connection with a novel, robust control unit concept. In case of an ECU failure, the evaluation of the lateral and longitudinal dynamics of the vehicle is of great importance, whereby the ECUs receive information from all four wheels. In the event of skidding (oversteer and understeer), the vehicle also leaves the linear range. The linear single-track model was therefore too simplified and not suitable for the desired purpose. However, driving comfort was also not in the focus of the work, i.e. there was no need to use a complex multi-mass model.

For this reason, an adapted model was developed, which extends the linear single-track model to a nonlinear model with two-track components. With this model, meaningful simulations for the evaluation of the novel ECU concept can be performed on the basis of an ADNA (especially regarding failure and robustness of ECUs), while the model complexity remains at a reasonably low level.

6 Conclusion

In this paper a vehicle simulator is presented as an evaluation tool of an organic control unit concept (represented by the AHS in combination with the ADNA) in the automotive field. For this purpose, physical models for steering, brake, engine and vehicle dynamics are developed, validated and implemented. With these models a comprehensive evaluation of driving situations for the evaluation of the organic ECU concept or robust ECU is possible. To enable reproducible experiments, a script language was developed and implemented. This allows flexible experiments under identical conditions and thus enables the comparison of different ECU concepts against each other. In future work, the failure behavior of more sophisticated control units for autonomous driving will be evaluated with this tool.

Bibliography

[Am13] Amelunxen, Hendrik: Fahrdynamikmodelle für Echtzeitsimulationen im komfortrelevanten Frequenzbereich. Dissertation, Universität Paderborn, 2013.

[Ar15] Arndt, Albrecht: Querregelung eines spurgeführten Modellfahrzeugs. chapter 5 Modellbildung, 2015.

[Br15] Brinkschulte, Uwe: An artificial DNA for self-descripting and self-building embedded real-time systems. In: Concurrency and Computation: Practice and Experience, volume 28. Wiley Online Library, 2015.

[Br19] Brinkschulte, Melanie: , Entwicklung eines Fahrzeugsimulators zur Evaluation eines neuartigen organischen Steuergerẽkonzepts im Automotiven Bereich, 2019. Masterthesis at Goethe Universität Frankfurt am Main.

[He09] Hensel, Enrico: Führungskonzept eines autonomen Fahrzeugs, Vorbetrachtung und Bewegungsmodell. Seminarbericht, Hochschule für Angewandte Wissenschaften Hamburg, 2009.

[MW14] Mitschke, Manfred; Wallentowitz, Henning: Dynamik der Kraftfahrzeuge. Springer, 2014.

[Se05] Sentürk, Fikret: Durchführen von Fahrversuchen hinsichtlich einer Optimierung von FHTW-Fahrdynamikfahrzeug. Diplomarbeit, Fachhochschule für Technik und Wirtschaft Berlin, 2005.

[SHB13] Schramm, Dieter; Hiller, Manfred; Bardini, Roberto: Modellbildung und Simulation der Dynamik von Kraftfahrzeugen. Springer, 2013.

[Un13] Unterreiner, Michael: Modellbildung und Simulation von Fahrzeugmodellen unterschiedlicher Komplexität. Dissertation, Universität Duisburg-Essen, 2013.

[vBP11] von Renteln, Alexander; Brinkschulte, Uwe; Pacher, Mathias: The Artificial Hormone System - An Organic Middleware for Self-organising Real-Time Task Allokation. In (Müller-Schloer, Christian; Schmeck, Hartmut; Ungerer, Theo, eds): Organic Computing - A Paradigm Shift for Complex Systems, chapter 4.4. Springer, 2011.

[WSK11] Wiedemann, Jochen; Schröck, David; Krantz, Werner: Fahrzeugdynamik, Themenheft Forschung, volume 7. Universität Stuttgart, 2010-2011.

R. Reussner, A. Koziolek, R. Heinrich (Hrsg.): INFORMATIK 2020,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2021 891

Effects of the Sampling Technique on Sender Identification Systems for the Controller Area Network

Marcel Kneib[1], Oleg Schell[2]

Abstract: As a result of the ongoing development of vehicle electronics and additional wireless communication interfaces, the possibilities for attacks and their negative consequences are increasing. Once an attacker has obtained access to the internal vehicle communication, in the case of the Controller Area Network (CAN) the attacker is able to forge all messages of the connected Electronic Control Units (ECUs) without a receiving ECU being able to recognize any suspicious behavior. The use of cryptographic methods is only possible to a limited extent due to restricted resources of the ECUs, which is why sender identification systems have been presented which are able to detect these kind of attacks. Presented approaches use different procedures to capture the analog signals on which the detection of attacks respectively the identification of the sender is based. This work shows that the impact on the performance of the sender identification system by the different sampling methods is minimal and therefore the selection of the appropriate technique can be mainly based on the available resources and the communication structure of the corresponding vehicle platform. This is shown on the one hand by the direct analysis of the analog signals captured from a real vehicle as well as by an evaluation of the previously introduced sampling methods using a recently published sender identification system. In addition, an assessment of the procedures based on different parameters shows which method is to be preferred for which application.

Keywords: Automotive Security; Sender Identification; Intrusion Detection

1 Introduction

The connectivity of modern vehicles, as well as the associated amount of interfaces, is constantly increasing. This trend does not only allow additional comfort functionalities and complex driver assistance systems, but also offers additional possibilities to attack a vehicle and its functions [HKD11; LL18]. Evidence that this is not only a theoretical threat was demonstrated by the attack of Miller and Valasek [MV15], as well as the latest research of the Tencent Keen Security Lab [Ca19]. Due to the absence of authenticity in the Controller Area Network (CAN) [Ro91], which is still the most commonly used bus technology in the automotive domain, an Electronic Control Unit (ECU) cannot check whether a received message was sent by a legitimate sender. This enables the forgery of messages, i.e. the execution of impersonation attacks. This problem still exists for its successors, CAN with flexible data rate (CAN-FD) [Ro12] and CAN-XL [CA20]. Unfortunately, the use of cryptographic methods is limited due to the constrained resources of the platforms used in

[1] Robert Bosch GmbH, Mittlerer Pfad 9, 70499 Stuttgart, Germany, marcel.kneib@de.bosch.com
[2] Bosch Engineering GmbH, Robert-Bosch-Allee 1, 74232 Abstatt, Germany, oleg.schell@de.bosch.com

doi:10.18420/inf2020_80

vehicles and the low payload and bandwidth of CAN. As an alternative or in combination with attack detection, methods have been presented in the past which provide sender identification on the basis of analog signals [Kn20]. Due to the static configuration of the internal vehicle communication, such systems allow to verify whether a message was sent by a valid ECU. For identification, however, the signals of CAN messages must first be recorded, for which the considered sender identification approaches suggest different procedures. While some methods capture the entire signal in order to extract the signal characteristics [Ch18; KH18], others concentrate on specific parts [Fo19] or individual points of a frame to determine the sender [KSH20]. The signal recording procedure has a corresponding effect on various properties, such as hardware requirements, cost, complexity and signal quality. In addition, the requirements and architecture of the actual system also have a major influence on the type of recording. This paper presents the different recording approaches and analyzes the associated effects on the relevant properties of sender identification systems for CAN. In addition, the associated performance is analyzed using the example of the recently presented work Edge-based Sender Identification (EASI) [KSH20] utilizing data from a series production vehicle. Furthermore, this work presents the individual application possibilities of the different sampling techniques, so that the reader is able to assess the optimal methodology with corresponding effects and constraints according to the respective requirements.

2 Sampling Approaches

For the CAN communication standard components are used, which can be produced in large quantities and very cost-effectively. These components only provide the connected microcontroller with access to the digital content of the message and not to the analog signals. For this reason, the actual recording of the signals must be independent of the existing hardware and therefore has to be considered and implemented by the respective sender identification approaches.

Since in principle every ECU can send a message at any time, it must be ensured that parallel transmission and thus corruption of the currently sent message does not occur. For this purpose, an ECU first checks whether the bus is free and then begins to send its message. In simplified form, the message consists of an unique message identifier, which also defines the priority of the message, and the associated content. During the transmission of the identifier it can happen that other ECUs start the transmission. The sending ECUs check whether their currently transmitted signal corresponds to the signal currently on the bus, and if not, the transmission of the respective ECU is stopped. The characteristics of the analog signals transmitted in this situation cannot be used for recognition as they contain characteristics of several ECUs. Therefore, all approaches focus on the segment succeeding the identifier.

As introduced in [KSH19], each frame consists of several symbols which represent the transmitted bits on the bus. As there are different kinds of symbols, the most approaches [Ch18; Fo19; KH18] first group those symbols according to the voltage transitions. There are four

different transition groups g, the rising ($g = 1$) and falling ($g = 2$) edges, and the stable high ($g = 3$) and low ($g = 0$) levels. The k-th symbol of a frame m sent by ECU e is defined by

$$S_k^{g,(e,m)} = (x_1, ..., x_l) \tag{1}$$

where, x_i, $i \in \{1, .., l\}$ are the individual voltage values of the symbol. Hence, elements of this l-tuple can be accessed according to the projection of the set theory with $S_k^{g,(e,m)}[i] = x_i$. Each symbol group, defined by

$$G^{g,(e,m)} = \bigcup_{k=1}^{K} S_k^{g,(e,m)}, \tag{2}$$

can contain a different amount K of symbols. Since the subsequent calculation of the characteristics from these symbols is computationally expensive, some approaches [Ch18; KH18] initially calculate an *average symbol* per group according to Equation (3), respectively use the averaged symbol directly as characteristic [Fo19].

$$\overline{S}^{g,(e,m)} = (\frac{1}{K}\sum_{k=1}^{K} S_k^{g,(e,m)}[1], \ldots, \frac{1}{K}\sum_{k=1}^{K} S_k^{g,(e,m)}[l]) \tag{3}$$

Another possibility is to utilize only a *single symbol* for the calculation of the characteristics. Without loss of generality, the first symbol of a group is considered for the further calculations, defined by

$$\mathring{S}^{g,(e,m)} = S_1^{g,(e,m)}. \tag{4}$$

The third variant, the *composite symbol* [KSH20], assembles the symbol from individual sample points of several symbols in a group. Based on the available number of samples per symbol L and the number of samples to be used per symbol P, the number of required symbols $K = \lceil \frac{L}{P} \rceil$, $P \leq L$ is given. For $L = 20$, $P = 2$, $K = 10$ according to

$$\tilde{S}^{g,(e,m)} = \bigcup_{p=1}^{P}\bigcup_{k=1}^{K} S_k^{g,(e,m)}[K * (p-1) + k], \tag{5}$$

the resulting sample points are illustrated in Fig. 1 for $g = 1$.

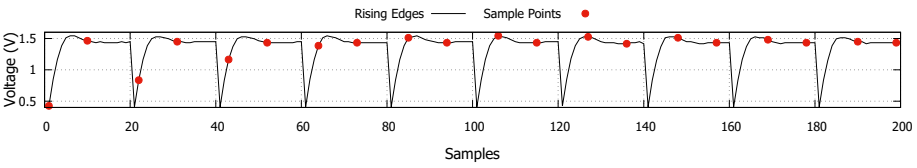


Fig. 1: Considered sampling points of the rising edges for the composite symbol.

3 Signal Analysis

3.1 Data set

For the initial analysis of the effect on the signal quality, a data set is reused which has already been utilized for the evaluation of sender identification approaches [KH18; KSH20]. The signals were recorded from a Fiat 500 which has six internal ECUs, each using up to seven different identifiers. In order to increase the number of ECUs, two additional Raspberry Pis were connected, each equipped with a CAN shield. One Raspberry Pi was connected to the bus in the trunk, while the other Pi was attached to the on-board diagnostics port together with a PicoScope 5204 at a sampling rate of 500 MS/s and a resolution of 8 bit. Since this data set is only slightly affected by changing environmental conditions, it allows the effects of the different sampling approaches to be analyzed as accurately as possible.

3.2 Metrics

In order to allow an approach-independent evaluation of the signal quality, the metrics *intra- and inter-distance* as well as their combination, the *inter-intra-distance* are used. The distance between two symbols $\mathcal{S}$ and $\mathcal{S}'$ regarding the considered sampling approaches, is defined by

$$Symbol\text{-}Distance(\mathcal{S}, \mathcal{S}', g, (e,m), (e',m')) = \frac{1}{L}\sum_{l=1}^{L}\left|1 - \frac{\mathcal{S}^{g,(e,m)}[l]}{\mathcal{S}'^{g,(e',m')}[l]}\right|. \tag{6}$$

The intra-distance, calculated by Equation (6) with $\mathcal{S} = \mathcal{S}', e = e'$ and $m \neq m'$, is used to evaluate the deviations of the symbol between all frames of a single ECU. In order to additionally analyze the symbol differences between the ECUs, the inter-distance is utilized, which is calculated by Equation (6) with $\mathcal{S} = \mathcal{S}'$ and $e \neq e'$ for all ECUs and their associated frames. Finally, the inter-intra-distance, i.e. the difference between the inter- and intra-distance, is used as the metric to assess the distance between the ECUs, taking into account the magnitude of the deviations of the ECUs with respect to the considered sampling approach.

Furthermore, the statement must be evaluated that the differences of the variations of the sampling approaches are negligible, since they are in the range of the natural variation of the symbols within a frame [KSH20]. Therefore, first the natural deviation of the symbols in the data set within a frame is calculated with $\mathcal{S} = S_k^{g,(e,m)}$ respectively $\mathcal{S}' = S_{k'}^{g,(e,m)}$ for $k \neq k'$. Following, for the comparison of the deviation and thus the verification of the statement, $\mathcal{S}'$ is replaced by the symbols created by the sampling approaches.

Tab. 1: Effect of the sampling technique on the signal quality.

	Intra-Distance	Inter-Distance	Inter-Intra-Distance
Average Symbol	0.3763 %	6.9193 %	6.5430 %
Single Symbol	0.6886 %	6.9351 %	6.2466 %
Composite Symbol	0.6886 %	6.9351 %	6.2466 %

3.3 Analysis

In Tab. 1 the calculated distances for the different sampling techniques are shown. The symbols $g = 1$, i.e. the rising edges, were used for the calculation, since these symbols contain the most important characteristics for distinguishability [KH18; KSH20]. For the average symbol it can be seen that it has the highest inter-intra-distance, mainly due to the lower intra-distance. This indicates that the use of the average symbol allows the best overall differentiation among all ECUs. However, it can also be seen that the single and composite symbols have no noticeable differences and, with an inter-intra-distance that is less than 0.3 % lower, the distinguishability is only minimally reduced.

Tab. 2: Effect of the sampling technique on the intra-frame deviation.

Data set	Average Symbol	Single Symbol	Composite Symbol
0.6425 %	0.4983 %	0.6230 %	0.6101 %

The deviations of the symbols within a single frame for the data set and the considered sampling techniques are shown in Tab. 2. Basically, it can be noticed that the data set shows the biggest and the average symbols the lowest deviations and the single and composite symbol again are close to each other and also close to the data set. All in all, the results confirm the claim that the differences due to the sampling approaches are negligible.

4 Sender Identification System Evaluation

The previous analysis is based on the signal itself respectively on the calculated distances. However, since the sender identification approaches use much more complex characteristics for classification, this chapter analyzes the effect of the different sampling methods on a real system. For this purpose, the Edge-based Sender Identification (EASI) [KSH20], which also uses only a single rising edge for identification, is considered. For the evaluation, the system uses the same configuration and data set used in the original work for the analysis of the behavior of the characteristics to environmental factors as well as the effect of electrical consumers. The utilized vehicle is the same as mentioned in Sect. 3.1, but without having the additional Raspberry Pis connected. The metrics considered for the evaluation of the approximately 55 000 frames are the *true positive* and *true negative rate*, the *identification rate* and the *confidence* of the system. A high true positive rate indicates the system's ability to detect forged frames, the true negative rate allows an assessment of the amount of wrong

alarms, the identification rate analyzes the general performance of sender identification and the confidence gives an indication on how well the learned model fits to the current situation.

Tab. 3: Effect of the sampling technique on the sender identification performance.

	Average Symbol	**Single Symbol**	**Composite Symbol**
True Positive Rate	99.82 %	99.16 %	99.59 %
True Negative Rate	100 %	100 %	100 %
Identification Rate	99.98 %	99.91 %	99.98 %
Confidence	99.81 %	99.57 %	99.87 %

In Tab. 3, it can be noticed that the use of the average symbol achieves the best results considering the real sender identification system. While no false alarms have occurred in any of the analyses, the usage of the single symbol leads to a slight decrease of the true positive rate. Assuming that an attack requires three messages which are not detected by the system, the probability of a successful attack is increased from 5.8^{-9} to 5.9^{-7} by using the single symbol instead of the average symbol. Accordingly, even by using EASI with the most lightweight configuration, the single symbol, a high probability of detecting potential attacks is achieved and thus still provides a high increase in security. Overall, as already determined during the direct signal analysis in Sect. 3.3, no significant differences can be observed for the different sampling techniques.

5 Assessment

Tab. 4: Assessment of the signal acquisition approaches.

	Performance	**Additional Hardware Requirements**	**Resource Requirements**	**Multi-Channel Capability**	**Complexity**	**Timing Restrictions**
Average Symbol	+	-	○	-	○	○
Single Symbol	-	-	+	+	+	+
Composite Symbol	○	○	+	-	-	-

An overview of the assessment is presented in Tab. 4, where the approaches are compared relatively in terms of the individual aspects. While the use of the average symbol provides the best results in the previous analyses, it is also expected to have the highest resource consumption. In particular, as with the single symbol, an external analog-to-digital converter (ADC) is required to record the entire signal or symbol at the appropriate sampling rate. For instance, a required sampling rate of 20 megasamples/second is assumed for the classic CAN, but due to the higher requirements respectively the shorter symbol duration of CAN-FD, higher sampling rates will be necessary. The acquisition of a composite symbol offers some advantages, as under certain assumptions regarding the used microcontroller it is possible to utilize the internal ADC for the acquisition. However, this requires a particularly fast comparator [KSH20] to be able to detect the individual level changes fast enough and the observed frames must have a certain amount of corresponding symbol transitions. In principle, both the acquisition of single and composite symbols require the least resources in

terms of calculation and storage, because in the case of the average symbol, it is necessary to store the entire signal in order to process it before the signal characteristics can be extracted. Provided that the computing capability of the implementing ECU is sufficient, however, this disadvantage can be compensated by calculating the running average. With the single and composite symbol this is omitted as the symbols can be used directly. Depending on the communication architecture of the vehicle under consideration, it may be necessary to analyze several CANs in parallel. For example, the networks of many vehicles are nowadays separated by domain or functions, which contributes to an increased security [RFS18]. In the case that multiple channels have to be observed, the usage of the single symbol is especially advantageous, as the sampling unit is only occupied for the time span of the symbol. For a single or a small number of bus segments, the composite symbol approach shows the lowest cost, but the complexity of time-critical sampling should not be underestimated. Reaching a high sender identification performance and a high robustness against fluctuations and signal changes potentially caused by environmental conditions and electrical consumers [KSH19], will allow to prevent attacks by disturbing ongoing transmissions of forged frames [Fo19; KH18]. If the possibility of preventing an attack is intended, the decision whether a forged frame is present must be made in a correspondingly short time. The type of signal acquisition has a considerable influence in this respect, since a certain number of symbols of the same type must have been transmitted for the generation of the composite symbol. A certain number of symbols of the same type are also used for the average symbol, whereby the amount can also be defined variably. For example, a time span can be defined after which all captured symbols are used for the calculation of the average symbol. In case only a single symbol is acquired during this time, this corresponds at least to the single symbol whose direct usage causes the least negative effect on the required processing time.

6 Conclusion

Basically, no significant differences in performance with respect to detection and identification rates could be determined by the different sampling methods. For this insight, on one hand the signals were analyzed directly and on the other hand the performance effects of the sampling method on a sender identification system were investigated. The small difference in performance enables the selection of the method based on the available resources and the underlying communication architecture. The average symbol, for example, not only shows slightly higher performance but also has the highest resource usage, while the recording of a single or composite symbol has advantages depending on the specific application. For single CAN buses the composite symbol is the most cost effective option, while the recording of a single symbol with an external ADC is advantageous for monitoring several CAN segments.

References

[Ca19] Cai, Z.; Wang, A.; Zhang, W.; Gruffke, M.; Schweppe, H.: 0-days & Mitigations: Roadways to Exploit and Secure Connected BMW Cars. Black Hat USA/, 2019.

[CA20] CAN in Automation: CAN XL is knocking at the door./, Jan. 2020, URL: https://www.can-cia.org/news/cia-in-action/view/can-xl-is-knocking-at-the-door/2020/1/3/, visited on: 01/03/2020.

[Ch18] Choi, W.; Joo, K.; Jo, H. J.; Park, M. C.; Lee, D. H.: VoltageIDS: Low-Level Communication Characteristics for Automotive Intrusion Detection System. IEEE Transactions on Information Forensics and Security 13/8, pp. 2114–2129, Aug. 2018, ISSN: 1556-6013.

[Fo19] Foruhandeh, M.; Man, Y.; Gerdes, R.; Li, M.; Chantem, T.: SIMPLE: Single-Frame Based Physical Layer Identification for Intrusion Detection and Prevention on in-Vehicle Networks. In: Proceedings of the 35th Annual Computer Security Applications Conference. ACSAC '19, Association for Computing Machinery, San Juan, Puerto Rico, pp. 229–244, 2019.

[HKD11] Hoppe, T.; Kiltz, S.; Dittmann, J.: Security threats to automotive CAN networks—Practical examples and selected short-term countermeasures. Reliability Engineering & System Safety 96/1, pp. 11–25, 2011, ISSN: 0951-8320.

[KH18] Kneib, M.; Huth, C.: Scission: Signal Characteristic-Based Sender Identification and Intrusion Detection in Automotive Networks. In: Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security. CCS '18, ACM, New York, NY, USA, pp. 787–800, 2018, ISBN: 978-1-4503-5693-0.

[Kn20] Kneib, M.: A Survey on Sender Identification Methodologies for the Controller Area Network. In (Reinhardt, D.; Langweg, H.; Witt, B. C.; Fischer, M., eds.): SICHERHEIT 2020. Gesellschaft für Informatik e.V., Bonn, pp. 91–103, 2020.

[KSH19] Kneib, M.; Schell, O.; Huth, C.: On the Robustness of Signal Characteristic-Based Sender Identification. 2019.

[KSH20] Kneib, M.; Schell, O.; Huth, C.: EASI: Edge-Based Sender Identification on Resource-Constrained Platforms for Automotive Networks. In: Proceedings of the 27th Network and Distributed System Security Symposium. 2020.

[LL18] Luo, Q.; Liu, J.: Wireless Telematics Systems in Emerging Intelligent and Connected Vehicles: Threats and Solutions. IEEE Wireless Communications 25/6, pp. 113–119, 2018.

[MV15] Miller, C.; Valasek, C.: Remote exploitation of an unaltered passenger vehicle. Black Hat USA 2015/, p. 91, 2015.

[RFS18] Ring, M.; Frkat, D.; Schmiedecker, M.: Cybersecurity Evaluation of Automotive E/E Architectures. 2. ACM Computer Science in Cars Symposium/, 2018.

[Ro12] Robert Bosch GmbH: CAN with Flexible Data-Rate Specification. 2012.

[Ro91] Robert Bosch GmbH: CAN Specification. 1991.

EAVE: Emotional Aerial Vehicle Evaluator

Marc Lieser[1], Ulrich Schwanecke[1], Jörg Berdux[1]

Abstract: Today, semi-autonomous quadrotors are already available at affordable prices and have the potential to become part of everyday life due to the variety of possible applications. To ensure that people feel safe around quadrotors and to optimize flight times, their size should be kept to a minimum which results in their appearances remaining purely functional. This reduces the possibility of adding anthropomorphic or zoomorphic features that are typically used in order to increase acceptability by conveying the robot's intent or emotion. Constrained by mechanical appearance, other non-verbal communication channels can be exploited instead, in particular robot motion. The application *EAVE* presented in this paper was developed with the idea to design and evaluate trajectories that breathe life into inanimate, mechanical quadrotors in order to improve interaction in human-robot companionships. It extends our existing quadrotor testbed *ICARUS*, which is capable of tracking arbitrary trajectories of real and simulated quadrotors that were designed using *EAVE*. We demonstrate that applying some of the established principles of character animation to the design of quadrotor trajectories opens up the possibility of conveying intent and improving interaction, though the appearance of the quadrotor remains purely functional.

Keywords: human-quadrotor interaction; emotional aerial vehicles; non-verbal communication; motion anticipation; quadrotor companion; principles of animation; quadrotor testbed; social robots.

1 Introduction

Quadrotors experienced a huge gain in popularity over the past twenty years. Thanks to their mechanical simplicity and affordability, they have found their way from the hobby sector via research into everyday applications. Most of them take advantage of their elevated view, whether for locating victims in disaster scenarios, for inspection of buildings or for filming movies. Private consumers mainly utilize quadrotors as self-flying cameras that accompany them during various leisure activities. Today, semi or even fully autonomous quadrotors are available and follow humans while avoiding obstacles and receiving commands via smartphones or gestures. This also enables them to be helpful companions in home environments, where in future scenarios they could accompany people through daily life as smartphones already do today. Whatever the exact purpose, the number of robot assistants — be it ground-based or of aerial nature — will most certainly increase and require research on possible communication channels for human-quadrotor interaction.

Even with today's state of technology and miniaturization, it is important and challenging to keep the size of quadrotors used in interaction scenarios to a minimum. Larger rotors

[1] Department of Computer Science, RheinMain University of Applied Sciences, Wiesbaden, Germany, {marc.lieser,ulrich.schwanecke,joerg.berdux}@hs-rm.de

doi:10.18420/inf2020_81

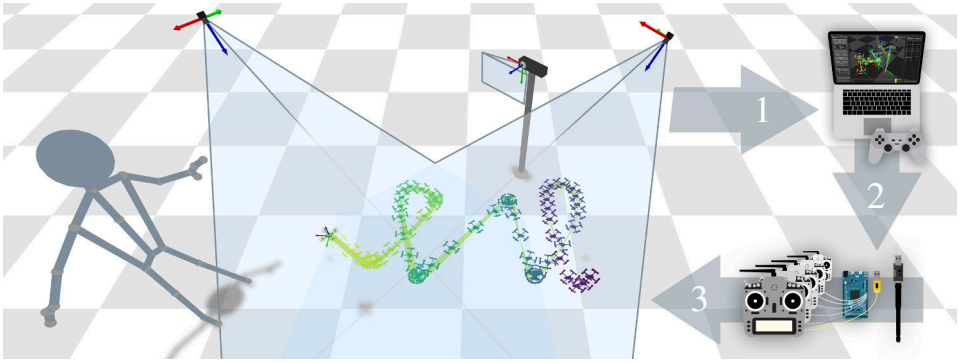


Fig. 1: Overview of our quadrotor testbed *ICARUS*, an affordable, portable indoor testbed: After optical pose estimation of the individual quadrotors (1), control variables are determined (2) and sent via different radio control systems (3) to the quadrotors. The depicted schematic shows a quadrotor in a human-quadrotor interaction scenario following a trajectory designed with *EAVE*.

of larger platforms, the generated noise and downwash are more likely to be perceived by people as disturbing [Ye17]. This introduces an inhibition threshold which is not desirable in companionships. For this reason, the quadrotors used in our scenarios are lightweight and cannot carry sensors or other electronics beyond ensuring their flight capability.

Introducing anthropomorphic features is often used in order to make the robot's inner state more identifiable for humans. This may be hard or even impossible to implement without adding to the already critical payload of quadrotors. Mediating motion intent without social cues such as gestures or gaze can be achieved by Augmented Reality (AR) applications [Wa18], but requires additional hardware. Instead of adding extra hardware, other channels of non-verbal communication can be further explored, in particular robot motion [HJ14], which is capable of transporting information beyond motion intent, namely the robot's emotions. Hence, for quadrotors that have no moving parts other than rotors, the key to convey intent lies in their trajectories and their augmentation with motions familiar to humans. In order to bring the purely mechanical remaining robots to life and to increase acceptance, we apply some of Disney's well established principles of character animation [TJ81] that have already been adapted for robots other than quadrotors [RP12, GT11]. The desire to animate quadrotors exists [Ka17], but published research remains on a conceptual level [De18] or lacks detailed implementation and parameterization [Ca16].

In this paper we present the application *EAVE* (Emotional Aerial Vehicle Evaluator) that allows to design and evaluate quadrotor trajectories. Following animation, trajectories are defined by keyframes, each consisting of position, velocity, acceleration and heading angle. Some principles of animation can be applied and parameterized in order to increase the expressiveness of the quadrotor's motion with the objective to enhance user interaction by enabling trajectories to communicate the quadrotor's emotion or intent. Based on the

defined keyframes, smooth trajectories are generated and tracked by quadrotors within our testbed *ICARUS* that allows for safe test flights of arbitrary quadrotor models in a simulated environment but also controls real quadrotors.

In the next section we give a short overview of the current state of our quadrotor testbed and point out where *EAVE* is extending the infrastructure. In Section 3 we describe the trajectory generation and provide some examples of applied principles of animation. Finally, we summarize our work in Section 4 and give an outlook on future research.

2 Testbed

This section describes the current state of our quadrotor testbed *ICARUS*, that is extended by *EAVE* through software. We outline the core components of the system and give an insight of the libraries used. A detailed description of *ICARUS* can be found in [Li17]. This low-cost quadrotor testbed was continuously expanded and improved over the past years. A schematic of the testbed is given in Fig. 1. The quadrotors used in our experiments and the overall data flow of *ICARUS* and *EAVE* are shown in Fig. 2.

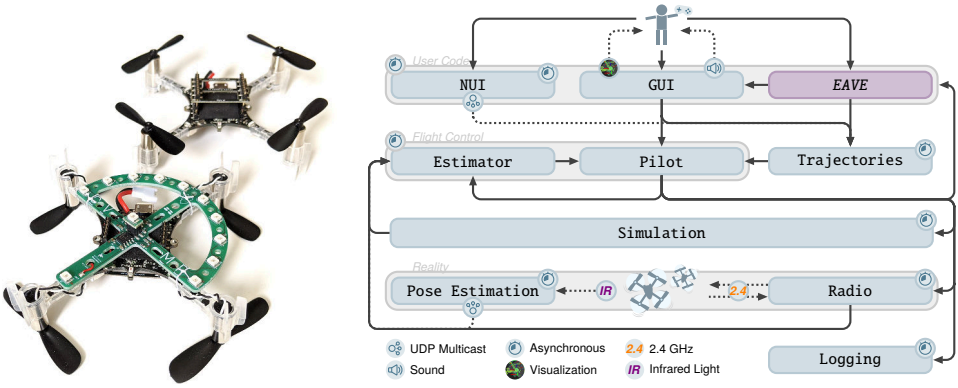


Fig. 2: Bitcraze Crazyflie 2.0 with the rotors mounted upside down so they do not obscure the LEDs of the attached tracking marker and an unmodified platform for comparison in the background (left). Data flow of *ICARUS* with *EAVE* as a component that extends the testbed by custom user code (right).

In general, all testbed software components are developed in `C++` using the `Boost` libraries for multithreading and networking. Serialization of data structures for UDP messages is implemented with `Protobuf`. The *User Code* layer consists of a classical Graphical User Interface (GUI) and an `OpenGL` visualization implemented with `Qt` and `glm`; a screenshot is shown in Fig. 3. Also, manual control using a joystick (implemented with `SDL`) and a gesture-based Natural User Interface (NUI) using a Microsoft Kinect 2 with the `Kinect 2 SDK` are implemented in the top layer. At this point, the existing infrastructure can also be extended by custom user code like *EAVE*. Next to visual feedback, the user interface also generates synthetic rotor sound. For this purpose, the sound of a single rotor was recorded

during hover and is played back spatially during the simulation using SFML. The sound of each rotor is pitched by the ratio of rotor revolutions to the revolutions required for hovering. Upper and lower limits were chosen by experiment. Below hover revolutions, the playback volume fades in linearly and remains constant above.

The implementation of the *Flight Control* layer uses a predictor–corrector estimator similar to the one described in [Lu14]. It filters noise from pose estimation and latency-compensates the quadrotors' states. Data from the estimator is used by the pilot implementation to hover in position or to track trajectories defined by the user. The control loop runs at a frequency of 100 Hz. For trajectory control a Model-Predictive Controller (MPC) with a discretization time of 0.1 s and a time horizon of 2 s is used. The implementation is based on the open-source code from [Fa18] who is using ACADO to set up the optimization problem and qpOASES to solve it. The hover controller was presented in [MK11]. Control variables are sent to the radio or simulation and back to the estimator for the next control step. A configurable range of data is also sent to an asynchronous logging implementation. Our *Simulation* is based on the quadrotor dynamics summarized in [Mi10, Lu14] and is implemented using Eigen.

For off-the-shelf (brushless) quadrotors in the *Reality* layer we use the Arduino-based serial remote control library serialrc[2] [Li17] in conjunction with a FrSky Taranis X9D radio system. Less effort but shorter flight times come with the brushed Bitcraze Crazyflie 2.0, which we integrated using the radio library that is part of [HA17]. The in-house developed OpenCV-based optical pose estimation is described in [Tj19]. The markers utilized in our system were optimized for the use with quadrotors and can be seen attached to a Crazyflie in Fig. 2. Pose estimation runs on a dedicated machine and multicasts UDP messages that consist of a six degrees of freedom pose and a timestamp. The estimator receives position and attitude from pose estimation and determines velocity and angular velocities by numerical differentiation. Telemetry data is also forwarded to the estimator and includes the current battery voltage to compensate the thrust command for varying voltages.

3 Trajectories

For most robots, it is completely sufficient to determine a trajectory on the shortest path between two points. This results in calculated straight or plain "mechanical" trajectories that render the identification of the robots intent or emotion impossible in a human-robot companionship. In our scenario, in which the appearance of the quadrotors is kept mechanical, their trajectories should distract from that design and be able to convey certain feelings to revalue quadrotors for humans in human-robot interaction. This section describes the representation of trajectories and our design approach to applying some of the basic principles of animation, with the goal of making quadrotors more alive. A screenshot of the user interface including the trajectory editor can be seen in Fig. 3.

[2] https://github.com/cvmrgroup/serialrc

3.1 Trajectory Representation

We describe trajectories as piecewise polynomial functions (see for example [Co13]). Quintic (fifth-order) polynomials are a common choice for the representation of trajectories, since their first and second temporal derivatives — velocity and acceleration — are continuous and thus result in smooth trajectories. Furthermore boundary conditions, such as position, velocity, acceleration and time, can easily be set. A scalar trajectory quintic polynomial together with its first and second order derivates, given as

$$\begin{aligned} p(t) &= at^5 + bt^4 + ct^3 + dt^2 + et + f \\ \dot{p}(t) &= 5\,at^4 + 4\,bt^3 + 3\,ct^2 + 2\,dt + e \\ \ddot{p}(t) &= 20\,at^3 + 12\,bt^2 + 6\,ct + 2\,d \end{aligned}$$

result in a linear system

$$\begin{pmatrix} p(0) \\ \dot{p}(0) \\ \ddot{p}(0) \\ p(T) \\ \dot{p}(T) \\ \ddot{p}(T) \end{pmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 2 & 0 & 0 \\ T^5 & T^4 & T^3 & T^2 & T^1 & 1 \\ 5T^4 & 4T^3 & 3T^2 & 2T^1 & 1 & 0 \\ 20T^3 & 12T^2 & 6T^1 & 2 & 0 & 0 \end{bmatrix} \begin{pmatrix} a \\ b \\ c \\ d \\ e \\ f \end{pmatrix} \tag{1}$$

with start boundary time $t = 0$ and end boundary time $t = T$, where T is the duration of the trajectory segment. For $T \neq 0$ the linear system (1) can be uniquely solved for the polynomial coefficient vector $\mathbf{c} = (a, b, c, d, e, f)^\top$. A trajectory $\mathcal{K}$ is represented by a list of $N + 1$ boundary conditions $\mathbf{p}_i(t_i)$ at prescribed time t_i, i.e. $\mathcal{K} = (\mathbf{p}_0(t_0), \ldots, \mathbf{p}_N(t_N))$. Scalar trajectories can be extended to the multi-dimensional vector case in a straightforward way. Three-dimensional vectors of position, velocity and acceleration are pooled with time and quadrotor heading into keyframes, a term borrowed from animation. When a trajectory is finally sampled for the controller, the piecewise polynomials are interpolated independently in each dimension and joined to obtain the final trajectory. The heading angle is interpolated linearly. As sampling rate we use 100 Hz according to the update rate of the controller used.

3.2 Trajectory Design

It is difficult, if not impossible, to predict the intent of a mechanically designed quadrotor or of robots in general. Disney's twelve basic principles of character animation [TJ81] have proven beneficial in robot motion design [HJ14] but have not yet been applied to quadrotors. These principles are the result of several decades of work of animators and serve the main purpose of creating the illusion that cartoon characters adhere to the basic laws of physics, but they also deal with more abstract topics such as emotional timing and character appeal.

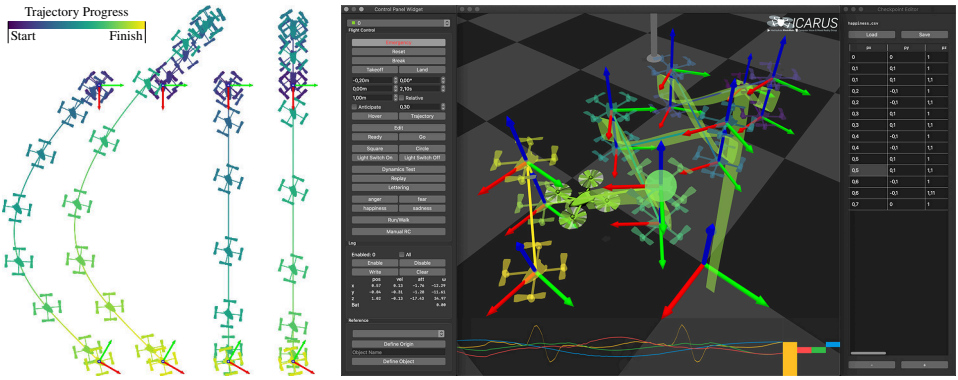


Fig. 3: Two pairs of trajectory segments, with and without anticipation (velocity and heading), which differ by the velocity at the target keyframe (left). *EAVE*'s GUI with the control panel on the left side, trajectory editor on the right side and visualization in the middle (right).

Since quadrotors are already subject to the laws of physics and have no moving parts other than rotors, not all principles are applicable, or at least not in their entirety. On the other hand, some principles can easily be applied by just setting the according boundary conditions of the linear system (1). The *Slow In and Slow Out* principle creates realistic acceleration of objects and can be adjusted by reducing the acceleration at keyframes. Velocity information generates arched trajectories and thus more natural motion, as described by the *Arcs* principle. On the contrary, a mechanical robot motion would be the direct path between two points. *Timing* can also be modified with a boundary condition and is probably the most powerful tool to change emotions. Velocity and acceleration between keyframes can determine whether a character is lethargic, excited or nervous. According to [TJ81], personalities are defined more by their movements than by their appearance, which substantiates our idea of distracting from a required mechanical design by natural motion. A *Secondary Action* is often used to emphasize the main action and thus give more life to a scene. Applied to quadrotor motion, it could yaw its heading to the next keyframe to give a more life-like impression. *Exaggeration* presents character features in a more extreme form in order to overcome the static appearance of drawings that perfectly imitate reality. A quadrotor could perform a tornado-like yaw spin right before a faster movement or through the whole trajectory. *Anticipation* helps the audience preparing for actions and to make them appear more natural. This is done by special moves that anticipate the upcoming action, just as a bow draw anticipates the arrow shot.

In order to avoid that users are surprised or frightened by sudden movements of the quadrotor and thus possibly intimidated in their interaction, we applied the principle of *Anticipation* along with the principles of *Slow In and Slow Out* and *Arcs*. The result is shown in Fig. 3, where the motion of the quadrotor is anticipated by extrapolating the first piecewise polynomial by a user-defined time coefficient. Thus a counter movement is created right before the start of the actual motion. Extrapolating by large numbers is not predictable, but

low coefficients achieve results well suited to the intend of anticipation. Using polynomial extrapolation over linear extrapolation does not only transport information about the next keyframe's position but also adds velocity and acceleration information. Beyond that, a counter rotation is added by extrapolating the heading angle. Evaluating improvement of interaction quality by allowing the user to anticipate the quadrotor's next movement as well as adding further animation principles will be part of our future work.

Rotor sound plays an important but subordinate role and should be kept in mind during the design process. Parameter tuning of the MPC influences smoothness, aggressiveness and accurateness of how a trajectory is tracked. With certain controller settings, it is even possible to fly a smooth trajectory with irritating, unsteady rotor noise, which can be unpleasant for humans. In order to evaluate this during the design stage, we have introduced synthetic sound as part of the user interface as described in Section 2.

3.3 Evaluation

While evaluation of trajectory controllers address the robots's spatial and temporal deviation from the planned trajectory, there are several options for the evaluation of trajectories in terms of quality of user experience and quadrotor acceptance. User tests can be carried out directly in the testbed using real hardware or remotely using videos. Testing users on site is more elaborate to conduct, but gives the users a more realistic experience than online surveys. Online surveys on the other hand are able to reach a much larger test group. A test video can be filmed from the user's or a third person's perspective or rendered from the visualization. Videos can also be extended by AR to gain immersion for more elaborate tests, where the use of real hardware would be too dangerous, costly or complex. How well the trajectory design described in Section 3.2 is understood and helps to improve interaction could be evaluated by rendering all the trajectory information of our 3D visualization during a flight in the testbed on the screen of an AR device, such as a head-mounted display.

4 Conclusion & Future Work

In this paper we presented *EAVE*, a software extension to design and evaluate trajectories along with an updated description of our quadrotor testbed *ICARUS* that is used to control real and simulated quadrotors. We provided an updated description of the core components of said testbed. *EAVE* was implemented with the aim to design trajectories that are able to convey intent and thus improve quality of human-quadrotor interaction. As an example, we have applied an animation technique called *Anticipation* to a given trajectory. Future work will include user tests on how well the interaction experience can be improved by adding anticipation along with further principles of animation to quadrotor trajectories. Also, we are currently working on a dynamic, situation-dependent human-quadrotor interaction scenario where the quadrotor's response is triggered by user behavior.

Bibliography

[Ca16] Cauchard, J. R.; Zhai, K. Y.; Spadafora, M.; Landay, J. A.: Emotion encoding in Human-Drone Interaction. In: HRI 2016. pp. 263–270, March 2016.

[Co13] Corke, Peter: Robotics, Vision and Control: Fundamental Algorithms in MATLAB. Springer Publishing Company, Incorporated, 1st edition, 2013.

[De18] Deng, Honghao; Li, Jiabao; Sayegh, Allen; Birolini, Sebastian; Andreani, Stefano: Twinkle: A Flying Lighting Companion for Urban Safety. In: TEI '18. pp. 567–573, 2018.

[Fa18] Falanga, D.; Foehn, P.; Lu, P.; Scaramuzza, D.: PAMPC: Perception-Aware Model Predictive Control for Quadrotors. In: IROS '18. pp. 1–8, Oct 2018.

[GT11] Gielniak, M. J.; Thomaz, A. L.: Generating anticipation in robot motion. In: 2011 RO-MAN. pp. 449–454, 2011.

[HA17] Hönig, Wolfgang; Ayanian, Nora: Flying Multiple UAVs Using ROS. In: Robot Operating System (ROS): The Complete Reference. Springer, pp. 83–118, 2017.

[HJ14] Hoffman, Guy; Ju, Wendy: Designing Robots with Movement in Mind. J. Hum.-Robot Interact., 3(1):91–122, February 2014.

[Ka17] Karjalainen, Kari Daniel; Romell, Anna Elisabeth Sofia; Ratsamee, Photchara; Yantac, Asim Evren; Fjeld, Morten; Obaid, Mohammad: Social Drone Companion for the Home Environment: A User-Centric Exploration. In: Proceedings of the 5th International Conference on Human Agent Interaction. HAI '17, p. 89–96, 2017.

[Li17] Lieser, M.; Tjaden, H.; Brylka, R.; Löffler, L.; Schwanecke, U.: A low-cost mobile infrastructure for compact aerial robots under supervision. In: 2017 European Conference on Mobile Robots (ECMR). pp. 1–6, Sept 2017.

[Lu14] Lupashin, Sergei; Hehn, Markus; Mueller, Mark W; Schoellig, Angela P; Sherback, Michael; D'Andrea, Raffaello: A platform for aerial robotics research and demonstration: The Flying Machine Arena. Mechatronics, 24:41–54, 2014.

[Mi10] Michael, Nathan; Mellinger, D.; Lindsey, Q.; Kumar, V.: The GRASP Multiple Micro-UAV Testbed. Robotics Automation Magazine, IEEE, 17(3):56–65, Sept 2010.

[MK11] Mellinger, D.; Kumar, V.: Minimum snap trajectory generation and control for quadrotors. In: ICRA 2011. pp. 2520–2525, May 2011.

[RP12] Ribeiro, T.; Paiva, A.: The illusion of robotic life: Principles and practices of animation for robots. In: HRI '12. pp. 383–390, 2012.

[TJ81] Thomas, F.; Johnston, O.: The Illusion of Life: Disney Animation. Disney Editions, 1981.

[Tj19] Tjaden, Henning: Robust Monocular Pose Estimation of Rigid 3D Objects in Real-Time. PhD thesis, Johannes Gutenberg University Mainz, Germany, 2019. https://nbn-resolving.org/urn:nbn:de:hebis:77-diss-1000025478, last accessed 06 May 2020.

[Wa18] Walker, Michael; Hedayati, Hooman; Lee, Jennifer; Szafir, Daniel: Communicating Robot Motion Intent with Augmented Reality. In: HRI '18. Association for Computing Machinery, New York, NY, USA, p. 316–324, 2018.

[Ye17] Yeh, Alexander; Ratsamee, Photchara; Kiyokawa, Kiyoshi; Uranishi, Yuki; Mashita, Tomohiro; Takemura, Haruo; Fjeld, Morten; Obaid, Mohammad: Exploring Proxemics for Human-Drone Interaction. In: HAI '17. ACM, New York, NY, USA, pp. 81–88, 2017.

R. Reussner, A. Koziolek, R. Heinrich (Hrsg.): INFORMATIK 2020,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2021 907

Citcom – Citation Recommendation

Melina Meyer[1] Jenny Frey[1] Tamino Laub[1] Marco Wrzalik[2] Prof. Dr. Dirk Krechel[2]

Abstract: Citation recommendation aims to predict references based on a given text. In this paper, we focus on predicting references using small passages instead of a whole document. Besides using a search engine as baseline, we introduce two further more advanced approaches that are based on neural networks. The first one aims to learn an alignment between a passage encoder and reference embeddings while using a feature engineering approach including a simple feed forward network. The second model takes advantage of BERT, a state-of-the-art language representation model, to generate context-sensitive passage embeddings. The predictions of the second model are based on inter-passage similarities between the given text and indexed sentences, each associated with a set of references. For training and evaluation of our models, we prepare a large dataset consisting of English papers from various scientific disciplines.

Keywords: citation recommendation; natural language processing; representation learning

1 Introduction

Writing scientific papers or any kind of work that requires a lot of citing can be very exhausting. A lot of research is necessary to find documents that are close to a specific topic and worth citing to prove a point. There already exist some tools to propose related work. However, there is potential for improvements in these tools and thus for research. The task focusing on this problem is called citation recommendation. It can be be divided into two categories: global approaches and context-based recommendation. While approaches of the first category pay attention on the whole document, context-aware attempts focus on sentences before and after a citation. After considering various options, three different context-related approaches to this task are pursued here, including the baseline. The first is to create a simple proof of concept application in order to get a benchmark for recommendation performance. The other models are based on neural networks. As a baseline experiment, we use Elasticsearch [3], a full-text search engine that can find documents due to their structural similarity. The second model uses feature engineering and a feed forward network. Additionally, we wanted to use a tool that learns characteristics of the language used in the documents and takes semantics into consideration, which brought us to BERT[De19] and is used in the third model. This model learns passage similarity using BERT embeddings to

[1] RheinMain University of Applied Sciences, Faculty of Design Computer Science Media, Unter den Eichen 5, 65195 Wiesbaden, Germany, <forename>.<surname>@student.hs-rm.de
[2] RheinMain University of Applied Sciences, Working Group LAVIS – Learning and Visual Systems, Unter den Eichen 5, 65195 Wiesbaden, Germany, lavis@hs-rm.de
[3] See https://www.elastic.co/

doi:10.18420/inf2020_82

predict references. For evaluation, we create a dataset which is based on a dump of arXiv documents of all scientific disciplines from recent years. In summary, our contribution is to prepare a comprehensive dataset suitable for citation recommendation, a baseline for comparison purposes and to develop two more advanced models.

2 Related Work

Citation recommendation is a complex task that has gained increasing attention in recent years. Besides some methods for proposing references for entire documents [CHY18; Kü12; Re14], many new publications refer to the context-based approach. Local recommendation, as the term was coined by the authors of [He10], uses contextual information. Based on this idea, [Hu12] proposed an approach via a machine translation system transferring keywords of the context into cited documents. They continued this work in [Hu15] by adding semantic embeddings of the words of the context as well as cited documents. Finally, a recommendation is carried out based on the semantic distance in vector space. [TWZ14] presented the first embedding-based approach for context-aware citation recommendation, which uses TF-IDF vectors to form cross-language embeddings and uses them for the proposals. Approaches without neural networks based on information retrieval techniques and metrics such as TF-IDF or BM25 also have been investigated. In their proposals, [Du16] annotate each sentence of the given documents with CoreSC classes, which are indexed in Lucene and used to determine similarity. On the other hand, [EF17] use BM25 as a baseline for their encoder-decoder framework, inspired by neural machine translation, which learns relations between text pairs of variable length. The approach is further supplemented by additionally analyzing the writing style of authors. Other procedures also include document metadata. [FS20] is a semi-genetic hybrid recommender system for citation recommendation. The authors combine embedding and information retrieval approaches using a fitness score to receive the top k recommendations. Recent approaches additionally include language models such as BERT. In [Je19], BERT is used as a context encoder for textual embeddings, supplemented by a Graph Convolutional Network (GCN) for metadata and reference relationship between papers as a citation encoder for the construction of graph embeddings. The concatenation of the output vectors is then used as input of a feed forward network.

3 Dataset

In order to recommend citations for scientific papers and to evaluate our model on a large dataset, we reuse the record provided in [SF20]. The dataset is based on an arXiv source dump including papers from 1991 to 2018 and of all scientific disciplines available on arXiv.org. It offers 29.2 million ready-made citation contexts, each consisting of three sentences: the sentence containing the citation and the two surrounding ones. We use these contexts to find the corresponding sections in the full text files, extract all references of the section and map them to the arXiv- and/or MAG-ID of the cited paper. We only consider

references that could be matched to arXiv- or MAG identifiers as only these can be verified. Furthermore, we use a dump of arXiv metadata[4] to obtain basic background information about the cited papers. Since MAG metadata are difficult to obtain for research purposes, we limited our experiment to passages that reference known arXiv papers as we need the metadata for one of our approaches. Finally, training and test data are divided in the ratio 80:20. Since our models require a lot of computing power, we create another dataset to reduce the computing time. Since we want to use the scientific categories as additional meta information, we cluster the passages into 176 clusters based on those categories. From each cluster, five percent of the references are randomly selected and citing passages are part of the reduced dataset. This dataset is also randomly split in the ratio of 80:20. Table 1 illustrates the details of our datasets containing the total number of disjoint references, training passages, disjoint references contained by the training passages and test passages.

Dataset	#Refs	#Train Pass.	#Train Refs	#Test Pass.
full	718,329	14,932,722	700,220	3,733,181
reduced	115,643	1,339,920	115,643	329,130

Tab. 1: An overview of the employed datasets

4 Model Architectures

4.1 Baseline Model

In order to get an initial orientation, a baseline experiment is first performed using Elasticsearch. Elasticsearch is a search engine based on the Lucene program library for full text searches. First, several documents are added to indices so that similar documents can be found by a query using a RESTful API. For both of the datasets, a new index is created. Due to the large amount of training data, the passages are grouped according to their content. This means that passages with the same context but different target references are identified and indexed only once. Each indexed document contains a text field with `english` language analyzer[5] as well as two keyword fields for lists of all associated passage IDs and target references. This reduces the amount of training data to a total of 8,700,191 indexed documents for the full dataset and 1,199,747 for the reduced one. Especially for the full dataset, this has a significant impact on the duration of the subsequent evaluation.

4.2 Feed Forward Model with Feature Engineering

In the first model, a feature-based approach is used to achieve better results in the prediction of citations. This model consists of two sub models and the architecture is presented in

[4] Downloaded from https://archive.org/details/arXiv-metadata-dump-2019-06-18.tar.xz
[5] See https://www.elastic.co/guide/en/elasticsearch/reference/current/analysis-lang-analyzer.html#english-analyzer

Figure 1. The features of each passage that we want to use for our embedding are TF-IDF scores and the passage length. To have equally sized feature vectors for each of the passages, only the scores of the 100,000 most common words across all passages, sorted by the occurrences in the corpus, are used. So, each passage is represented as a vector of 100,001 dimensions, where 100,000 dimensions made up of the TF-IDF values of the words. The remaining dimension represents the passage length, which is calculated by dividing the length of the current passage by the average passage length of our corpus.

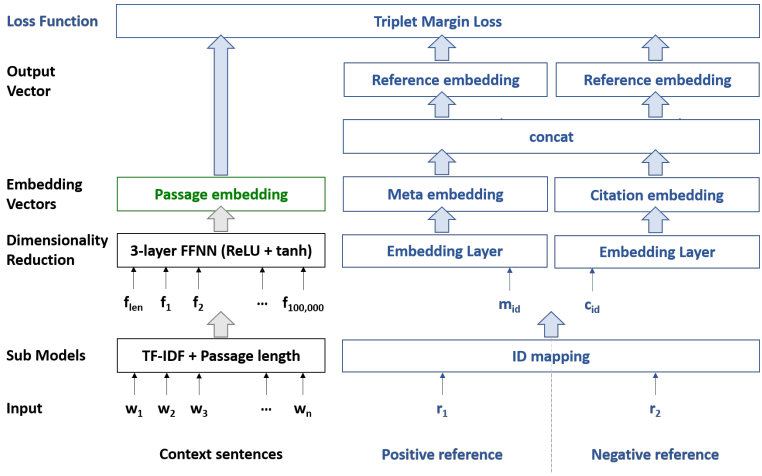


Fig. 1: Architecture of our feature based approach

We chose a three-layer Feed Forward Neural Network (FFNN) to convert the vectors down to 300 dimensions. The network consists of three linear layers, the first two are provided with a ReLU and *tanh* is utilized for the third layer. Since we use a Triplet Margin loss for training, two references per passage are needed to improve network performance. For the first, we use the correct reference featured in the original text passage and the second one is a randomly chosen reference from the corpus. The required embeddings of the references are obtained by the second sub model. The second sub model is used to encode the references. In addition, the meta-information is taken into account to improve the prediction of suitable references. As meta information, the scientific category of the cited paper is considered. For this reason, each embedding of a reference consists of a meta embedding and a citation embedding. The meta embedding refers to a feature that is shared by several references, they are grouped by this feature. The citation embedding refers to the reference itself. The reference data is used as input, which is filtered according to the required features. Based on this, identifier for metadata and the citation are determined and converted into vectors via embedding layers. Both embedding types possess their own lookup table and are concatenated for each reference. The meta embedding has a dimension of 100 and the citation embedding's dimension is 200, leading to an output embedding with a dimension of 300.

4.3 BERT Passage Model

For our third model, we use a pre-trained english BERT model [Wo19] to encode a passage. BERT (Bidirectional Encoder Representations from Transformers) is a language representation model that achieves state of the art results on different natural language processing tasks. Training a BERT model consists of two steps: pre-training and fine-tuning. In addition, BERT uses some special tokens, these are, among others, the [CLS] token and the [MASK] token. In the first step, we use a BERT tokenizer [Wo19] to convert the passage into an id sequence. As [MASK] tokens can be used to hide special tokens, we replace the references in the passages by this. Additionally, the encoded sequence has the classification token [CLS] at the beginning. Like all tokens of the sequence, this token is transformed into an embedding and corresponds to the sequence representation for different classification tasks. The converted sequence is fed into BERT to receive the word embeddings and the classification embedding of the [CLS] token of the passage. We further only use the classification embedding which is used for fine-tuning to predict references. Our first idea was to reuse the reference sub model described in the previous section. The aim was to train these sub models again by Triplet Margin Loss and fine-tune BERT, so that the encoded [CLS] embedding refers to the positive reference embedding. As this model did not perform well on a first trial dataset, we decide to use other passages for training instead of the references. The aim is to predict references using similar passages as illustrated in Figure 2.

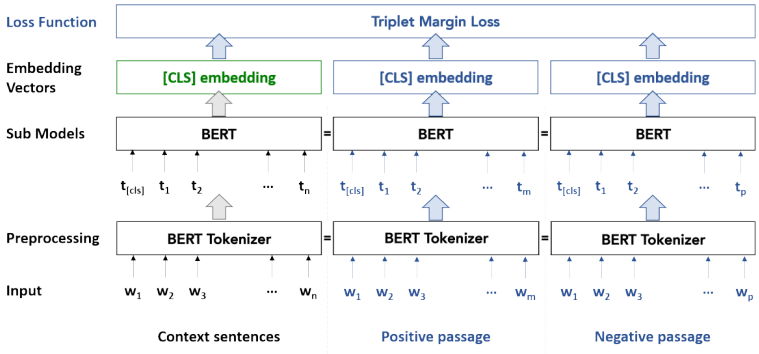


Fig. 2: Architecture of our passage-based BERT approach

For this reason, a data sample consists of a passage for which we want to predict the reference (relevant passage), a passage with the same reference (positive passage), and a passage with another reference (negative passage). As the sample requires two passages with the same reference, we only use passages if at least one other passage with the same reference exists. All passages are fed into BERT to receive the [CLS] embeddings. BERT is fine-tuned so that it learns a similarity between the relevant and the positive passage. To predict the reference of the relevant passage, the cosine similarity between the classification embedding of this passage and all other passages is calculated. The final reference is determined by the known reference of the passage that is most similar to the relevant passage.

5 Training and Evaluation

Most of citation recommendation tasks use well-known metrics such as Mean Average Precision (MAP), Recall, Mean Reciprocal Rank (MRR), Normalized Discounted Cumulative Gain (NDCG) and hits@k for evaluation. We followed these approaches and decided to use MRR@10, hits@10 and hits@5.

5.1 Baselines

As there are many training passages, the training data was grouped by contexts and inserted one single time for each content yielding corresponding passage IDs and target references. As search query, a *More Like This Query*[6] is formed which promises a better timing performance than usual queries for large indices. The query is mapped to the content field and gets the context of a test passage as *like* statement. In addition, due to the brevity of the context, the minimum term frequency is reduced to 1 and the minimum document frequency to 3. The maximum number of query terms is decreased to 10 to further reduce response times. Using different thread workers, each query is posted separately also returning the top ten results. For each result, the proposed target references are collected and afterwards sorted by their frequency of occurrence. Finally, the topmost ten references are extracted and searched for the ground truth reference, which serves as the relevant item for calculation of MRR, hits@10 and hits@5. Table 2 illustrates the evaluation results for the reduced dataset. As the table shows, the baseline already provides respectable results.

Model	MRR@10	Hits@10	Hits@5
Elasticsearch	0.546393	0.793804	0.758596
Feature FFNN	0.000042	0.000125	0.000073
BERT Passage	0.582673	0.763704	0.713639

Tab. 2: Evaluation of the models for the reduced dataset

5.2 Feed Forward Network with TF-IDF

For training the feature embedding network we use a GeForce GTX 1080 with a batch size of 8. We train the model for three epochs using the Adam optimizer for both sub models with a learning rate of 1e−3 on the reduced dataset. We also evaluate this model by MRR@10, hits@10 and hits@5. The results of this approach are shown in Table 2. It is easy to see that the scores achieved by this approach are way lower than the other numbers. A reason for this could be that the model only takes limited context into account when looking for similar passages, mostly focusing on the reference itself. Additionally, these extremely low results may caused by the drastic reduction of the dataset due to computation time which results in having only a few passage examples per unique reference.

[6] See https://www.elastic.co/guide/en/elasticsearch/reference/current/query-dsl-mlt-query.html

5.3 BERT Passage Model

For fine-tuning BERT we use AdamW [LH17] optimizer with a learning rate of 1e−6. AdamW is an optimizer with decoupled weight decay is suitable for fine-tuning BERT. Due to the amount of data, we train the model on four GPUs, three GeForce RTX 2080 Ti and a GeForce GTX 1080 Ti, in parallel with a total batch size of 10 for 3 epochs on the reduced dataset. Table 2 shows the evaluation using Mean Reciprocal Rank, hits@10, and hits@5. While hits@10, and hits@5 is lower than the baseline, MRR@10 provides the best results on the dataset. When training a further epoch, it has been shown that the results of the metrics change only at the third decimal place, which shows that the model converges stably.

6 Conclusion

Recommending citations for a given query is a complex task that will keep researchers engaged for a long time. In our paper we proposed some context-based attempts that presented us with many challenges. These models are based on different approaches, which all have shown their advantages and disadvantages. We presented them in the context of our work and attempted to improve them continuously. The baseline evaluation was particularly striking, since even a simple term-based similarity already yields respectable results. Based on this, we have tried to improve these results with our models, which are based on different approaches, and to compare the results. The results of the presented models provided some surprises, especially the model using TF-IDF. The model based on learning citation embedding and meta embedding gives worse results in comparison to the Elasticsearch baseline. As mentioned in the evaluation, the worse results can be caused by different reasons and should be examined in detail in some future work. On the contrary, reference prediction by learning passage similarities using BERT embeddings finally gives the best results on MRR but provides poorer results on hits@10 and hits@5 than the baseline calculation. In general, the models using passage similarity for reference prediction provide the best results for our use case and a well-functioning citation recommendation solution.

References

[CHY18] Cai, X.; Han, J.; Yang, L.: Generative Adversarial Network Based Heterogeneous Bibliographic Network Representation for Personalized Citation Recommendation. In: Proc. 32nd AAAI Conf. on Artificial Intelligence. AAAI Press, pp. 5747–5754, 2018.

[De19] Devlin, J.; Chang, M.-W.; Lee, K.; Toutanova, K.: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In: NAACL-HLT. 2019.

[Du16] Duma, D.; Liakata, M.; Clare, A.; Ravenscroft, J.; Klein, E.: Rhetorical Classification of Anchor Text for Citation Recommendation. D-Lib Magazine 22/, Sept. 2016.

[EF17] Ebesu, T.; Fang, Y.: Neural Citation Network for Context-aware Citation Recommendation. In: Proc. 40th Int. ACM SIGIR Conf. on Research & Development in Information Retrieval. Pp. 1093–1096, 2017.

[FS20] Färber, M.; Sampath, A.: HybridCite: A Hybrid Model for Context-Aware Citation Recommendation. arXiv preprint arXiv:2002.06406/, 2020.

[He10] He, Q.; Pei, J.; Kifer, D.; Mitra, P.; Giles, L.: Context-aware Citation Recommendation. In: Proc. 19th Int. Conf. WWW. Pp. 421–430, 2010.

[Hu12] Huang, W.; Kataria, S.; Caragea, C.; Mitra, P.; Giles, C. L.; Rokach, L.: Recommending Citations: Translating Papers into References. In: Proc. 21st ACM Int. Conf. on Information & Knowledge Management. Pp. 1910–1914, 2012.

[Hu15] Huang, W.; Wu, Z.; Liang, C.; Mitra, P.; Giles, C. L.: A Neural Probabilistic Model for Context based Citation Recommendation. In: 29th AAAI Conf. on Artificial Intelligence. 2015.

[Je19] Jeong, C.; Jang, S.; Shin, H.; Park, E.; Choi, S.: A Context-Aware Citation Recommendation Model with BERT and Graph Convolutional Networks. arXiv preprint arXiv:1903.06464/, 2019.

[Kü12] Küçüktunç, O.; Kaya, K.; Saule, E.; Çatalyürek, Ü. V.: Fast Recommendation on Bibliographic Networks. In: 2012 IEEE/ACM Int. Conf. on Advances in Social Networks Analysis and Mining. Pp. 480–487, 2012.

[LH17] Loshchilov, I.; Hutter, F.: Decoupled Weight Decay Regularization. arXiv preprint arXiv:1711.05101/, 2017.

[Re14] Ren, X.; Liu, J.; Yu, X.; Khandelwal, U.; Gu, Q.; Wang, L.; Han, J.: ClusCite: Effective Citation Recommendation by Information Network-based Clustering./, Aug. 2014.

[SF20] Saier, T.; Färber, M.: unarXive: A Large Scholarly Data Set with Publications' Full-text, Annotated in-text Citations, and Links to Metadata. Scientometrics/, pp. 1–24, 2020.

[TWZ14] Tang, X.; Wan, X.; Zhang, X.: Cross-language Context-aware Citation Recommendation in Scientific Articles. In: Proc. 37th Int. ACM SIGIR Conf. on Research & Development in Information Retrieval. Pp. 817–826, 2014.

[Wo19] Wolf, T.; Debut, L.; Sanh, V.; Chaumond, J.; Delangue, C.; Moi, A.; Cistac, P.; Rault, T.; Louf, R.; Funtowicz, M.; Brew, J.: HuggingFace's Transformers: State-of-the-art Natural Language Processing. ArXiv abs/1910.03771/, 2019.

Bidirectional Transformer Language Models for Smart Autocompletion of Source Code

Felix Binder[1] Johannes Villmow[1] Adrian Ulges[1]

Abstract: This paper investigates the use of transformer networks – which have recently become ubiquitous in natural language processing – for smart autocompletion on source code. Our model *JavaBERT* is based on a RoBERTa network, which we pretrain on 250 million lines of code and then adapt for method ranking, i.e. ranking an object's methods based on the code context. We suggest two alternative approaches, namely unsupervised probabilistic reasoning and supervised fine-tuning. The supervised variant proves more accurate, with a top-3 accuracy of up to 98%. We also show that the model – though trained on method calls' full contexts – is quite robust with respect to reducing context.

Keywords: smart autocompletion; deep learning; transformer networks

1 Introduction

AI-based support in software engineering has recently emerged as a research field, and recommenders for software commits [Da16], prediction of code changes [Zh19] or semantic code search [Hu19] have been developed. These are usually trained on vast amounts of source code and documentation from open-source platforms such as GitHub. Another challenge – and the subject of this paper – is *smart autocompletion*: As the developer types source code, a *neural network* suggests names for methods to use next. We refer to this challenge of ranking an object's method names by their plausibility in a given code context as *method ranking*. Figure 1 illustrates this, where a neural network has learned a suggestion from GitHub projects including code passages similar to the target context.

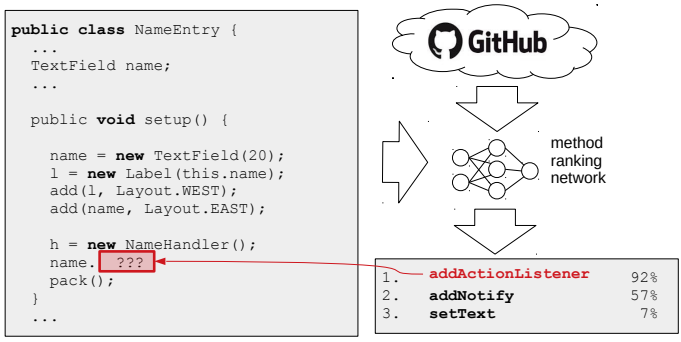


Fig. 1: A method ranking network analyzes a position in Java source code (red, left), and infers that – out of the class `TextField`'s methods – `addActionListener()` seems most plausible.

[1] RheinMain University of Applied Sciences, DCSM Department, Wiesbaden/Germany,
felix.d.k.binder@gmail.com, [johannes.villmow, adrian.ulges]@hs-rm.de

doi:10.18420/inf2020_83

While previous work on method ranking has used *n*-grams [Hi12] or recurrent networks [Wh15], we evaluate the transformer-based *masked language model* BERT [De18] (more precisely, its RoBERTa variant [Li19]). This approach has been very successful in natural language processing, but there has been – to the best of our knowledge – only one recent publication on smart autocompletion in source code [Ki20]. We call our model *JavaBERT* (since our focus lies on the Java programming language).
To utilize JavaBERT for method ranking, we propose two alternatives addressing the fact that method names may consist of multiple tokens (e.g., `add/Action/Listen/er`).:

1. *JavaBERT-unsup*: The pretrained (unsupervised) JavaBERT model is applied by masking out varying numbers of tokens. JavaBERT's predictions on token level are then combined in a probabilistic reasoning to predictions on method level.
2. *JavaBERT-sup*: JavaBERT is fine-tuned supervisedly as a binary classifier, estimating whether a certain method call is plausible or not in a given code context.

We evaluate both models in quantitative experiments on random samples from the GitHub Java Corpus [AS13]. Our results indicate that masked language modeling is surprisingly accurate, with a top-3 accuracy of up to 98%. We also study the impact of different contexts, e.g. only the code up to the target method call, or shorter vs. larger pieces of code.

2 Related Work

Smart autocompletion The task of code completion has been addressed since 2012 by using n-gram models [Hi12, AS13], cached *n*-gram models for improved localization [Fr15, TSD14, HD17] and graph-based statistical language models [NN15]. More recently, the availability of large code bases has facilitated the creation of neural network language models, including recurrent neural networks [Wh15, Ra16, Li17], gated recurrent neural network models [KS19] and LSTM models [Da16]. Most recent code completion models for Java use a single layer gated RNN [Ka20] model with Byte-Pair Encoding [SHB16].

Transformer networks in NLP Transformer networks [Va17] use the concept of attention [BCB14] to derive contextualized representations for the single tokens from a sequence (i.e. a sentence or paragraph). Probably the most prominent model is BERT [De18], which applies masked language modeling, i.e. the model is trained to predict random masked tokens in the training text. Other variants use generative transformers (GPTs) trained by left-to-right language modeling [Ra18], optimize hyperparameters such as model depth and learning rate (RoBERTa [Li19]), reduce the amount of parameters (ALBERT [La20]) or perform an adversarial training (ELECTRA [Cl20]). While transformer models have been extremely successful and intensely studied in processing *natural language* recently, we are only aware of one recent publication employing them for code completion [Ki20]. While this work uses an autogenerative model, we employ *masked* language modeling on pieces of source code.

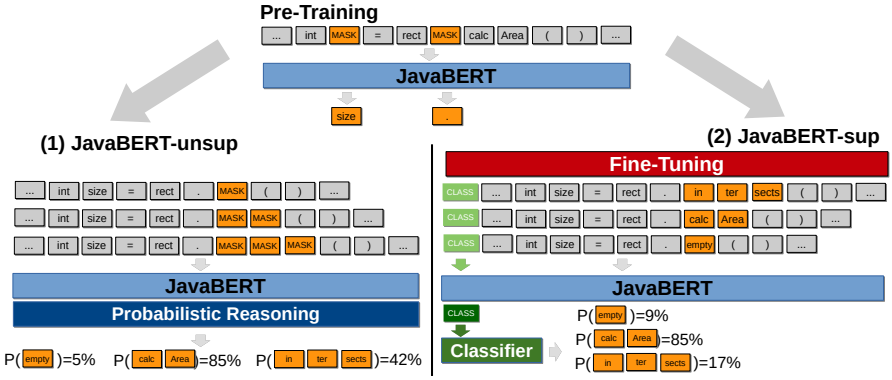


Fig. 2: Our approach JavaBERT is first pretrained using Masked Language Modeling (top). Afterwards, method ranking can either use masking with probabilistic reasoning (JavaBERT-unsup, left) or fine-tuning with a binary classifier (JavaBERT-sup, right).

3 Approach

As shown in Figure 2, our approach pretrains an encoder (JavaBERT) by masked language modeling. The resulting model can either be applied in an unsupervised fashion (using probabilistic reasoning) or by fine-tuning it into a supervised binary classifier. We discuss these three processing steps in depth in the following subsections.

3.1 Pretraining

Our approach starts with training a RoBERTa model on the Github Java Corpus [AS13] using the *fairseq* library [Ot19]. RoBERTa replicates and improves key hyperparameters of the well-known BERT model [De18] for a more robust training.

The training set of the Github Java Corpus consists of around 1.1 billion tokens of Java source code, from which we separate the last 5% into a validation set. We tokenize all Java code with a language parser[2] to separate natural language identifiers from syntax symbols. Thereby, we also remove multiple whitespace and replace string, character, float, and integer literals with a respective constant (e.g. `<INT>`). As source code may contain unicode identifiers, we use the *unidecode* library to transcribe any non-ASCII letters into ASCII.

Afterwards, we train and apply a Byte-Pair Encoding [SHB16] of V=10,000 sub-words on the tokenized source code. We perform neither lower-casing nor camel-case splitting as it is often done in machine learning on source code [ALY18]. A vocabulary of $10K$ tokens is used, which we assume is sufficient for source code (most identifiers are rather short or combined with camel-casing) while improving training speed. For example, our preprocessing tokenizes `public float myFloat = 10.0;` into `public/float/my/Float/=/<FLOAT>/;`.

[2] We use the *javalang* library for tokenization.

Model and Training We use a configuration similar to the RoBERTa$_{BASE}$ model (12 layers, 768-dimensional embeddings, 12 attention heads, 110M params total). Like ToBERTa, we use GELU activation functions [HG16], learned positional embeddings and a dropout of 0.1 throughout all layers and activations. The model is optimized with the Adam optimizer [KB14] (β_1=0.9, β_2=0.98, ϵ=10^{-6}, weight decay 0.01) using a linear warm-up of the learning rate for $6K$ steps up to $6{\times}10^{-4}$ followed by a cosine decay over $30K$ steps.
For efficiency reasons, we first train $15K$ steps on shorter code blocks of 128 tokens each (batch size $8K$) and then increase the block length to 512 (batch size $1,500$). The sampled blocks do not cross document (i.e. source file) boundaries. We use *gradient accumulation* to mimic larger batch sizes on our limited hardware (6 NVIDIA GeForce GTX 1080 Ti).
After a total training time of about two weeks, the JavaBERT model reached a validation perplexity of 1.16 for predicting masked out tokens, which is significantly better than common results for *natural* language (which is more ambiguous and unstructured).

3.2 Unsupervised Method Ranking (JavaBERT-unsup)

We assume an incomplete piece of code to be given, which is a sequence of n tokens $T{:=}(t_1,\ldots,t_s,\ldots,t_n)$ containing a slot t_s for the missing method call, and a set of candidate method names $\{C^1,\ldots,C^l\}$ which contains the correct method call C^* as well as other method names from the same class as C^*. The task is to maximize the probability $P(C^*|T)$. Note that – after tokenization – each candidate method name may consist of multiple tokens, i.e. $C{=}(c_1,\ldots,c_L)$ with $L{>}1$.
Note that the JavaBERT model's original training is *similar* to method ranking: The original source sequence is transformed into a sequence T', in which – similar to our input sequence – random tokens t_i have been masked out by replacing them with a `<mask>` token. During training the probability $P(T_i{=}t_i|T')$ of predicting the masked token is maximized. The key difference with method ranking is that *multi-token* method names have to be predicted. To do so, we calculate the probability of a candidate method (e.g., $C{=}(c_1,..,c_L)$) by replacing the slot token t_s with L `<mask>` tokens, obtaining a sequence T^L. Then, the overall probability of candidate C is defined as

$$P(C|T) = P(L)\cdot\prod_{j=1}^{L} P(T_{s+j-1}{=}c_j|T^L) \tag{1}$$

$P(L)$ acts as a prior on method name length, exploiting the fact that shorter names are more likely (an estimate on the training set is given in Table 1).

L	1	2	3	4	5	6	7	8	9	10+
$P(L)$	26.6%	23.9%	21.0%	12.4%	6.9%	3.9%	2.2%	1.2%	0.7%	1.2%

Tab. 1: The prior $P(L)$ indicates the probability of different method name lengths L.

3.3 Supervised Method Ranking (JavaBERT-sup)

Our second approach *fine-tunes* the pretrained JavaBERT model using a supervised training. The idea is to insert candidate method names into code blocks and estimate their plausibility with a binary classifier. To do so, we use a classifier token t_{CLASS} (as is common practice), replace the slot token t_s with the candidate C, and add markers t_{START} and t_{END} before the bound object t_{OBJ} and after the method call, resulting in the input sequence

$$T^C := (t_{CLASS}, \ldots, t_{START}, t_{OBJ}, t_{dot}, \overbrace{c_1, \ldots, c_L}^{=C}, t_{END}, \ldots, t_n).$$

We encode this sequence with the JavaBERT model and feed the resulting contextualized classifier token into a binary classifier, which is trained to predict whether T^C contains the correct method name. The classifier first projects the encoded representation with a linear layer into another embedding space of the same dimensionality as the JavaBERT model, followed by layer normalization and a second projection to our binary output space.
A training set of $3.3M$ labeled code blocks is constructed of 2,649 repositories from the GitHub Java Corpus' test split. Each positive sample (containing the true method call) is complemented with six negative samples, three of which feature another method name from the same class and the other three containing another random method name from the corpus. The model was fine-tuned for 5 days on 4 GPUs.

4 Experiments

This section compares the models JavaBERT-sup and JavaBERT-unsup in quantitative experiments on held-out test data from the Github Java Corpus. We also analyze how different amounts of context information affect the model's accuracy.
For this, we use another part of the original test split, which consists of 969 repositories that are not overlapping with the repositories used for pretraining or fine-tuning from Section 3. From these test projects, we sample $14K$ random code blocks of up to 504 tokens each. In each block, a randomly selected method call is chosen as slot t_s, and a list of candidate method names to be ranked is extracted from the method call's bound object's class. The median length of those candidate lists is 39.

Comparison Unsupervised vs. Supervised We compare the supervised and unsupervised model by measuring the hits@1, hits@3 and hits@5 rates, and the mean reciprocal rank (MRR). For example, a hits@3 of 98% indicates that for 98% of our $14K$ test blocks, the model ranks the correct method name among the top 3. Table 2 illustrates the results. Both approaches surpass a baseline that ranks the target methods randomly and the supervised approach outperforms the unsupervised model by a significant margin, especially for the top 1 predictions (difference $\approx$ 15%) and mean reciprocal rank (difference $\approx$ 10%).
Figure 3 illustrates an example for a random piece of code. Here, both models (supervised and unsupervised) rank the 8 candidate methods from Class `Scanner`, and the method `nextInt()` is ranked highest correctly. Overall, we found the model to prefer methods

	JavaBERT-unsup	JavaBERT-sup	Random guessing
hits@1	77.5	**92.3**	7.4
hits@3	92.5	**98.0**	20.8
hits@5	94.6	**98.9**	29.9
MRR	85.4	**95.2**	18.3

Tab. 2: Results of method ranking: The supervised approach significantly outperforms the unsupervised one and shows remarkable accuracy (hits@3 is 98%). We report all values as a percentage.

```
import java.util.Scanner;

public class EvenOdd {
  public static void main(String[] args) {
    Scanner reader = new Scanner(System.in);
    System.out.print("Enter a number: ");
    int num = reader . SLOT  ();
    String evenOdd = (num % 2 == 0) ? "even":"odd";
    System.out.println(num + "is" + evenOdd);
  }
}
```

JavaBERT-unsup	JavaBERT-sup
1: nextInt	1: nextInt
2: nextLong	2: nextShort
3: nextShort	3: close
4: nextByte	4: hasNext
5: skip	5: nextByte
6: close	6: skip
7: hasNext	7: locale
8: locale	8: nextLong

Fig. 3: Given this example code (left) with a left out target method (`SLOT`), both JavaBERT variants rank the correct method (`nextInt()`) out of 8 candidate methods highest.

with the correct parameters and return types (e.g., boolean methods are ranked high in if-statements). Note that this is not inferred from a static code analysis but only from the method name (e.g., `isEmpty`, `hasConnection`). Also, we found methods that have already been defined or used in the context to be ranked higher.

Context Analysis So far, we have trained and tested our model on *full* code contexts, including the code before and after the target method as well as their arguments. In practice, e.g. when typing code from left to right, only the code before the target may be available. Also, it is interesting how much context is required for a stable inference. Therefore, we evaluate JavaBERT-sup (trained with full contexts) on various forms of reduced context:

- **Original**: uses the complete context.
- **PC** (**P**receding **C**ontext): all tokens after the candidate method are removed, the context only consists of preceding tokens.
- **FC** (**F**ollowing **C**ontext): all tokens before the candidate method call (more precisely, before the bound object) are removed.
- **PC+ParaC** (**P**receding **C**ontext plus **Para**meter **C**ontext): Most tokens after the candidate method token are removed. Only eight complete words or symbols following the candidate method are kept, which can contain up to four parameters.
- **FAM** (**F**ew words **A**round **M**ethod): Only eight complete word or symbols preceding and following the method call are kept.
- **MAM** (**M**ore words **A**round **M**ethod): Only 40 complete word or symbols preceding and following the method call are kept.

	Original	**PC**	**FC**	**PC+ParaC**	**FAM**	**MAM**
hits@1	**92.3**	70.5	73.2	86.0	61.9	77.7
hits@3	**98.0**	86.1	85.5	94.3	73.8	87.3
hits@5	**98.9**	90.5	88.8	96.1	77.0	89.8
MRR	**95.2**	79.0	80.0	90.0	69.0	83.0

Tab. 3: Comparing JavaBERT's ranking accuracy with different context windows.

These experiments are based on the same test set as before. Since the location of the target method call in a code block is chosen randomly, the amount of text for different context forms varies accordingly. Table 3 shows the results of this experiment. As expected, using the full context performs best. The follow-up run is PC+ParaC, indicating that the parameters of a method call form an important source of information for the ranking model.
Comparing the results from FAM to PC, FC and MAM and from original to MAM showcases the influence of input size on the method ranking. An observation on the three best runs (Original, PC+ParaC and MAM) is that those are a combination of preceding and following content and are ordered descendingly by the size of their input. The results from MAM show, however, that combining the preceding and following content has a larger influence on the method ranking than the input size when compared to PC and FC. In conclusion, using surrounding content rather than only the preceding content like left-to-right models does have an impact on the ranking of candidate methods.

5 Conclusions

In this paper, we have shown that transformer networks pretrained by masked language modeling are a promising approach towards method ranking. Particularly, we have demonstrated the benefits of supervised fine-tuning and studied different context windows, whereas surprisingly small context windows combining a bit of preceding and following code suffice for an accurate inference. Our future work will focus on enhancing JavaBERT (which is a token-only model) with syntax trees to obtain richer code representations, as well as tackling other challenges in automated source code understanding, such as code search and summarization.

Bibliography

[ALY18] Alon, Uri; Levy, Omer; Yahav, Eran: code2seq: Generating Sequences from Structured Representations of Code. CoRR, abs/1808.01400, 2018.

[AS13] Allamanis, Miltiadis; Sutton, Charles: Mining Source Code Repositories at Massive Scale using Language Modeling. In: Proc. MSR. pp. 207–216, 2013.

[BCB14] Bahdanau, Dzmitry; Cho, Kyunghyun; Bengio, Yoshua: Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473, 2014.

[Cl20] Clark, Kevin; Luong, Minh-Thang; Le, Quoc V.; Manning, Christopher D.: ELECTRA: Pre-training Text Encoders as Discriminators Rather Than Generators. In: Proc. ICLR. 2020.

[Da16] Dam, Hoa Khanh; Tran, Truyen; Grundy, John; Ghose, Aditya: DeepSoft: A Vision for a Deep Model of Software. In: Proc. 24th ACM SIGSOFT FSE. p. 944–947, 2016.

[De18] Devlin, Jacob; Chang, Ming-Wei; Lee, Kenton; Toutanova, Kristina: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. 2018.

[Fr15] Franks, C.; Tu, Z.; Devanbu, P.; Hellendoorn, V.: CACHECA: A Cache Language Model Based Code Suggestion Tool. In: Proc. ICSE. pp. 705–708, 2015.

[HD17] Hellendoorn, Vincent J.; Devanbu, Premkumar: Are Deep Neural Networks the Best Choice for Modeling Source Code? In: Proc. ESEC/FSE 2017. pp. 763–773, 2017.

[HG16] Hendrycks, Dan; Gimpel, Kevin: Gaussian error linear units (gelus). arXiv preprint arXiv:1606.08415, 2016.

[Hi12] Hindle, Abram; Barr, Earl T; Su, Zhendong; Gabel, Mark; Devanbu, Premkumar: On the Naturalness of Software. In: 2012 34th ICSE. IEEE, 2012.

[Hu19] Husain, Hamel; Wu, Ho-Hsiang; Gazit, Tiferet; Allamanis, Miltiadis; Brockschmidt, Marc: CodeSearchNet Challenge: Evaluating the State of Semantic Code Search. arXiv preprint arXiv:1909.09436, 2019.

[Ka20] Karampatsis, Rafael-Michael; Babii, Hlib; Robbes, Romain; Sutton, Charles; Janes, Andrea: Big Code != Big Vocabulary: Open-Vocabulary Models for Source Code. arXiv preprint arXiv:2003.07914, 2020.

[KB14] Kingma, Diederik P; Ba, Jimmy: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980, 2014.

[Ki20] Kim, Seohyun; Zhao, Jinman; Tian, Yuchi; Chandra, Satish: Code Prediction by Feeding Trees to Transformers. arXiv preprint arXiv:2003.13848, 2020.

[KS19] Karampatsis, Rafael-Michael; Sutton, Charles A.: Maybe Deep Neural Networks are the Best Choice for Modeling Source Code. CoRR, abs/1903.05734, 2019.

[La20] Lan, Zhenzhong; Chen, Mingda; Goodman, Sebastian; Gimpel, Kevin; Sharma, Piyush; Soricut, Radu: ALBERT: A Lite BERT for Self-supervised Learning of Language Representations. In: Proc. ICLR. 2020.

[Li17] Li, Jian; Wang, Yue; King, Irwin; Lyu, Michael R.: Code Completion with Neural Attention and Pointer Networks. CoRR, abs/1711.09573, 2017.

[Li19] Liu, Yinhan; Ott, Myle; Goyal, Naman; Du, Jingfei; Joshi, Mandar; Chen, Danqi; Levy, Omer; Lewis, Mike; Zettlemoyer, Luke; Stoyanov, Veselin: RoBERTa: A Robustly Optimized BERT Pretraining Approach. arXiv preprint arXiv:1907.11692, 2019.

[NN15] Nguyen, A. T.; Nguyen, T. N.: Graph-Based Statistical Language Model for Code. In: Proc. ICSE. pp. 858–868, 2015.

[Ot19] Ott, Myle; Edunov, Sergey; Baevski, Alexei; Fan, Angela; Gross, Sam; Ng, Nathan; Grangier, David; Auli, Michael: fairseq: A Fast, Extensible Toolkit for Sequence Modeling. In: Proc. NAACL-HLT. pp. 48–53, 2019.

[Ra16] Raychev, Veselin: Learning from large codebases. PhD thesis, ETH Zurich, 2016.

[Ra18] Radford, Alec; Narasimhan, Karthik; Salimans, Tim; Sutskever, Ilya: Improving language understanding by generative pre-training. OpenAI Blog, 2018.

[SHB16] Sennrich, Rico; Haddow, Barry; Birch, Alexandra: Neural Machine Translation of Rare Words with Subword Units. In: Proc. ACL. Berlin, Germany, pp. 1715–1725, August 2016.

[TSD14] Tu, Zhaopeng; Su, Zhendong; Devanbu, Premkumar: On the Localness of Software. In: Proc. 22nd ACM SIGSOFT FSE. ACM, pp. 269–280, 2014.

[Va17] Vaswani, Ashish; Shazeer, Noam; Parmar, Niki; Uszkoreit, Jakob; Jones, Llion; Gomez, Aidan N; Kaiser, Łukasz Kaiser; Polosukhin, Illia: Attention is All you Need. In: NIPS 2017, pp. 5998–6008. 2017.

[Wh15] White, Martin; Vendome, Christopher; Linares-Vásquez, Mario; Poshyvanyk, Denys: Toward Deep Learning Software Repositories. In: Proc. MSR. pp. 334–345, 2015.

[Zh19] Zhao, Rui; Bieber, David; Swersky, Kevin; Tarlow, Daniel: Neural Networks for Modeling Source Code Edits. arXiv preprint arXiv:1904.02818, 2019.

A Decade of Energy Awareness Technology Evolution for Sensor Nodes

Marcus Thoss[1]

Abstract: Energy awareness is an important aspect of the design of sensor node hard- and software, particularly for battery-powered or energy-harvesting node architectures. Architectural design choices of such systems must regard a multitude of aspects, including size, weight, memory and processing power. Therefore, energy-related design aspects have in recent years become a feature that is being honoured throughout sensor node design. As a result, various technological solutions and strategies have evolved to facilitate energy awareness and energy management aspects.

This paper looks back at the evolution of sensor node technology during the 2010s. In the course of research activities, the state of the art of research and industrial solutions aiming at improving the support of energy awareness was being monitored. Advances were observed for various aspects and levels of relevant technological facets, including electronic measurement and control circuitry, harvesting facilities, power-saving mechanisms at both hard- and software level, energy management strategies and algorithms, networking aspects, and advances and extensions related to operating systems for sensor nodes. A conclusion of these observations, given in this paper, identifies technological increments, leaps and sidesteps that have occurred along the way.

For the overall time span of the decade observed, a short qualitative and quantitative analysis of the technological advances achieved is presented, including typical examples of actual sensor node designs. The paper concludes with an outlook on further evolution of advances in energy awareness technology for sensor nodes to be expected in the near future and to be desired in the long run.

Keywords: Energy Awareness; Sensor Node; Technology Review

1 Motivation

The Internet of Things (IoT) is being rolled out with increasing speed and in a multitude of application areas with competing and complementing communication technologies like 5G networks, massive satellite fleets, LoRA, WPANs and, ultimately, the classical Internet connections. Yet, current reports like Microsoft's "IoT Signals" [Mi19] clearly show that there is not only a strong willingness to adopt IoT technologies and many platforms to choose from, but also a huge percentage of projects (30% according to the report) failing to reach deployment.

It can be assumed that IoT scenarios relying on widely distributed small-scale self-powered nodes, with or without energy harvesting, are rather common. Engineering of this type of

[1] RheinMain University of Applied Sciences, Department of Design – Computer Science – Media, Unter den Eichen 5, 65195 Wiesbaden, Germany, marcus.thoss@hs-rm.de

doi:10.18420/inf2020_84

device requires a focusing on the aspects of size and weight, node lifetime, connectivity and maintenance. All of those benefit from a better handling of the energy resources available, sometimes in an indirect manner, as when connectivity is being improved by designing a system around the goal to offer the power to transmit just when an application requires it.

When [Th10] was published 10 years ago, the IoT was still in its infancy. Looking back at this point in time, the decade that has passed since then is referred to in this paper as "the decade". A precursor of the IoT had been the previous 10-year long increase in RFID technology usage during the 2000s. Subsequently, especially in Europe, numerous research programmes for then-topical Ambient Assisted Living were funded, multi-national research cooperations like the European Research Cluster on the Internet of Things[2] began to assemble and the first platform solutions like UniversAAL [Fe15] emerged.

From then on, interest in sensor network technology, both among industry and researchers, boomed, and technological advances, some of which will be described here, helped to increase reliability, applicability to more domains, and efficiency of the solutions. Besides connectivity improvements, both at sensor network and internet-based cloud levels, an increase of the regard for energy awareness issues was a major factor of the overall development observed.

2 Technological Facets of Energy Awareness

Design measures and system components contributing to (or hindering) energy awareness can be found throughout any systems architecture. Therefore, challenges and the potential of different technological facets should first be looked at separately to be considered for an overall, system-wide design strategy for energy awareness.

2.1 Energy Sources

Energy sources can be rated by many factors, the most relevant here being voltage range and maximum power output. They can be further classified into storage (primary and secondary cells, capacitors and solid state storage) and on-line sources (wired or harvesting sources).

The necessity to rate energy consumed by sensor nodes based on a cost factor, which decreases from primary cells to harvesting solutions, had been observed at least by the beginning of the decade [Th10]. With regard to cost as a general concept, rechargeable storage components can be neglected as they do not change the overall energy balance, energy harvesting supplies can be tagged with a no-cost factor, wired power sources add a moderate, constant cost, and primary cells would be considered the most expensive form of energy source both in terms of supply logistics and environmental impact.

[2] http://www.internet-of-things-research.eu

Consequently, the overall development over the decade moved from primary cells to swappable secondary cells to energy harvesting approaches. Sadly, many industry solutions are still deliberately designed to use primary cells because their lifetime often exceeds that of the application, and cost per cell is, at least monetarily, cheap.

As to the technology of harvesting sources, the most interesting development happened and is to be further expected among solar cell technology. With some cell types now nearing an efficiency of 30%, while 25% had been a good rating by the beginning of the decade (according to [Gr10], the report has been updated regularly since), size of the cells for a given node uptime could be decreased and harvesting nodes can now be utilised in a much broader range of environments.

2.2 Energy Storage

Storage-type energy sources introduce additional parameters like capacity, possibly recharge characteristics, wear-out, size, weight and operating temperature range. Regarding purely economic motivations, secondary cells of the Li-Ion and related types have become favoured over the decade, as small form factors like coin cells were available and the technology had advanced to allow energy harvesting to provide enough recurring energy input for many applications. The other trend – that did not gain as much momentum – is the use of supercapacitors as rechargeable storage element. From an environmental perspective, the use of carbon-electrode supercapacitors would be highly preferable compared to the more problematic secondary cell technologies often employing scarce or toxic substances. The problems of still minor capacities and more challenging output voltage curves, although partially outweighed by a high number of possible recharge cycles, leaves this technology so far a candidate for future advances.

2.3 Power Controllers

The evolution of microcontroller technology has led to single-supply voltage solutions, as modern controllers are capable of internally generating themselves diverse voltages required. Although this had been the case by the beginning of the decade already, input voltage ranges now become broader and single-supply architectures are even more common.

On the supply side, power management becomes more complex when storage and harvesting strategies are included. Here, technological evolution has produced an increase in efficiency of conversion and the availability of highly integrated power controllers. These can manage multiple harvesting and wired sources, intermediate energy storage, output voltage and power control, and require few external components. Thus, the complexity of node hardware designs could be reduced dramatically, while efficiency was increased.

2.4 Microcontrollers

In the case of a sensor node, the processor is almost certain to be a microcontroller. A consolidation of instruction set architectures could be observed among semiconductor manufacturers from a multitude of proprietary designs towards classical Intel 8051 Cores for 8-bit processors and ARM architectures covering 32-bit. Notable exceptions are Microchip (including former Atmel), remaining successful with their 8-bit PIC and AVR architectures, and Texas Instruments with their 16-bit MSP430 processor family.

Microcontrollers aid energy aware designs by the fact that power domains of processor core and peripherals are closely coupled and compatible, reducing transfer and conversion losses. A notable improvement has been introduced by controllable power gating of embedded peripherals and an increase of detail in sleep state modelling, both regarding the number of levels and the granularity of control that can be exerted.

2.5 Memory

The choice of RAM type for low-energy designs is of course static RAM offering negligible idle-state power losses. Thus, driven by the need for more complex sensor node firmware, RAM size offers have been increased from a few KB to some 100 KB without greater penalty, as relevant power drains occur only during RAM access, which is a matter of limiting the amount of code execution.

Not yet frequently available in common architectures are the follow-up non-volatile, near-RAM-access-speed technologies FRAM and MRAM, with FRAM being available in some MSP430 variants as the only mainstream platform. A major disadvantage of SRAM as main memory is the loss of information in power-off and deep sleep states. Since zero-power sleeping is a desirable state, though, energy aware designs must currently apply measures to persist or re-initialise RAM contents during these states. Should non-volatile memory technologies become the standard for main memory designs, powerless sleeping would become available with less overhead and complexity. In [Re19], energy consumption measurements for various sleep states and application aspects were analysed.

2.6 Energy Measurement

Energy measurements can basically be taken with an external meter or, the node can measure its own energy flows. In any case, voltage and current must be determined to calculate a momentary $P = U \cdot I$ and then integrate it over a time span Δt to get $E_{\Delta t} = \int_{t_0}^{t_0+\Delta t} P dt$. For external measurements, the problem of mapping samples taken to the network node's time domain poses a challenge. So, although external measurements can often afford more precise metering equipment, internal measurements are desirable, i.e. the note would measure itself.

Thus, initially, analog input pins of the microcontroller were, and still are, used to measure battery and operating voltage, and a shunt resistor is inserted in the supply path to determine the current by measuring the voltage drop created.

Over the decade, energy aware node designs emerged that were equipped with dedicated on-board power metering solutions, improving accuracy and sometimes allowing the microcontroller to sleep while measurements continue. Yet, such additional circuitry would consume power itself, degrading node uptime and falsifying measurement values.

There is a kind of "holy grail" of such measurement solutions beginning to be realised more frequently. It is the monitoring of the embedded switched-mode power supply (SMPS) circuits managing the supply voltages of the node, as noted in Section 2.3. An SMPS performs demand-based regulation by transferring chunks of energy from a supplier to the consumer. Thus, the amount of energy transferred is proportional to the number of chunks converted. The trick applied here is to have the microcontroller itself count the switching events by adapting the pulses through appropriate decoupling and probing circuitry.

Such solutions promise to minimise measurement overhead, both in terms of node circuit complexity and energy investment. As most modern microcontrollers are capable of counting events in deep sleep modes, even the requirement of continuous monitoring can be fulfilled.

2.7 Operating Systems

There has been a notable evolution of embedded operating systems (OS) meeting sensor network technology requirements. Foremost, strong connectivity support is obviously a necessity. Secondly, to be applicable to a broad range of projects, multi-platform support, also scaling from low-end architectures for small sensor nodes to those found in rather well-equipped router and gateway nodes, would contribute to the popularity of a sensor network OS.

Energy awareness did not seem to be of primary concern by the beginning of the decade, and has actually begun to be fully integrated into the OS architectures and APIs only recently. To some extent, the open source platforms Contiki OS [DGV04] and Tiny OS [Le05] pioneered the sensor node research community, and Contiki considered on-line, OS support for energy measurements early on, providing the important feature of attributing measurement data with the execution context of the measurement. During the decade, the open source embedded OS scene saw Mbed OS (2009), Apache MyNewt (2015), Linux Foundation's Zephyr (2016), and only very recently, RIOT (2018) appear. Mbed is being driven by the ARM company, targeting their own architectures, while Texas Instruments has provided their own closed-source solution TI RTOS for their hardware platforms, notably the MSP430 family. They all show activities towards providing better support for energy awareness, with RIOT having a strong research background forstering its energy awareness aspects [Ha16].

2.8 Testbeds

By the middle of the decade, the survey published in [HH14] showed that a large number of approaches had accumulated, and researchers and industry had been busy investigating sensor networks through testbeds. The authors of the survey themselves had been part of the community from around 2010 on, and their SANDbed setup [HWM10] explicitly focused on energy monitoring. Testbeds continue to evolve regularly within the communities developing platform solutions, a recent, notable example being the RIOT testbed [GEN19].

Basically, there are three methods for assessing energy budgets, that is, model-based prediction (optionally based on previous off-line measurements), and internal and external on-line measurement. With an approach reasonably limiting the overhead of the measurement action, on-line measurements of the actual operation of the sensor network would be preferable. It requires a measurement infrastructure, though, that is able to retrieve and collect the measurement data and offers sufficiently comfortable testbed setup, experiment control and evaluation support. These requirements are likely to be the reason why many solutions still rely on off-line measurements and simulated or modelled energy assessments.

3 Example Designs

3.1 Harvesting Iris Mote

An early, rather simplistic energy aware harvesting sensor node solution was presented and analysed in 2011 in [RMT12], providing all basic ingredients based on the Iris mote platform from MEMSIC. The design and the analysis paper anticipate many problems and approaches that have been discussed in the community in the following years. The researchers added a solar cell, supercapacitor and power management to the base platform, considered energy measurement by counting switched mode power supply events and looked at the pros and cons of simulated off-line versus measurement-based on-line assessment strategies.

3.2 WieDAS

In this research project, the sensor node box and its solar panel measure 8 cm by 12 cm each, being voluminous by today's standards. The size was caused by the use of AAA Batteries and a single-layer carrier board for prototype production. Still these nodes were successfully operated in an AAL demonstrator scenario described in [Kr14]. Energy awareness was initially limited to battery voltage measurement using the main microcontroller and attaching a current probe in the supply line. A major step towards power awareness was the addition of a TI INA219 current sensor. This device is capable of determining power consumption

measuring both voltage current and even averaging power values over programmable window sizes. Thus, with constant sampling frequency f and $N_{samples}$ per window, energy transfers can can be approximated by $\Delta E = P \cdot \Delta t$ with $\Delta t = N_{samples} \cdot \frac{1}{f}$.

Adding this external measurement component allowed leaving the microcontroller in deep sleep modes without stopping measurements, if the CPU was woken up in time to collect buffered measurement data. One drawback was the late addition of solar energy harvesting, which did not match the inflexible power domain scheme consisting of jumpers that had to be removed manually to disconnect single sensors of the board from the power supply.

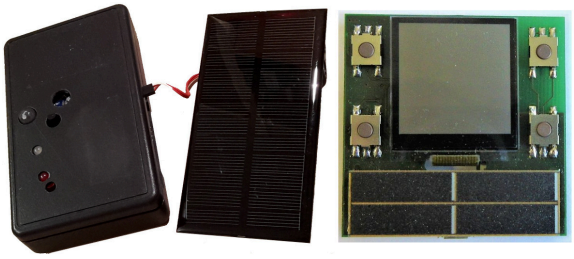


Fig. 1: WieDAS (left) and BASE MOVE (right) nodes compared

3.3 BASE MoVE

By the end of the decade, the BASE MoVE project [Ak20] had created custom sensor nodes, too, which measured less than half the size of a WieDAS node, but included comparatively tiny solar cells and a large display (shown in Figure 1, nodes are not to scale). The project created node hardware with programmable power gating of sensor peripherals, energy harvesting realisations for solar and thermal sources, and an automated testbed integrated with build automation for regression tests. Automated distributed energy measurements and energy management, that had been planned, finally exceeded the scope of the project.

4 Conclusion and Outlook

Within the given context, the project LONG MOVE[3], a successor to BASE MoVE, focuses on node-local and distributed energy management topics for sensor networks. Its mission statement calls for “energy awareness by design”, taking this aspect into account from the beginning on. It thus reflects the insight that previous designs could have benefited from earlier, consistent involvement of energy-related questions. With OS and hardware platforms offering better support in this area, it can be expected that this will become an overall trend. In fact, energy awareness could be an integral part of designs when self-adaptation features embedded at platform level can obliterate to some degree the need to regard energy within the design of the application proper.

[3] https://www.hs-rm.de/de/fachbereiche/design-informatik-medien/forschungsprofil/long-move

Bibliography

[Ak20] Akelbein, J.-P.; Beckmann, K.; Hoss, M.; Schneider, S.; Seyfarth, S.; Thoss, M.: BASE MoVE - A Basis for a Future-proof IoT Sensor. In: INFORMATIK2020. Gesellschaft für Informatik, Bonn, 2020. to be published.

[DGV04] Dunkels, A.; Gronvall, B.; Voigt, T.: Contiki - A Lightweight and Flexible Operating System for Tiny Networked Sensors. In: 29th Annual IEEE International Conference on Local Computer Networks. pp. 455–462, 2004.

[Fe15] Ferro, E.; Girolami, M.; Salvi, D.; Mayer, Ch.; Gorman, J.; Grguric, A.; Ram, R.; Sadat, R.; Giannoutakis, K.; Stocklöw, C.: The UniversAAL Platform for AAL (Ambient Assisted Living). Journal of Intelligent Systems, Jan 2015.

[GEN19] Günes, M.; Engelhardt, F.; Nothnagel, K.: Technical report - Designing a Testbed for Wireless Communication Research on Embedded Devices. In: 18. GI/ITG KuVS FachGespräch SensorNetze, FGSN 2019. pp. 41–44, 2019.

[Gr10] Green, M. A.; Emery, K.; Hishikawa, Y.; Warta, W.: Solar Cell Efficiency Tables (v. 35). Progress in Photovoltaics: Research and Applications, 18(2):144–150, 2010.

[Ha16] Hahm, O.: Enabling Energy Efficient Smart Object Networking at Internet-Scale : Experimental Tools, Software Platform, and Information-Centric Networking Protocols. PhD thesis, University of Paris-Saclay, Dec 2016.

[HH14] Horneber, J.; Hergenröder, A.: A Survey on Testbeds and Experimentation Environments for Wireless Sensor Networks. IEEE Communications Surveys & Tutorials, 16(4):1820–1838, April 2014.

[HWM10] Hergenröder, A.; Wilke, J.; Meier, D.: Distributed Energy Measurements in WSN Testbeds with a Sensor Node Management Device (SNMD). In: Workshop Proceedings of the 23th International Conference on Architecture of Computing Systems. VDE Verlag, pp. 341–438, February 2010.

[Kr14] Kröger, R.; Lux, W.; Schaarschmidt, U.; Schäfer, J.; Thoss, M.; von Fragstein, O.: The WieDAS AAL Platform: Architecture and Evaluation. In: 7. Deutscher AAL-Kongress mit Ausstellung, 21.-22. Jan. 2014, Berlin. VDE Verlag GmbH, Berlin/Offenbach, 2014.

[Le05] Levis, P.; Madden, S.; Polastre, J.; Szewczyk, R.; Whitehouse, K.; Woo, A.; Gay, D.; Hill, J.; Welsh, M.; Brewer, E.; Culler, D.: TinyOS: An Operating System for Sensor Networks. In: Ambient Intelligence. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 115–148, 2005.

[Mi19] Microsoft Corporation: IoT Signals. Technical report, Microsoft Corporation, 2019.

[Re19] Reinhard, J.: Energiebetrachtungen für Sensorknoten mit FRAM (Energy analysis of FRAM-equipped sensor nodes). Lab report, Embedded Systems Course 2018/19, Rhein-Main Univerity of Applied Sciences, 2019.

[RMT12] Renner, C.; Meier, F.; Turau, V.: Holistic Online Eenergy Assessment: Feasibility and Practical Application. In: Ninth International Conference on Networked Sensing, INSS 2012, Antwerp, Belgium, June 11-14, 2012. IEEE, pp. 1–8, 2012.

[Th10] Thoss, M.: Supporting Energy Awareness in Distributed Embedded Systems. In (Dörner, Ralf; Krömker, Detlef, eds): Self Integrating Systems for Better Living Environments: First Workshop, Sensyble 2010. Shaker Aachen, pp. 117–124, Nov 2010.

R. Reussner, A. Koziolek, R. Heinrich (Hrsg.): INFORMATIK 2020,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2021 931

BASE MoVE - A Basis for a Future-proof IoT Sensor

Jens-Peter Akelbein[1], Kai Beckmann[2], Mario Hoss[3], Samuel Schneider[4], Stefan Seyfarth[5], Marcus Thoss[6]

Abstract: For a long time, the Internet of Things was considered the vision of interconnecting every device, leading to fundamentally new and pervasive application scenarios. In practice, however, the projected growth and realisation of IoT scenarios is often impeded by practical problems. The BASE MoVE research project, a cooperation between universities and industrial partners, took a holistic look at requirements and inhibitors for investing in IoT solutions, using Ambient Assisted Living as an application domain example. The perspectives of all stakeholders involved were taken into account during the design of a solution architecture, from the user to the manufacturer to the service provider and housing association. This paper presents the resulting modular base platform for IoT applications. Power supply through battery and energy harvesting reduces installation costs. The use of open source software and the support of several common smart home protocols also prevents a lock-in effect and dependency on a single manufacturer. This makes it possible to protect investments from market-driven changes from one manufacturer's ecosystem to another. Here, the paper takes an in-depth look into the choice of protocols. Over-the-air updates allow for secure operation as well as remote maintenance, no longer requiring expensive in-person maintenance. Finally, the manufacturing of the solution as a hardware module, as realised in BASE MoVE, also allows for easier creation and certification of new sensor devices in a company's product portfolio. To evaluate the developed solution, an apartment was equipped with different sensor devices, and a smart home scenario was implemented. The feasibility study could demonstrate that it is indeed possible to create a base platform that meets today's requirements of the stakeholders involved and allows for a sustainable, future-proof usage by offering adaptability to new technologies. In addition to the scientific results, the project also gave an assessment about component maturity and cost, which is valuable for the commercial project partner and its market entry strategy.

Keywords: Internet of Things; Ambient Assisted Living; Home Automation

1 Motivation

The vision of an „Internet of Things" (IoT) describes a technological revolution to be created by operating interconnected devices that pervade environments of personal life, work, and nature. Although distribution of such devices has been realised to a notable degree already, and solutions for many application areas, commercial and research platforms

[1] Darmstadt University of Applied Sciences, Schoefferstraße 8b, 64295 Darmstadt, jens-peter.akelbein@h-da.de
[2] RheinMain University of Applied Sciences, Unter den Eichen 5, 65195 Wiesbaden, kai.beckmann@hs-rm.de
[3] Darmstadt University of Applied Sciences, Schoefferstraße 8b, 64295 Darmstadt, mario.hoss@h-da.de
[4] Thermokon Sensortechnik GmbH, Platanenweg 1, 35756 Mittenaar-Offenbach, samuel.schneider@thermokon.de
[5] Thermokon Sensortechnik GmbH, Platanenweg 1, 35756 Mittenaar-Offenbach, stefan.seyfarth@thermokon.de
[6] RheinMain University of Applied Sciences, Unter den Eichen 5, 65195 Wiesbaden, marcus.thoss@hs-rm.de

doi:10.18420/inf2020_85

abound, there are numerous problems hindering further growth and flawless operation. Foremost, sustainability is still questionable, as the market has not yet stabilised, and new protocols and interfaces are still emerging. This, and the existence of so many variants calls for adaptability as a major feature of a design if it is meant to prevail. Without it, the growth of the IoT landscape is bound to decrease, or even cease.

It is therefore necessary to identify the relevant impeding factors that could endanger the future of the IoT. Functional correctness or applicability to the solution domain is rarely a problem. Instead, non-functional and platform-level factors must be considered critical for the success of a market solution. As, with increasing distribution of sensor nodes, most of these cannot use wired powering any more, and regular battery changes become impractical and too costly, self-sustained energy harvesting must become a mainstream technology. To further support sustainability, and the success of a platform solution, technological changes must be reacted upon by providing easy integration into future scenarios, should the protocol landscape change. This requires a high degree of adaptability and the possibility to update the firmware, and thus, possibly, support new protocols.

The project did not attepmt to re-invent base technologies readily available today. Nor was there a focus on offering a consistent application-level protocol and modelling solution, as investigated thoroughly in previous research projects. Instead, this project evaluated how a viable, flexible IoT platform could be created by integrating existing hardware, (embedded) operating systems (OS), and communications technologies. For the evaluation, an application scenario was implemented. Emphasis was also put on rendering the approach more future-proof with regard to technological changes by facilitating the exchange of communications technologies and protocols.

The acronym contained in the project name BASE MoVE states the main aspects regarded in the project. A dedicated sensor node hardware platform should be created as the **b**asis for the solution. **A**daptability is to be achieved by allowing over-the-air (OTA) firmware updates. **S**ecurity should be considered as a first-class design objective and thus be regarded from the very start, and finally, design for **e**nergy awareness at hard- and software level must lead to a viable self-powered sensor node architecture utilising energy harvesting technology. Application scenarios regarded in the project were meant to serve as technology test show cases for the validation of the fulfilment of these goals, whereas application level modelling was not considered a prime objective.

2 Related Work

Enabling real multi-protocol support in IoT-scenarios depends on the capabilities of the underlying hardware. If the transceivers are locked to specific radio protocols, the freedom to change or replace a protocol is limited. Most flexibility can be achieved if the radio transceivers are generic and the implementation of a specific protocol is a matter of hardware configuration and software.

There are hardware platforms like the EFR32 series from Silicon Labs [Sib] providing System-on-a-Chip (SoC) solutions embedding generic radio transceivers which allow for the usage of different protocols with the same device. Moreover, a subset of the EFR32 series embeds transceivers for different radio frequencies, like 2.4GHz and the sub-GHz band (915 resp. 868MHz or 433MHz). With their RAIL library, Silicon Labs provides several proprietary Smart-Home protocol stacks for their SoC family, like ZigBee, Thread, BLE, Z-Wave etc., and supports the creation of firmware running two stacks in parallel, like BLE and Zigbee [Sia].

Most hardware vendors provide proprietary protocol stack implementations for the application areas they are targeting. For the broader area of the IoT this can only cover a subset. There are many surveys like [SJ17], [Za18] or [Di19] gathering and categorising the protocols proposed and used, and solutions in the IoT or the smart home sector. Regarding the term „IoT“, there is a consolidating trend towards IP-based protocols noticeable in recent surveys. The emergence of IP for IoT protocols becomes apparent, considering the Zigbee Cluster Library having been made available for IP [Do], or the work towards IP over BLE [Ni15]. This is a significant change from the former situation, with manufacturers selling products combining hardware and software. Open vendor-independent protocol stack implementations for the IoT are provided by open source IoT operating systems, like RIOT OS [Ba18], mbed, Zephyr, Contiki or others [Qu18].

3 Architecture

Figure 1 shows the architecture for the IoT platform developed. It consists of the modular hardware platform and the software layer composed of an IoT OS, exchangeable protocol stacks, management and firmware update functionality and the top-level applications. The aim is to provide a flexible platform for fast and cost-efficient development of smart home and related IoT devices, supporting operation based on energy harvesting. The IoT protocols are kept exchangeable within the base platform to prevent lock-in effects for the OEM and customers. Furthermore, the usage of an open source OS is proposed to protect sold and installed smart home devices in households from becoming abandon-ware.

One important requirement for the *hardware* is its featuring low power modes that allow for energy harvesting devices. The transceiver should support arbitrary protocols, like BLE and 802.15.4, and should be able to support future protocols by using a generic transceiver. The selected Mighty Gecko EFR32MG13P7-33F512GM48 from Silicon Labs meets these requirements. It was manufactured as a PCB module to simplify certification and re-usability. Several iterative versions of four different devices were created. 1) A relay actuator with a permanent power supply. 2) A window contact with environment sensors and a reed contact. 3) A room control unit additionally equipped with a low power display and control buttons. 4) An occupancy sensor, replacing the display with a passive infrared (PIR) sensor. The types 2) to 4) are passive (end devices), reacting on external or timer events. They feature an energy harvesting subsystem and a monocrystalline solar cell for indoor use.

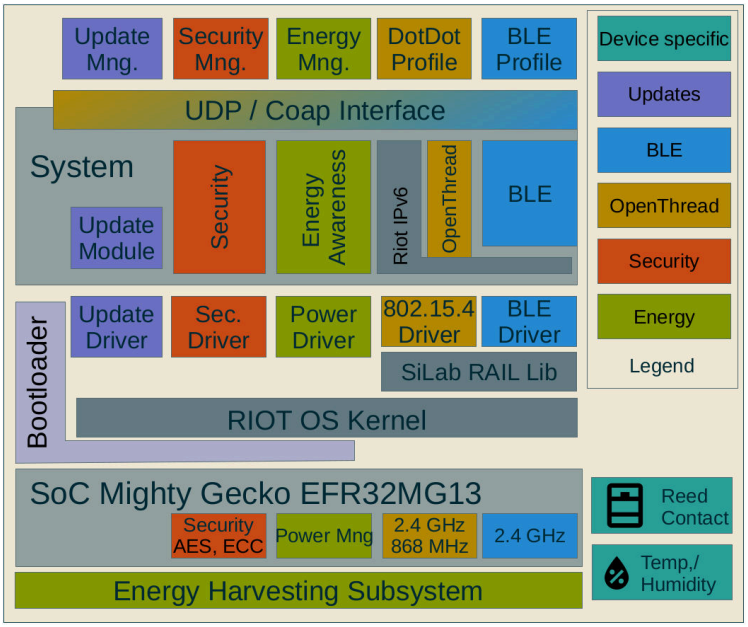


Fig. 1: Architecture of the "BASE MoVE" IoT platform

The OS as the *software* foundation was chosen based on a requirements analysis. Important aspects were the support of an open compiler like GCC, debugging support and it being an active project. Collaboration with the developer community and the ability to extend the OS with new features had to be possible. Satisfying these requirements, RIOT OS was chosen, even if it did not yet support all of the required hardware and protocol stacks. This was possible as the authors were familiar with RIOT OS, and estimated the development and integration of the missing functionality to be possible within the project.

The modular base platform supports exchangeable IoT protocols by flashing the device with different application images. The efforts to support multiple protocols within the same application, like Silicon Labs allows with their proprietary protocol stacks, were analysed and deferred due to the complexity and necessary modification of the OS. Furthermore, replacing the protocols through firmware update reduces the resource requirements to keep the devices viable, but requires firmware over the air (FOTA) functionality and the management of devices. To this aim, an additional management protocol was added, enabling the use of unaltered IoT protocol stacks. It is foreseeable that this approach will become obsolete, once a standard for FOTA is adopted by IoT protocols. On a device level an OS bootloader is used for a standard two slot approach.

The IoT platform supports wireless devices without a permanent energy source to reduce installation and maintenance costs. This comes at the cost of additional cross cutting

concerns, limiting the possible functionality. In addition, the limited resources of the hardware platform and the required security level for OTA updates have to be considered throughout the system. To be a viable product, security requirements have to be met, while still allowing the operation with limited energy and memory requirements. Therefore, the platform requires both Security- and Energy-Awareness by Design.

For the management protocol, there was no established solution for constrained devices yet. Lightweight Machine to Machine (LWM2M) was evaluated[SA19], but could not be used due to resource requirements. Instead of a dedicated management protocol with the associated overhead, an update functionality is triggered over the Constrained Application Protocol (CoAP). A similar approach was also followed at the Software Updates for Internet of Things (SUIT)[So] IETF Taskforce which started work shorty before the beginning of the BASE MoVE project. The transmission of the image with parsing information in the form of a SUIT manifest should also enable protocol-independent transmission.

3.1 Multiprotocol

In the smart home sector, no consolidation of protocols is noticeable at this moment. On the contrary, new protocols are proposed, like the relatively new protocol „Thread“ developed by Google [Th]. To be able to demonstrate the multi-protocol approach of the BASE MoVE platform, a qualified subset of protocols had to be selected. This selection is differentiated between transport protocols, which distribute data between nodes (summarised OSI layers 1 - 6) and application protocols, defining the structure and semantics of the data exchanged (OSI layer 7). The requirements for this selection are: they have to be usable in smart home scenarios, support low power operations (potential utilisable with energy harvesting) and state-of-the-art security features. Furthermore, there have to be open source protocol stacks available, which can be used on the selected Mighty Gecko SoC hardware platform. For the application protocols, there should be standardised device profiles available enabling interoperability and connectivity between smart home devices.

The first selected group of protocols satisfying these requirements are IP-based. The 6LoWPAN protocol is the IETF standard to run „IPv6 over Low -Power Wireless Personal Area Networks“ [Mo07]. It is set on top of the IEEE 802.15.4 protocol, like ZigBee, and allows transparent mapping to standard IPv6 networks. There are several protocol stacks available, provided by different IoT OS. Using RIOT OS, its native 6LoWPAN stack is used. Additionally, the „Thread“ protocol is part of this group. Originally developed by Google Nest, it is now supported by different software and hardware vendors [Th]. Thread is based on 6LoWPAN, but replaces some parts and adds functionality, especially regarding security, deployment and routing. For this IoT platform, the open source stack „OpenThread“ [Op] is used, as there is a port to RIOT OS available. As second, different type of protocol BLE is selected. It is widely used in practice in many different smart home products, and virtually every smartphone offers connectivity and support. Furthermore, BLE Mesh, as a new standard extension, is going to provide the necessary mesh routing for more complex smart

home scenarios in the future. For this IoT platform, the open source stack nimBLE from the Apache mynewt project is utilised, which was integrated by the RIOT OS community by the end of the BASE MoVE project. The selected application protocols are applied on top of the IP-based transport protocols. Again, there are several possibilities (see [SJ17]), but in this work, focus was on CoAP, another IETF standard. It was used for the application as well as the management part, because it is relatively lightweight and uses UDP. It can be integrated in edge and regular IT networks and there are several smart home and management protocols utilising CoAP.

For the support of application profiles based on CoAP, two approaches were incorporated. First, probably the most popular approach in practice is to define something new for a particular use case. As a first quick solution, the sensor data provided by the concrete devices were manually mapped to a REST structure and served by the RIOT OS CoAP implementation. The obvious limitation of the approach is that compatibility and interoperability are limited to the CoAP layer. As a second approach, a subset of Dotdot [Do] was prototypically implemented, which is a ZigBee Alliance standard mapping the well-established ZigBee Cluster Library (ZCL) and the ZigBee Device Profiles (ZDP) to a CoAP REST interface. Since the official specification of Dotdot was not openly accessible for most of the project time, the proof-of-concept realisation was based on information gathered from presentations, white papers and commercial implementations. It is limited to poll sensor data from end devices over CoAP and 6LoWPAN or the Thread protocol.

Regarding BLE, the application layer is part of the standard itself (Generic Attribute Profile - GATT). The Bluetooth Special Interest Group (SIG) defines the structure and semantics of data as Characteristics, the specific behaviour of a device functionality as Services, which are composed to Profiles defining the functionality of a type of device [GOP12]. Since the BLE stack within RIOT OS was only usable at the very end of the project time, very simple GATT services were implemented for the hardware platform as a proof-of-concept.

4 Application Scenario

To evaluate the base platform, an apartment was retrofitted to test its functionality in exemplary scenarios. The apartment was provided by one of the supporting industry partners, Vonovia SE.

The usage of BLE was planned, with IP over BLE for the OTA functionality, but due to delays for BLE support in RIOT, 6LoWPAN was used as a replacement protocol in the apartment instead. The behaviour and logic of the scenarios consist of a set of rules, executed on a smart home service platform. As such, the OpenHAB platform was chosen, necessitating the implementation of protocol bindings for CoAP and the application semantics defined for this project.

For the application scenarios, an exemplary ambient assisted living scenario (S1) as well as a smart home scenario (S2) were realised using the devices presented in chapter 3. For both

scenarios, window sensors were attached to windows and doors reporting state changes (open/closed). In S1, a resident is warned when leaving the apartment if the windows are still open. So, if any window and the front door is open, a red warning light is switched on, and the forgotten window is displayed on a map next to the door. In S2, the room temperature is controlled depending on the resident's preferences, and the room lights are switched on or off depending on the room being occupied. Residents are using the room control unit to set a desired temperature, and, if the reported temperature from the window contacts surpasses this thresh-hold, an actor turns on a ventilator.

For a practical evaluation, these scenarios were implemented in three different rooms of the apartment. The first two rooms realised the scenarios with Thread and 6LoWPAN, respectively, to demonstrate the functionality. In the third room, the adaptability test case with FOTA functionality was evaluated, which needed two parallel OpenHAB deployments like in the first two rooms, each in a different IoT network and running a placeholder application. The devices were updated with firmware images, supporting the two different protocols, rotating them between the three applications.

5 Conclusion

The technical evaluation of the application scenarios has shown that the modular platform approach is feasible. The modular platform approach reduced the implementation effort for each new type of smart home device. In the end, the primary effort was on the hardware part. For the software, only drivers for new peripherals parts, some configuration and the additional application functionality were needed, which could largely be generated from ZigBee device profiles. Replacing an IoT protocol with another by firmware update was straightforward. The driver and protocol abstraction of the IoT OS allowed an easy creation of application images with new protocols. The issues encountered often stemmed from immature implementations. The early adaptation of Thread and DotDot, before the specification was available, provided some challenges. The implementation used, OpenThread, still had memory leak and energy consumption issues. While RIOT fulfilled the requirements set for this research project and offered great collaboration with the developer community, usage in a production environment would be premature. However, over the course of the project, clear improvements could be observed. The overall neglect of aspects regarding energy management throughout the design and the implementation, though, made the application in this scenario difficult. As a lesson learned, an adequate IoT-OS requires a holistic "Energy Awareness by Design". This topic is further explored in a new research project.

Furthermore, on a conceptual level, the evaluation of the application scenario has demonstrated that retrofitting existing living spaces with future-proof sensor devices is possible. It shows that sensor devices can be designed to adapt to new technologies, preserving the investment made despite being uncertain which technology will end up becoming the industry standard. Functional problems in the form of power consumption and stability issues with updates over Thread were caused by the use of immature implementations. In

the apartment setup, an additional inhibitor for the general use of smart home scenarios was identified in the configuration time and complexity in OpenHAB. It is unlikely that the average end-user, facility manager or electrician installing the devices could implement the apartment-specific rules. Improvements in smart home service platforms are thus still needed for widespread adoption.

Bibliography

[Ba18] Baccelli, Emmanuel; Gundogan, Cenk; Hahm, Oliver; Kietzmann, Peter; Lenders, Martine S.; Petersen, Hauke; Schleiser, Kaspar; Schmidt, Thomas C.; Wahlisch, Matthias: RIOT: An Open Source Operating System for Low-End Embedded Devices in the IoT. IEEE Internet of Things Journal, 5(6):4428–4440, 2018.

[Di19] Dizdarević, Jasenka; Carpio, Francisco; Jukan, Admela; Masip-Bruin, Xavi: A survey of communication protocols for internet of things and related challenges of fog and cloud computing integration. ACM Computing Surveys, 51(6):1–29, 2019.

[Do] Dotdot, https://zigbeealliance.org/solution/dotdot/.

[GOP12] Gomez, Carles; Oller, Joaquim; Paradells, Josep: Overview and evaluation of bluetooth low energy: An emerging low-power wireless technology. Sensors (Switzerland), 12(9):11734–11753, 2012.

[Mo07] Montenegro, G.; Kushalnagar, N.; Hui, J.; Culler, D.: Transmission of IPv6 Packets over IEEE 802.15.4 Networks. RFC 4944, September 2007.

[Ni15] Nieminen, J.; Savolainen, T.; Isomaki, M.; Patil, B.; Shelby, Z.; Gomez, C.: IPv6 over BLUETOOTH(R) Low Energy. RFC 7668, October 2015.

[Op] OpenThread, https://openthread.io/.

[Qu18] Qutqut, M. H.; Al-Sakran, A.; Almasalha, F.; Hassanein, H. S.: Comprehensive survey of the IoT open-source OSs. IET Wireless Sensor Systems, 8(6):323–339, 2018.

[SA19] Schmelzer, Paul; Akelbein, Jens-Peter: Evaluation of Hardware Requirements for Device Management of Constrained Nodes Based on the LWM2M Standard. In: CERC. 2019.

[Sia] Silicon Labs - Dynamic Multiprotocol, https://silabs.com/wireless/multiprotocol.

[Sib] Silicon Labs EFR32 Wireless Gecko Technology Features, https://www.silabs.com/products/wireless/technology.

[SJ17] Salman, Tara; Jain, Raj: A Survey of Protocols and Standards for Internet of Things. Advanced Computing and Communications, 1(1), March 2017.

[So] Software Updates for Internet of Things (suit), https://datatracker.ietf.org/wg/suit/about/.

[Th] Thread Stack Fundamentals, https://threadgroup.org/ourresources.

[Za18] Zaidan, A. A.; Zaidan, B. B.; Qahtan, M. Y.; Albahri, O. S.; Albahri, A. S.; Alaa, Mussab; Jumaah, F. M.; Talal, Mohammed; Tan, K. L.; Shir, W. L.; Lim, C. K.: A survey on communication components for IoT-based technologies in smart homes. Telecommunication Systems, 69(1), 2018.

R. Reussner, A. Koziolek, R. Heinrich (Hrsg.): INFORMATIK 2020,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2021 939

Modelling of Interaction in Interweaving Systems as Ontology Mapping Adaption

Matthias Jurisch[1], Bodo Igler[1]

Abstract: Interweaving Systems (IWS) are systems that have not been designed to cooperate, but can influence each other at runtine through a shared environment. In this paper, we present a general approach that enables IWS to gain an explicit view of the shared environment. The approach is based on ontology alignment and model driven techniques. The core idea is to build a *Directory Ontology* that shows what systems can know about their environment, and local *Connector Ontologies* that organize information routing between systems and connect these via ontology alignments. Changes are addressed using ontology mapping adaption. This approach is presented using an application example from the area of traffic control systems. The key benefit of the approach is that it allows supporting and sustaining data integration in Interweaving Systems using explicit and domain-independent rules.

Keywords: Interweaving Systems; Ontology-Driven Development; Ontology Mapping Adaption

1 Introduction

Interweaving Systems [To16] are systems that have not been designed to cooperate, but can influence each other at runtime by interacting through a shared environment. Interweaving systems work under conditions that require some kind of soft or hard real-time constraints. Therefore, they are first and foremost optimized to satisfy these constraints. Optimization for domain-specific goals is usually regarded as less important. However, deliberate cooperation between interweaving systems can potentially improve these soft aspects without touching the real-time/safety critical part.

Whether and how a shared view of the environments of Interweaving Systems improves cooperation and performance regarding domain-specific goals is still an open question. A first step to address this question consists of designing and evaluating approaches for environment sharing. An approach that allows sharing of data between interwoven systems would change the nature of interactions from accidentally interwoven systems to deliberate interactions supported by a central authoritative knowledge, which might also support the systems in dealing with emergent stability issues. As the typical environments of Interweaving Systems are prone to changes and as parts of such systems can fail, appropriate approaches must consider these issues, too.

[1] Department of Design – Computer Science – Media, RheinMain University of Applied Sciences, Unter den Eichen 5, 65195 Wiesbaden, {matthias.jurisch,bodo.igler}@hs-rm.de

doi:10.18420/inf2020_86

In this work, we present a general approach to support Interweaving Systems in creating a shared view on their environment. It employs knowledge based techniques such as ontologies and inference. To allow systems to access the shared view, a modification of these systems is required. What kind of information a system can gather about the environment (e.g., what sensors it can use to observe it) is modeled via a *Directory Ontology*. Each system can view the Directory Ontology and use it to request more information regarding specific aspects of the environment. In this way, systems will only receive data concerning the environment that is relevant to them. What kind of information is relevant to a system is modeled in a local *Connector Ontology* that is connected to the Directory Ontology using ontology alignments. Changes in the environment and system failures require that the alignment is constantly adapted. This issue is addressed with a technique called Ontology Mapping Adaption [Gr13a]. The key benefit of the approach is that it allows supporting and sustaining data integration in Interweaving Systems using explicit and mostly domain-independent rules.

In our previous work, we presented a domain-specific technique that addresses this problem for interweaving systems in the domain of autonomous traffic-control [JI17]. The main contribution of this paper is twofold: (1) We present an abstract, i.e., domain-independent approach that shows how the domain-specific technique can be applied to interweaving systems in general. (2) We show how the case study of [JI17] fits into the application of the abstract approach. This includes an evaluation of how the approach needs to be tailored to the specific use case and which model transformations need to be implemented. We also demonstrate, how autonomous systems can use this data in a domain specific use case and how the shared view on the environment can be useful.

The remainder of this work is structured as follows: Section 2 gives an overview of the background and related work and identifies the research gap. An overview of the approach and the models that are involved and how these models are used to create a shared view is shown in Section 3. Section 4 describes an application and shows how the approach can be tailored to domain-specific problems. A discussion is given in Section 5. The paper ends with a conclusion and an outlook to future work in Section 6.

2 Background and Related Work

Interweaving Systems are systems that are designed in some kind of technical context that is characterized by several aspects: The systems in this context influence each other, both via *defined interactions* and *not defined interactions*. Also, the systems are heterogeneous in a sense that no central instance controls all of them and they were not designed to take system's influences into account. The environment of the systems in general is uncertain, but can be partly observed. This has severe consequences for their effectiveness, as determining the outcome of interactions becomes very difficult. Another important aspect is that the application domain requires the systems to operate under real-time conditions. [To16]

Interweaving Systems are often supported using so called organic computing approaches [MSSU11] and often focus on making systems themselves more fault-tolerant and more able to deal with changing environments. Organic computing techniques draw inspiration from biological systems and are commonly used in the area of self-managing systems. In this work, we approach the area of interweaving systems from a different perspective: We focus on allowing Interweaving Systems to communicate about their environment and making this environment explicit. The messages exchanged in this model contain information on what data is available. This communication is supported by formal conceptual models (ontologies). This leads to several areas of related work, including (1) Model Driven Software Engineering for communication middleware, since our approach can be regarded as belonging to this area (2) ontology-driven software engineering and (3) ontology alignment adaption.

In the area of Model Driven Software Engineering for communication middleware, several works have shown that model driven software development can be used to facilitate the usage of middleware by using models to represent properties of data and middleware aspects [Ed04, BT10]. While these approaches show that a model driven approach can significantly reduce the lines of code needed to be written, the models and the modeling languages are domain-specific and do not allow automated reasoning, which makes them less flexible. This shortcoming leads to the area of ontology-driven software engineering (2) that uses ontologies as a modeling language for model driven software engineering. One of the benefits of using ontologies in this context is that ontologies can be easily linked to other ontologies and the declaration of these links allows for an automatic generation of model transformations [Pa12]. While approaches from this area can be used to model systems working in a real-time domain [St17], these models are static and do not account for changes in the environment or aligned ontologies.

Reacting to changes in aligned ontologies is an area of research that has been approached using several methods: Both applying rules to changes directly [Gr13b, do15] as well as applying rules to changing inferences [Ju16] have been proposed. In addition to these rule-based approaches, machine learning approaches have been explored [JI19]. While a rule-based approach has been proposed for a use case in the area of Interweaving Systems, this approach is tied to the domain-specific aspects of the use case [JI17]. To our knowledge, no domain-independent approach exists which facilitates environment sharing based on ontologies. This issue is at the core of the research presented in this paper.

3 Approach

The main goal of the approach presented in this work is to allow interweaving systems to cooperatively create a shared view of their environment. This approach needs to fulfill a set of non-functional requirements: (1) it needs to be fault-tolerant in the sense that it can still support a shared view on the environment when some systems fail, and (2) it needs to be fault-tolerant in the sense that it allows dealing with partial communication failures, and (3) it must not interfere with real time properties of the interweaving systems. Such an approach

is beneficial for IWS which observe their environment and whose performance (including functional aspects) can be improved by sharing these observations. As environment sharing requires additional computational resources, the systems also need some kind of isolation between real-time critical aspects and the program that implements our approaches, for example implemented by a mixed-criticality system [BD13].

The core aspects of our approach consist of three main ideas: (1) An architecture for supporting data sharing in an interweaving systems context, (2) an implementation strategy for this model based on ontologies and (3) an algorithm that is used to adapt the ontology model to changes. This paper focuses on the first two ideas. Although the approach has to be adapted to the respective domain specifics, the core ideas are domain-independent. These ideas are discussed in the following subsections.

3.1 Architecture

A simplified example of an application of the framework with two systems is shown in Figure 1. For the sake of simplicity, ontologies are represented as storages, while they can also be implemented as a view on concepts that exist in multiple other storages. Pre-existing components of interweaving systems are shown in yellow, our additions to these concepts is depicted in blue with dashed lines. In this paper, we will focus on the Connector Ontology, the Directory Ontology and the reaction to changes. The models used in the framework for knowledge sharing mainly need to fulfill two purposes: (1) let systems discover what information on the environment is available (2) bind information from remote locations to local needs. The first purpose is implemented using a model called directory ontology. The Directory Ontology contains information on two aspects: It represents which interweaving systems are able to exchange data. Each system can provide a set of sensors that can be used to observe the environment. These sensors are also represented in the Directory Ontology.

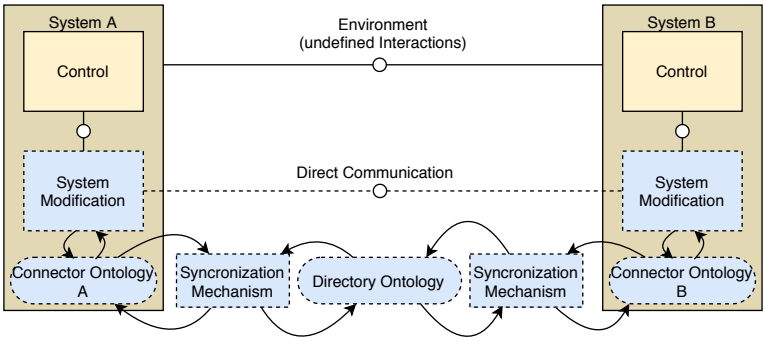


Fig. 1: Approach to Augmentation of IWS

The second purpose is fulfilled by a Connector Ontology for each system. The connector ontology, which is specific to each system, can label a sensor as local, or as a remote. Remote

sensors are deployed on other systems. For remote sensors, connection information is stored. An example for connection information can be a DDS topic ID or a URL. In the Directory Ontology, sensors can be related to data needs that are used to represent the requirements of data to improve the self-management capabilities of a system. The Connector Ontology also contains a Similarity Ontology that stores similarity scores. These scores are used to associate remote sensors to other stand-in sensors which can replace them in case of failing remote sensors. (This includes failure of other systems and communication failures, which lead to missing remote sensor data.) Pairs of sensors are assigned a similarity score that represents how well readings from a sensor are suited to replace readings from the other sensor.

A synchronization mechanism is used to integrate the Directory Ontology with local Connector Ontologies. This synchronization mechanism needs to fulfill two purposes: (1) ensure that the directory contains the most up-to-date information on sensors available on a system and (2) allow each system to access what sensors are available on other systems. Hence, this mechanism can be implemented by systems broadcasting their sensor capabilities in a predetermined interval. Based on this information, local systems can interact with each other to exchange data according to the definitions used in local and global ontologies. This information exchange can be implemented using a peer-to-peer framework such as DDS [Pa03] or any topic-oriented middleware that does not rely on a central broker. The same middleware can also be used to exchange Directory Ontology information.

3.2 Ontology Design

All models and, as far as possible, the change mechanism should be implemented using standardised ontology languages, so that standard tools can be used to process the models. Wherever possible, only subclass reasoning should be used. Subclass-based reasoning requires only low computational complexity. While the usage of subclass reasoning limits the expressiveness of the ontologies, the expressiveness suffices for the model presented in this paper.

These design guidelines lead to implementation strategies for the Directory and Connector Ontology. An example for these ontology types is shown in Figure 2. The class *System* models all sensor readings. Therefore, if a sensor is part of a system, the class representing the sensor is a subclass of the class representing the respective system. A sensor reading would be an instance in this ontology. The connector part of the ontology represents the *Need* concepts and shows what sensors are *External*. Every sensor that is a subclass of External is a sensor at a remote system that provides data to the local system. All connections between the connector and Directory Ontology are alignment connections. The Similarity Ontology part of the Connector Ontology, which is shown at the bottom of Figure 2, is based on *Similarity Entries* that relate pairs of local and remote sensors to similarity scores. Due to design choices regarding the Connector Ontology, Connector Ontology classes need to be represented as instances in the Similarity Ontology.

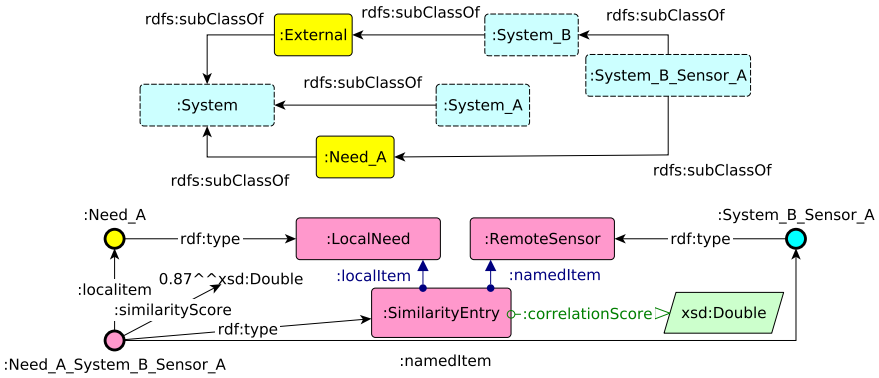


Fig. 2: Example Connector Ontology (yellow) with mapped Directory Ontology classes (blue with dashed lines) including Similarity Ontology (bottom)

3.3 Adapting to Changes

To adapt the mapping between Directory and Connector Ontologies when the Directory Ontology changes, we use a rule-based technique published in [Ju16, JI17]. The basic idea of this approach is to compute inferences before and after a change occurs. A comparison of the inferences leads to a set of deleted and added inferences named Δ_{inf}. Then two kinds of rules are applied to entries in Δ_{inf}: (1) rules that identify whether a change in the ontologies requires an adaption of the alignment and (2) rules that compute the actual changes to the alignment.

The first kind of rules can be implemented by simply searching for changes in inferences that are alignment statements. For each changed alignment statement, replacements need to be computed. This is the task of the second kind of rules: if an alignment statement needs to be adapted, the second kind of rule searches the Similarity Ontology for replacements using SPARQL queries. These replacements are then applied to the ontology alignment.

4 Transformation to a Domain-Specific Example

This section outlines how the general concept can be transferred to a concrete application in traffic control systems. A more detailed description of this application was published in [JI17]. In the example, traffic control systems controlling the traffic lights at one intersection have a fixed set of signaling plans that they can change based on traffic flow data. The approach is used to exchange traffic flow data between intersections, so that traffic control systems can incorporate traffic flow data measured at remote intersections into their decision making process.

Applying the approach requires some domain-specific artifacts to be created: change rules, the modification of the IWS and the incorporation of environment data into the traffic control process are always domain specific and need to be created for each use case. Other aspects are transformed in the sense of model driven software engineering. The specialization transformation is used to add the required domain-specific information to the abstract ontology model. In this case, the class *System* is used to represent traffic control systems. The class *Data* needs to be specialized so that it can hold data for traffic flow. Hence, a datatype is added to the class. *Data Needs* are used to represent data needs specific to intersection entries and exits. The remainder of the model remains unchanged. As described, when reacting to changes, rules are used to search Δ_{inf} for entries with instances of the class *External*. In this use case, a new alignment statement is generated so that the best candidate for a data need can be used. This process is based on a SPARQL query, that explicitly represents the change rule.

5 Discussion

The core contribution of this work is to present an approach that shows how to build a data structure that can use mapping adaption to deal with changes in the environment of interweaving systems. The approach supports the separation of concerns on several levels: domain-specific concepts are separated from domain-independent ones, and creating a shared model of the environment is separated from the regular functionality of the systems. Also, rules for changes, domain-specific rules and rules for connecting systems are separated. In addition, the approach allows for rules concerning the systems to be formulated in an explicit, short and domain-independent way. This also means that these rules are not hidden in the source-code of the system or as a part of assumptions of a domain-specific language.

6 Conclusion and Outlook

This paper shows, how an ontology-alignment-based approach can support data integration in interweaving systems and sustain integration when changes in the environment occur. Using these technologies allows a modularization of several aspects that are important to the scenario and an explicit formulation using standardized, domain-independent modeling languages. Future work could explore more application areas for the approach such as the industrial manufacturing domain. Also, the quantitative benefit of the approach could be evaluated and compared to other methods of systems management.

Bibliography

[BD13] Burns, Alan; Davis, Robert: Mixed criticality systems-a review. Department of Computer Science, University of York, Tech. Rep, pp. 1–69, 2013.

[BT10] Beckmann, Kai; Thoss, Marcus: A model-driven software development approach using OMG DDS for wireless sensor networks. In: IFIP International Workshop on Software Technolgies for Embedded and Ubiquitous Systems. Springer, pp. 95–106, 2010.

[do15] dos Reis, Julio Cesar; Pruski, Cédric; Silveira, Marcos Da; Reynaud-Delaître, Chantal: DyKOSMap: A framework for mapping adaptation between biomedical knowledge organization systems. Journal of Biomedical Informatics, 55:153 – 173, 2015.

[Ed04] Edwards, George; Deng, Gan; Schmidt, Douglas C.; Gokhale, Aniruddha; Natarajan, Bala: Model-Driven Configuration and Deployment of Component Middleware Publish/-Subscribe Services. In (Karsai, Gabor; Visser, Eelco, eds): Generative Programming and Component Engineering. Springer, Berlin, Heidelberg, pp. 337–360, 2004.

[Gr13a] Groß, Anika; Dos Reis, Julio Cesar; Hartung, Michael; Pruski, Cédric; Rahm, Erhard: Semi-automatic Adaptation of Mappings between Life Science Ontologies. In (Baker, Christopher J. O.; Butler, Greg; Jurisica, Igor, eds): Data Integration in the Life Sciences. Springer, Berlin, Heidelberg, pp. 90–104, 2013.

[Gr13b] Groß, Anika; Dos Reis, Julio Cesar; Hartung, Michael; Pruski, Cédric; Rahm, Erhard: Semi-automatic Adaptation of Mappings between Life Science Ontologies. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 7970 LNBI:90–104, 2013.

[JI17] Jurisch, Matthias; Igler, Bodo: Knowledge-Based Self-Organization of Traffic Control Systems. In: 47. Jahrestagung der Gesellschaft für Informatik, Informatik 2017, Chemnitz, Germany, September 25-29, 2017. pp. 947–954, 2017.

[JI19] Jurisch, Matthias; Igler, Bodo: Graph-Convolution-Based Classification for Ontology Alignment Change Prediction. In (Mehwish Alam, Davide Buscaldi et al., ed.): Proceedings of the Workshop on Deep Learning for Knowledge Graphs (DL4KG2019) Co-located with ESWC 2019, Portoroz, Slovenia, June 2, 2019. CEUR-WS.org, pp. 11–20, 2019.

[Ju16] Jurisch, Matthias: Managing Ontology Mapping Change based on Changing Inference Sets. In: Knowledge Engineering and Knowledge Management - EKAW 2016 Satellite Events. Springer, pp. 255–262, November 2016.

[MSSU11] Müller-Schloer, Christian; Schmeck, Hartmut; Ungerer, Theo: Organic Computing - A Paradigm Shift for Complex Systems. Springer-Verlag, Berlin, Heidelberg, 2011.

[Pa03] Pardo-Castellote, G.: OMG Data-Distribution Service: architectural overview. In: 23rd International Conference on Distributed Computing Systems Workshops, 2003. Proceedings. pp. 200–206, May 2003.

[Pa12] Pan, Jeff Z.; Staab, Steffen; Amann, Uwe; Ebert, Jürgen; Zhao, Yuting: Ontology-Driven Software Development. Springer Publishing Company, Incorporated, 2012.

[St17] Steinmetz, Charles; Schroeder, Greyce; dos Santos Roque, Alexandre; Pereira, Carlos Eduardo; Wagner, Carolin; Saalmann, Philipp; Hellingrath, Bernd: Ontology-driven IoT code generation for FIWARE. In: 2017 IEEE 15th International Conference on Industrial Informatics (INDIN). IEEE, pp. 38–43, 2017.

[To16] Tomforde, S.; Rudolph, S.; Bellman, K.; Würtz, R.: An Organic Computing Perspective on Self-Improving System Interweaving at Runtime. In: 2016 IEEE International Conference on Autonomic Computing (ICAC). pp. 276–284, July 2016.

A Tangible Object for General Purposes in Mobile Augmented Reality Applications

Linda Rau[1], Robin Horst[1], Yu Liu[1], Ralf Dörner[1], Ulrike Spierling[1]

Abstract: Smartphones and tablets are common technologies within today's private living environments. They are well-suited to serve as a platform for mobile Augmented Reality (AR). Tangible AR is a subclass of AR which includes tangible objects and can make interactions intuitive. With this, new options for human-computer interaction become available at home. Based on literature research and design rationale, we identify requirements that help to develop a tangible object which can intuitively be used as tangible user interface (TUI) for mobile AR applications. Users should be able to handle the tangible object comfortably. Additionally, it needs to be reliably trackable with today's tracking algorithms. The tangible object should also offer affordances to the users. We strive to develop a single, versatile object that is usable in different application scenarios at home. Our approach is to design a tangible object that combines different surfaces and shapes to offer various affordances and interaction possibilities. A physical instance of this object can be created with a 3D printer. We argue that this allows users to trigger actions intuitively in an AR environment or to manipulate virtual content.

Keywords: Tangible Augmented Reality; Mixed Reality; Tangible User Interface

1 Introduction

Current smartphones and tablets are suitable devices to serve as mobile Augmented Reality (AR) platform because they are equipped with a camera and have sufficient computing power. Furthermore, AR supporting software can be preinstalled with the operating system, e.g., ARCore on Android or ARKit on iOs. The wide distribution of such devices in private life offers new options for human-computer interactions that are available at home. With this, the number of available mobile AR applications is also increasing. Compared to traditional 2D interfaces, mobile AR applications offer other possibilities for user interactions. One category for such user interactions is physical interaction. Literature such as [BKP08] shows that the intuitiveness of user interactions can be improved when a tangible real-world object serves as user interface. Using this object, users can manipulate objects in their real-world to manipulate virtual content. AR experiences can be improved with the interaction techniques that tangible objects allow, because they can offer more affordances than solely virtual interfaces [Bu99]. The notion of affordance is discussed as properties of objects which invite for specific actions by Gibson [Gi77], e.g., a button suggests pressing or a door handle

[1] RheinMain University of Applied Sciences, Faculty of Design Computer Science Media,
Unter den Eichen 5, 65195 Wiesbaden, Germany. E-mail: firstname.lastname@hs-rm.de

doi:10.18420/inf2020_87

suggests levering it. Therefore, tangible objects could help users to have an intuitive tangible AR experience without frustration resulting from a non-intuitive interface.

In general, a large variety of objects can be utilized as tangible objects for a tangible user interface (TUI) in an AR application. Images or paper cards are commonly used as image targets and are popular for AR applications for private use. One 3D tangible object that is used for AR applications in private life is a cube, e.g., the merge cube [Me20]. While a cube can be a suitable object for viewing virtual 3D content, other objects could suggest more or other affordances. Thus, tangible objects providing affordances can help users to understand interaction possibilities for a mobile AR application without or with a short learning phase beforehand.

Our contributions in this paper are the following. Based on literature research and the elaboration of a design rationale, we identify and explore requirements for a tangible object that can be used as TUI for AR applications. We propose a generic tangible object that can be used in diverse AR application scenarios. It provides several geometrical forms which offer different affordances. With a 3D printer, we create a physical instance from our digital model and discuss which of its properties are conforming to our proposed requirements.

This paper is organized as follows. In the next section, we present related work. Following this, we identify and describe our three requirements. We applied our findings and designed our tangible object, which we introduce in section 4 and evaluate in the fifth section. We draw conclusions in section 6 and share our thoughts on future work.

2 Related Work

Early ideas for TUIs have been discussed in [WMG93]. Following this, Fitzmaurice et al. [FIB95] introduced their concept of graspable interfaces, where they use objects, e.g., handles, to manipulate digital content. These inspired the Tangible Media Group [IU97] for their vision of tangible bits, where the users' physical surrounding becomes an interface itself as objects or surfaces are linked to digital information.

Billinghurst et al. [BKP08] describe Tangible AR interfaces as an intuitive way to interact with an AR interface where users can manipulate a physical tangible object to manipulate a virtual object which is registered to the physical one. They describe design principles for TUIs which can help to develop an intuitive TUI for AR applications. They state that a Tangible AR interface that follows their design principles is intuitive to use and facilitates seamless display and interactions. However, they do not cover design principles for tangible objects but concentrate on functional requirements to the interface, e.g., support for multiple handed-interactions or multiple activities and objects. Furthermore, they propose four prototype AR applications with tangible interfaces: Shared Space, ARgroove, Tiles, and VOMAR. Planar images are used as tangible interfaces in Shared Space and ARgroove. For Tiles, the images have various shapes that are mapped to different semantics. The unique

functionality of each image target is similar to the unique functionality of each icon and tool on a computer desktop interface. In VOMAR, they use a trackable cardboard paddle as tangible interaction device.

Different tangible objects have been explored to manipulate AR content. For example, a trackable pen is used for MARS [Hö99] and Studierstube [SFH00]. Other tangible objects for AR applications can be a cup, which allows users to modify virtual object [Ka03], or a cube [Ji15; Me20].

The tangible AR applications and interfaces named above demonstrate that different shapes and objects can be conceivable as tangible interfaces. Some tangible objects offer affordances for specific interactions but are dependent on the use case and cannot be reused in another context, e.g., images or objects that represent one specific real-world object. Generic tangible objects can offer affordances for general purposes and hence be used in several use cases. However, for some use cases, their affordances are not sufficient. For example, a pen offers affordances for tasks like writing and selecting, but no affordances for examining 3D content. In contrast, a cube offers the affordance to examine 3D content but no affordances for writing.

3 Requirements for a Generic Tangible Object in Tangible AR

We identified three requirements for a tangible object that is supposed to be used as generic AR-based TUI. In contrast to tangible objects that are developed for one specific use case, a generic object can be used for several applications. Therefore, it can contribute to a simple TUI where no switching between several objects is necessary. In this section, we go into detail for the three requirements.

This first category is the object's attributes to make it reliably detectable and trackable with current tracking algorithms. This is helpful to provide a frustration-free and immersive user experience. For example, when immersed in an AR experience, the users can feel present in the AR to the point that virtual content is not perceived as virtual but as part of the physical world. Then, whenever the tracking is lost, this illusion can be disturbed or completely disrupted. Additionally, users need to wait and eventually need to put either the tracked object or the AR device in another position or to change the lighting conditions before the tangible object can be detected and tracked again which further disturbs the AR experience. Therefore, the tangible object needs to consist of reliably detectable and trackable textures or shapes and surfaces. This is dependent on the used tracking algorithms because not all tracking algorithms detect or track the same properties. Such properties of an object can be its shape, features in its texture, colors, or a combination of these. Several tracking algorithms make use of more than one property because this adds to tracking stability. For example, the tracking software Vuforia detects objects by shape, but additional information about the material, such as colors, significantly improves the robustness [Pa20]. Therefore, a reliably trackable object should ideally combine multiple well detectable and trackable

properties. When users hold the tangible object, they might occlude these properties in part or entirely, which can make the object hard to detect or contribute to a tracking loss. To provide enough detectable and trackable properties in despite of occlusion, each part or each side of the tangible object should be reliable trackable.

The second category we identified is the tangible objects' handling which includes its size, weight, and shape. A study by Sheridan et al. [Sh03] finds that tangible objects, in their case cubes, should naturally fit in the user's hand. We conclude this makes a recommended size for the tangible object dependent on the user's hand size. For example, a child can have smaller hands and require a smaller tangible object than an adult. A size of 8 cm × 8 cm × 8 cm is specified as suitable by Jimenez et al. [Ji15]. They use a webcam with a resolution of 2 megapixels and a focal length of 3.7 mm and find the resolution to be sufficient for target recognition with a distance of up to 1.5 m between object and camera. Furthermore, they state this is a mean value between sizes suggested by AR software developers. In their scenario, the user has both hands available and can choose to grab the object with one or both hands. However, in mobile AR users might be required to hold their AR device and therefore have only one hand left to use a tangible object as TUI. In this case, a smaller size can make the handling with one hand easier. Beside a tangible object's size, its weight influences its handling. Holding a tangible object can become exhausting for the user, especially if it is heavy. Therefore, we determined a tangible object should preferably be as light as possible, but yet durable, to make its usage less exhausting for users. A lightweight object is usually achieved without further effort due to a relatively small size. For example, a 3D printed cube of size 8 cm × 8 cm × 8 cm and with 25% infill (a common density setting) weighs about 150 grams. This value is dependent on the printing material, but common 3D printing materials have similar densities, except for metals. Furthermore, the tangible object's shape has an impact on its handling. The study by Sheridan et al. [Sh03] finds that an object's geometry affects how well users can grasp it. The authors state that curves in an object's geometry as well as a high surface area can enhance its grip. For example, they explain that a rhomboid or star-shaped object is easier to grasp than a cube. They also find that the object's material has an impact on its grip and that a flexible material can aid in grip.

We identified an object's affordances as third category. If specific properties of an object invite for certain actions, they offer affordances. This suggests that we can provide affordances for certain actions, that users can perform in an AR environment, by the tangible object's design. For instance, a hemisphere shape can be perceived as a button and therefore afford to push it. This indicates that various shapes can afford distinctive actions. Therefore, we suggest designing the tangible object consisting of various shapes, that each offer affordances for specific or several actions. General interactions, that a general tangible object could support and that can be useful in several AR applications, could be selecting, viewing 3D content, navigating, or scrolling.

4 Proposal of a Generic Tangible Object

Based on the requirements identified above, we designed an object that can serve as TUI for general purposes in AR applications and describe it in this section. The digital model of our tangible object is shown in Figure 1.

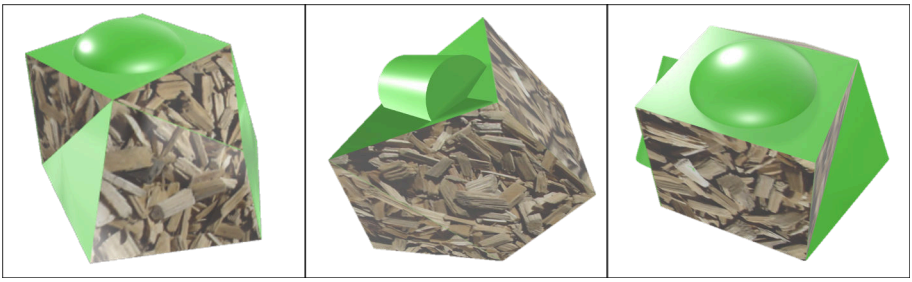


Fig. 1: Digital model of a tangible object for intuitively manipulating virtual content of an AR

We used Vuforia's image target example *chips* to create a texture for the tangible object. It was printed on adhesive film and glued to the object's surfaces. Because the image target is specifically designed to be detected and tracked, it provides a high number of trackable features. Besides this, trackable properties of our tangible object can be its edges and overall shape. The hemisphere and cylindrical shaped parts contribute to an asymmetric shape.

We use PLA material for 3D printing, which has a density of avg. 1.3 g/cm³. The PLA printing material is rigid and allows a firm grip without damaging or compressing the tangible object. Our object was printed at a size of 5.5 cm × 5.5 cm × 4 cm. The chosen infill density of 20% results in a weight of 23 g. The surfaces of our object include one pentagon, four quadrangles, and five triangles. The pentagon has a size of 5.5 cm × 5.5 cm at its widest place. The edges of two quadrangles are 2.8 cm, 2.1 cm, 3.6 cm, and 3.4 cm. A third quadrangle, which is a rectangle, has a size of 3.2 cm × 3.6 cm. From the fourth rectangle surface with size 3.6 cm × 3.0 cm, a hemisphere shaped part with 2.8 cm diameter and height 0.7 cm sticks out. Two of the triangles are isosceles. The first triangle has a base of 3.6 cm and a height of 2.0 cm while the second one has a base of 4.4 cm and a height of 2.2 cm. A third triangle measures 3.6 cm, 3.3 cm and 4.4 cm on its edges. A cylindrical shape (2.5 cm height and 0.7 cm diameter) sticks out from two further triangle surfaces which have the same sizes as the second and third triangle. This results in 28 edges, 16 vertices, and two round shapes. These various surfaces and shapes offer several affordances, depending on the use case. Our surfaces (triangles, rectangles, and one pentagon) can be augmented with different shaped images or virtual content. The cylindrical shape can suggest user interactions like scrolling, zooming, or turning over a page in a virtual book. Pushing a button or buzzer can be suggested with the hemisphere shape.

5 Evaluation and Discussion

We created a physical instance of our model using a 3D printer. We inspected how our proposed tangible object meets the requirements. Our results are described and discussed in this section.

Initially, the tangible object was printed with a 3D printer from a single material in one color. This provided few trackable features and the shapes' edges did not contrast well from the rest of the object. Therefore, the object could not be detected by the tracking software. To improve the object's tracking conditions, we applied a texture to both, the tangible object and its digital model. The digital model with UV mapped texture can be used as a reference for the tracking algorithm. However, gluing the texture on the tangible object is imprecise. With these differences between the physical object and the digital one, the tracking performance is insufficient. Another approach is to scan the physical object to use the scan data as a reference for the tracking algorithm. In this case, suitable lighting conditions must be provided during the scan process. With this approach, our tangible object can be tracked well while the lighting conditions are similar to the ones during the scan. The surface that the object rested on during the scan process is not included in the scan data. Therefore, it cannot be detected or tracked. A second scan from another side is required to make the tangible object reliably trackable from all sides. This process is sufficient but time-consuming.

Our 3D printed object is designed for one-handed interactions in mobile AR and therefore relatively small. Depending on the user's hand size, this can result in occlusion of a substantial number of trackable features or properties and therefore disturb the tracking quality. We find the object can comfortably be grabbed on its edges so that only a few parts are occluded. This is visualized in Figure 2.

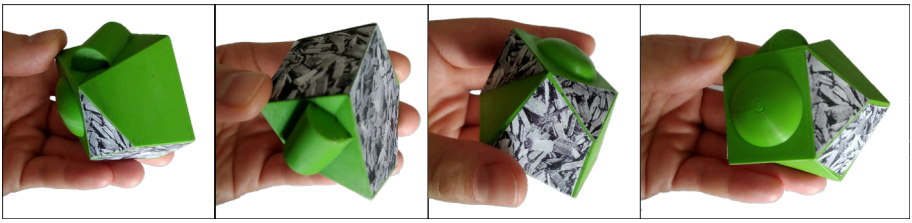


Fig. 2: Views of a tangible object for intuitively manipulating virtual content of an AR

6 Conclusion

With this work, we explored three requirements for intuitive tangible objects serving as TUI for mobile AR applications: tracking conditions, handling (weight, size, shapes), and affordances. We proposed an example for such a tangible object where we considered these. Based on our three requirements, our object is indicated to be suitable for intuitive interactions in an AR environment.

In future work, our proposed tangible object can be improved regarding its size and trackable properties or features. We found our size of 5.5 cm × 5.5 cm × 4.0 cm acceptable, but a greater size would result in less occlusion and therefore can contribute to a stable tracking. We applied a texture to our object to make it reliable trackable, but this made a time-consuming scanning task to set the object with texture as reference for the tracking software necessary. Future work can include developing 3D-printable textures that can easily be detected and tracked with current tracking algorithms.

Acknowledgments

This work has been funded (in part) by the German Federal Ministry of Education and Research (BMBF), funding program Forschung an Fachhochschulen, contract number 13FH181PX8.

References

[BKP08] Billinghurst, M.; Kato, H.; Poupyrev, I.: Tangible Augmented Reality. In: ACM SIGGRAPH ASIA 7. 2008.

[Bu99] Butz, A.; Hollerer, T.; Feiner, S.; MacIntyre, B.; Beshers, C.: Enveloping Users and Computers in a Collaborative 3D Augmented Reality. In: Proceedings 2nd IEEE and ACM International Workshop on Augmented Reality (IWAR'99). IEEE Comput. Soc, pp. 35–44, 20-21 Oct. 1999.

[FIB95] Fitzmaurice, G. W.; Ishii, H.; Buxton, W. A. S.: Bricks: Laying the Foundations for Graspable User Interfaces. In: Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '95. ACM Press, pp. 442–449, 1995.

[Gi77] Gibson, J. J.: The theory of affordances. In (Robert E Shaw, John Bransford, ed.): Perceiving, acting, and knowing: toward an ecological psychology. Hillsdale, N.J. : Lawrence Erlbaum Associates, p67–82, 1977.

[Hö99] Höllerer, T.; Feiner, S.; Terauchi, T.; Rashid, G.; Hallaway, D.: Exploring MARS: Developing Indoor and Outdoor User Interfaces to a Mobile Augmented Reality System. Computers & Graphics 23/6, pp. 779–785, 1999.

[IU97] Ishii, H.; Ullmer, B.: Tangible Bits. In: Proceedings of the SIGCHI conference on Human factors in computing systems. ACM Press, pp. 234–241, 1997.

[Ji15] Jiménez Fernández-Palacios, B.; Nex, F.; Rizzi, A.; Remondino, F.: ARCube-The Augmented Reality Cube for Archaeology. Archaeometry 57/, pp. 250–262, 2015.

[Ka03] Kato, H.; Tachibana, K.; Tanabe, M.; Nakajima, T.; Fukuda, Y.: MagicCup: A Tangible Interface for Virtual Objects Manipulation in Table-top Augmented Reality. In: ART 2003. IEEE, pp. 75–76, 2003.

[Me20] Merge Labs, Inc.: Merge Cube | AR / VR Lernen & Erstellen, accessed: 07.04.2020, URL: https://mergeedu.com/cube.

[Pa20] Parametric Technology GmbH: Model Targets Supported Objects & CAD Model Best Practices, accessed: 03.04.2020, URL: https://library.vuforia.com/content/vuforia-library/en/articles/Solution/model-targets-supported-objects.html.

[SFH00] Schmalstieg, D.; Fuhrmann, A.; Hesina, G.: Bridging Multiple User Interface Dimensions with Augmented Reality. In: Proceedings IEEE and ACM International Symposium on Augmented Reality (ISAR 2000). IEEE, pp. 20–29, 5-6 Oct. 2000.

[Sh03] Sheridan, J. G.; Short, B. W.; van Laerhoven, K.; Villar, N.; Kortuem, G.: Exploring Cube Affordance: Towards a Classification of Non-Verbal Dynamics of Physical Interfaces for Wearable Computing. In: IEE Eurowearable '03. IEE, pp. 113–118, 4-5 Sept. 2003.

[WMG93] Wellner, P.; Mackay, W.; Gold, R.: Back to the real world. Communications of the ACM 36/7, pp. 24–27, 1993.

Integration of Game Engine Based Mobile Augmented Reality Into a Learning Management System for Online Continuing Medical Education

Robin Horst[1], Dennis Fenchel[2], Reimond Retz[2], Linda Rau[1], Wilhelm Retz[1], Ralf Dörner[1]

Abstract: Physicians must participate in continuing medical education (CME) as part of the medical quality assurance. One possibility is to take online courses in their private living environment. These courses are mostly text- or video-based. Novel technologies such as mobile Augmented (AR) or mobile Virtual Reality (VR) are not yet established although their usage is not out of bounds in private homes anymore. Game engines can facilitate the authoring of applications that utilize VR/AR, as they provide many crucial functionalities out of the box. However, integrating the resulting VR/AR software in online CME courses is not trivial. In this paper, we investigate this integration into an existing learning management system (LMS) for online CME. In the example of a mobile AR application, we propose a system design that extends a course by a mobile AR part. We describe our implementation and how we transition users from their familiar web-interface on the desktop PC to a mobile AR application.

Keywords: Professional Education in Private Living Environments; Online Continuing Medical Education; System Design; Augmented Reality; Virtual Reality; Games Engineering; E-Learning

1 Introduction

Many professionals such as health professionals use their private living environments for continuous education in their field. Continuing medical education (CME) comprises training measures that serve to maintain and permanently update the professional competence of the medical profession. CME also serves as a part of medical quality assurance. These compulsory training activities for physicians are demanded by governmental-related organizations, such as the Accreditation Council for Continuing Medical Education (USA) or the Bundesärztekammer (Germany). Physicians are required to obtain a specific amount of credits, which can be earned through different activities. One of these activities is online training. Online CME is mostly conducted using technology that is already available in the private living environments of physicians, such as common desktop PCs, tablets, or smartphones. Different established media are used here to mediate information, for example, text, images, audio files, or videos. In the context of lifelong learning, private living environments are also expected to support modern technologies for learning, such as Augmented (AR) and Virtual Reality (VR) on mobile devices. Game engines such as Unity

[1] RheinMain University of Applied Sciences, Kurt-Schumacher-Ring 18, 65197 Wiesbaden, Germany, firstName.lastName@hs-rm.de
[2] health&media GmbH, Dolivostraße 9, 64293 Darmstadt, info@health-media.de

doi:10.18420/inf2020_88

[Un20] are suitable tools to support authors in creating applications for such technologies. However, integrating mobile AR and VR applications developed with game engines into existing systems for online CME and courses involves challenges for authors.

In this paper, we make the following contributions:

- We investigate the integration of mobile AR into an existing CME course that is based on a learning management system (LMS). Our goal is to extend a traditional online course with novel technology rather than replacing it entirely.
- In the example of a mobile AR application (app) that we built with Unity, we highlight crucial aspects of its integration in a CME course. Based on a prototype, we report and discuss lessons learned from the implementation process.
- Physicians usually participate in online courses on their desktop PC. To perceive the AR part, they must switch to our mobile AR app within the course. We propose a transitioning technique that guides users from the web-course to our mobile AR app.

This paper is organized as follows. The next section discusses related work. In Section 3, we describe the LMS that we integrate the mobile game engine technology into. In the fourth section, we present our system design that realizes the integration. Section 5 provides a conclusion and points out directions for future work.

2 Related Work

Web-based e-learning courses are widely used for CME. Various technologies can facilitate the authoring of such courses, including LMSs [CGB09]. There exists a variety of LMSs that serve different purposes. Moodle [DT03] (Modular Object-Oriented Dynamic Learning Environment) is an established multi-purpose and open source LMS. There are also more specific LMSs such as OLE [PDG20] (Open Learning Environment). OLE was developed to serve individual requirements of the local learning context of a university.

Different media can be integrated into LMS courses. Persia et al. [PDG20] show that courses using a variety of media highly support learning activities compared to solely text-based education. The user satisfaction was particularly improved when different multimedia were exploited. They also show how to integrate educational videos and smart text in their LMS courses but did not investigate AR or VR technology. Recent work by de Paiva Guimarães et al. [Pa17] propose a tool that utilizes educational browser-based AR content in the SCORM learning object standard. It creates AR learning objects and provides packaged AR applications to Moodle courses that can be executed utilizing a desktop PC and a webcam. It decouples LMSs and AR applications by distributing the applications through online repositories [Pa18]. Therefore, each AR application can be built with separate tools itself, as long as it is packaged with their tool in the SCORM standard. However, each time a course requires AR content, a novel AR application must be downloaded. With a shared

runtime environment, users only would have to download the content of an application. This facilitates the content delivery of AR applications within LMS courses.

Contrary to this approach by de Paiva Guimarães et al. [Pa17, Pa18], Coma-Tatay et al. [Co19] integrate the AR learning content without external applications or plug-ins. They provide a tool (FI-AR) that is based on the open-source software framework FIWARE [FI20]. It utilizes universal web-technology access to visualize AR content within their online courses. The AR content can be executed from both a desktop or a mobile AR runtime environment. Visualizing AR or VR content within an LMS environment brings the challenge of transitioning from traditional media to these immersive technologies. Dodd and Antonenko [DA12] propose to use signaling methodologies to facilitate the transition for users that enter the virtual worlds at the example of a desktop VR app. They provide users visual cues to guide them during this transition and use cues during the virtual learning activities until learners return to the traditional LMS content.

Generally, existing work on LMS and virtual technologies focuses on supporting separate custom-made AR and VR applications. Established authoring tools, such as game engines, are not considered. Furthermore, desktop PCs are targeted. Concepts for integrating AR and VR that runs on a mobile device within online courses are lacking.

3 Existing Continuing Medical Education System

Our existing *arztCME* [hG] e-learning framework for CME comprises both LMS and content management system (CMS) functionality aligned with specific requirements to comply with the certification of CME. The foundation of the CME system is a CMS based on WordPress that offers PHP and MySQL interfaces. It was extended with plugins to provide LMS functionality. The LMS features support authors to create CME courses and tests for assessing the learning success of physicians to give them credits. One CME course takes approximately 45 minutes and is consists of several 'pages' of learning content.

Physicians usually participate through web-browsers on a desktop PC. Some use mobile devices. In addition to static course representations as PDF documents, courses can be represented as multimedia HTML realization within our e-learning framework (Fig. 1). Established media in this realization are images, videos, and texts. Our web-technology can directly integrate them in the HTML environment.

4 Integration of a Mobile Unity Augmented Reality App

There exist ways to include AR and VR technologies directly in a web-browser (e.g., [Li04, La19]). As our online CME courses already take place in a web-environment, a naive integration of these technologies makes them available directly in the web-browser. Unity also offers such a solution. It provides a web-player which can run Unity applications directly

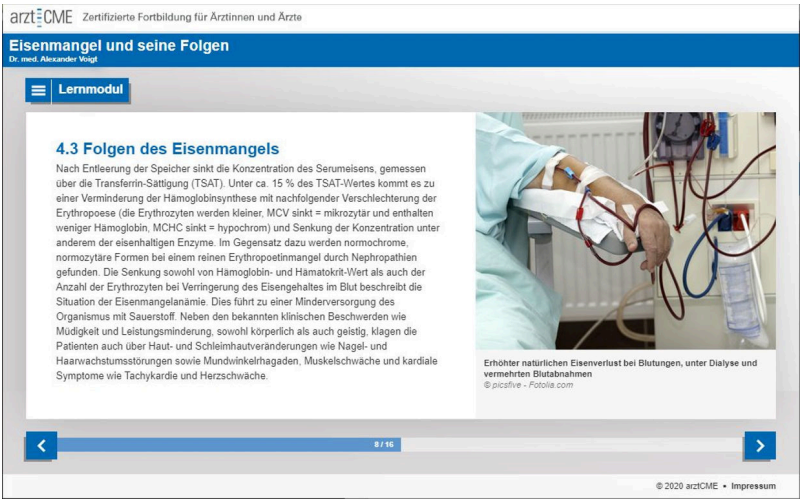


Fig. 1: The view on one page of a traditional online CME course for German-speaking physicians.

in the web-browser. However, this player and specific AR and VR features are designed for web-browsers that run on desktop PCs. Access to these players through web-browsers on mobile devices is possible but not entirely supported. The performance can suffer when using computationally intensive AR or VR functionalities through a mobile web-browser due to the complexity of additional rendering steps that the browser needs to go through [BRR16]. Running these technologies in dedicated apps on a mobile device can increase the performance.

To provide an AR experience at a specific point during the course, users are required to switch from the traditional CME course that runs on their desktop PC to our mobile app. To facilitate the transition between the two environments and to induce the user to do so, we integrated an easily accessible onboarding approach within our system design (Fig. 2). Using QR codes (Fig. 3), an established linking technique within the mobile domain, we can transfer users from a page within the course to the AR app. Furthermore, using dynamic QR codes, created with JavaScript at runtime, we are also able to include context-specific information during the transitional phase. This provides an opportunity to attach a dynamic session ID, that can be used in further steps of the onboarding process to retrieve user-specific data, such as the name of the course that the user is in and the current state within the course.

To proceed in the course, users scan the QR code with their mobile device. In case the users do not have the corresponding AR app already installed on their mobile device, they are redirected to the store app of their mobile operating system (OS) (e.g., Appstore for iOS and Google Play Store for Android). Here they can download our AR app. In case the app already exists on the mobile device, it is opened directly instead. We use Firebase Dynamic

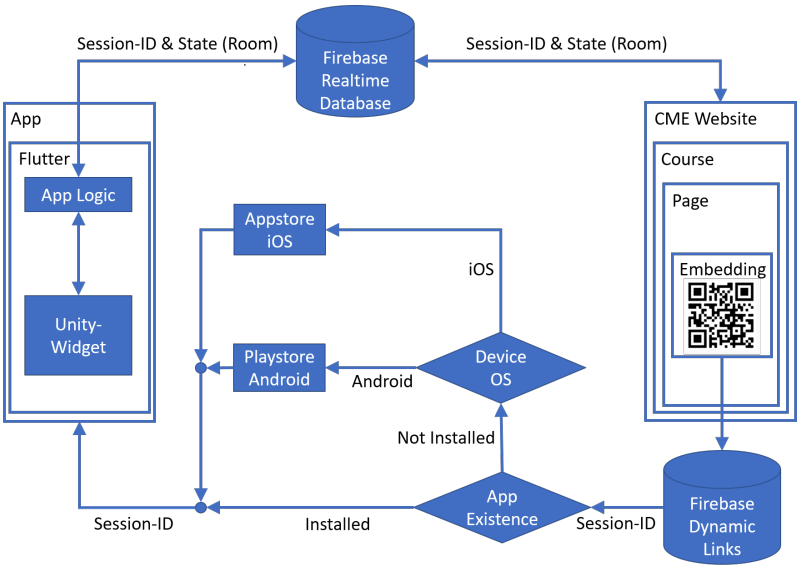


Fig. 2: Our system design extends an existing course on one specific page with mobile AR content. An onboarding process facilitates the transitioning from the website on a desktop PC to the mobile app.

Links [Go20a] to realize these re-directions. Firebase Dynamic Links (FDL) can store the session IDs from the QR code dynamically and transfer them through the installation process to our mobile app. The app can use the session ID to get further information about the user from a corresponding Firebase Realtime Database (FRTD) [Go20a] that is also connected to our arztCME LMS (Fig. 2). It uses an open web socket subscription model for the connection. Both the app and the website can access and manipulate entries in the database in realtime. A basic session management system and IDs that track the logged-in users were already included in the original LMS. This has been extended to include current AR content and the information required for it.

Now users can be directed from a course to a separate AR application for each course that extends it. However, it can be cumbersome for users when they have to download a dedicated AR app for each specific course they participate in. To counteract this, we decided to build a single app that serves as an AR platform for all courses. When using one application that provides the content for all available courses, it can still be tiresome for users to open the app and search the right content for their current course. We make use of the realtime connection of our app to the web-based LMS to open the app at a specific state directly after scanning the QR code instead of just starting the app. The session ID that we transferred throughout the onboarding process with FDLs can be used to query necessary information such as the specific location within a course. Alternatives without FDL would require scanning the QR code multiple times (e.g., at first for linking to the store or opening the app and then for opening the right content within the app).

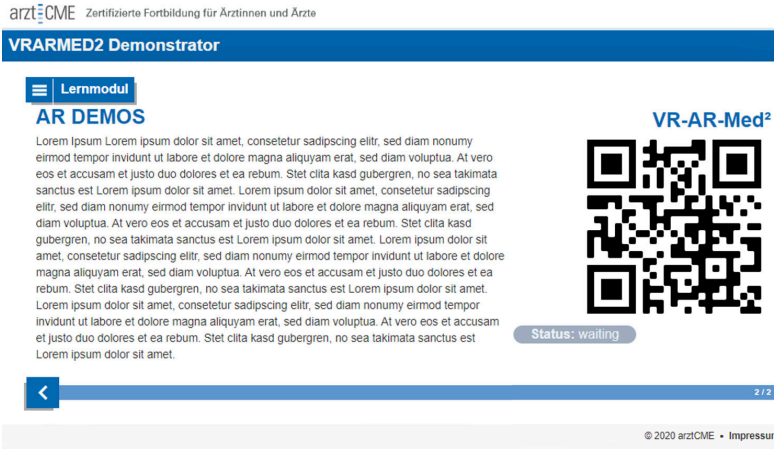


Fig. 3: A screenshot of a page within a course that includes a QR code used for transitioning between the desktop web-browser and our mobile game engine app.

The app development itself can be divided into two separate blocks. All AR-related functionality was developed in a Unity project, whereas all mobile-related functionalities (e.g., mobile app front end) were developed in a separate Flutter [Go20b] project. The advantage of using Flutter for mobile app development is that it supports cross-platform development. It also has the advantage that user interfaces (UIs) developed with Flutter look more conventional to UIs of regular mobile apps in contrast to mobile game UIs build with Unity. Furthermore, Flutter offers interfaces to a variety of common tools that facilitate mobile app development. However, splitting the development requires the integration of the Unity-based AR functionality within the Flutter app environment. This is a challenging task since Unity development for mobile devices will normally build separate APK or IPA files which cannot easily be integrated into a Flutter app. As of Unity version 2019.3, Unity allows to build a Unity project as a library to be included within other apps. In a Flutter-App, the embedded Unity library can directly be used as a Widget to display its content. This Unity Widget can be developed in the Unity authoring environment without further restrictions except one. Unity Widgets are not directly suitable for integration in a Flutter multi-platform code basis. For example, it is only supported to display AR Widgets in full-screen- mode. We also had to establish an asynchronous communication interface between the Flutter environment and the Unity Widget that is used as a plugin in Flutter. This communication is used to start the app directly with the content while correctly initializing the application's state. For example, separate Unity scenes can be addressed, or the Flutter app can tell the Unity Widget which image targets for the AR functionality to use. The Unity Widget also sends information through the Flutter environment back to the LMS. For example, this is used to share quiz results for later use in the evaluation of the CME course to award the credits for physicians when the course is successfully finished. The resulting app is illustrated in Fig. 4.

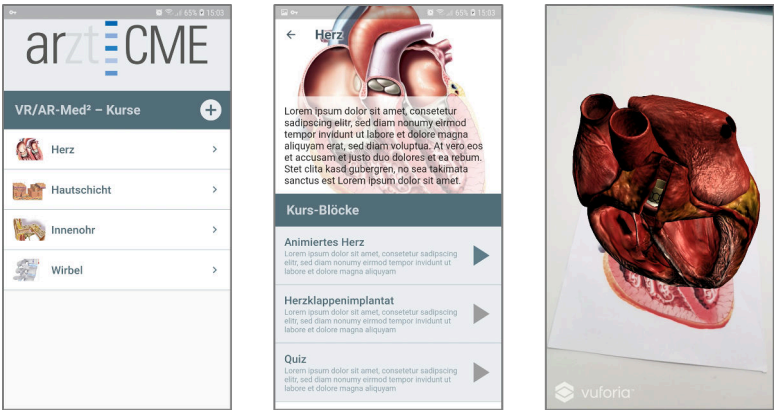


Fig. 4: Left: Start screen of our AR app. Users can select a course. Middle: A course about the functionality of the human heart is selected. Right: A heart displayed with AR technology.

5 Conclusion and Future Work

In this paper, we investigated the integration of mobile AR into an existing LMS for online CME. In the example of extending one of our courses with an AR functionality, we described how we realized this integration. We propose a system design and a suitable onboarding approach that facilitates the transition from the traditional online CME course on a desktop PC to its AR part on our mobile AR app.

Future work relating to our contribution can be divided into two categories. Now that the mobile AR can technically be integrated within our courses, a user test with physicians can be conducted to draw conclusions on the acceptance of such technologies within the CME domain. As these technologies are novel to CME, certification of such content must also be addressed and also privacy aspects of our onboarding approach must be considered. As for now, we do not know of any privacy issues regarding Google's Firebase technology. Secondly, we will explore how new AR content can be delivered to users efficiently. As of now, our app includes the AR content for all courses but when more and more courses get an AR extension, the size of our app will grow vastly. Content can be delivered on-demand. However, delivering novel content to a ready-built Unity application is not a trivial task. Furthermore, virtual assets such as 3D models can have large file-sizes so that downloading on demand may interrupt the learning flow of a course depending on the Internet connection of the mobile device.

Acknowledgments
This project (HA project no. 690/19-10) is financed with funds of LOEWE – Landes-Offensive zur Entwicklung Wissenschaftlich-ökonomischer Exzellenz, Förderlinie 3: KMU-Verbundvorhaben (State Offensive for the Development of Scientific and Economic Excellence).

Bibliography

[BRR16] Butcher, Peter WS; Roberts, Jonathan C; Ritsos, Panagiotis D: Immersive analytics with webvr and google cardboard. Posters of IEEE VIS, pp. 30–32, 2016.

[CGB09] Cirulis, Arnis; Ginters, E; Brigmanis, K: Virtual reality's technologies use in e-learning. In: Proceedings of the 8th WSEAS International Conference on E-Activities, WSEAS Press, Puerto De La Cruz, ESP. pp. 148–153, 2009.

[Co19] Coma-Tatay, Inmaculada; Casas-Yrurzum, Sergio; Casanova-Salas, Pablo; Fernández-Marín, Marcos: FI-AR learning: a web-based platform for augmented reality educational content. Multimedia Tools and Applications, 78(5):6093–6118, 2019.

[DA12] Dodd, Bucky J; Antonenko, Pavlo D: Use of signaling to integrate desktop virtual reality and online learning management systems. Computers & Education, 59(4):1099–1108, 2012.

[DT03] Dougiamas, Martin; Taylor, Peter: Moodle: Using learning communities to create an open source course management system. In: EdMedia+ Innovate Learning. Association for the Advancement of Computing in Education (AACE), pp. 171–178, 2003.

[FI20] FIWARE Foundation: FIWARE. https://www.fiware.org/, 2020. Accessed: 13.04.2020.

[Go20a] Google: Firebase Realtime Database and Dynamic Links. https://firebase.google.com/docs, 2020. Accessed: 13.04.2020.

[Go20b] Google: Flutter. https://flutter.dev/, 2020. Accessed: 13.04.2020.

[hG] health&media GmbH: arztCME. https://www.arztcme.de/year={2020},. Accessed: 13.04.2020.

[La19] Lam, Kit Yung; Lee, Lik Hang; Braud, Tristan; Hui, Pan: M2A: A Framework for Visualizing Information from Mobile Web to Mobile Augmented Reality. In: 2019 IEEE International Conference on Pervasive Computing and Communications (PerCom. IEEE, pp. 1–10, 2019.

[Li04] Liarokapis, Fotis; Mourkoussis, Nikolaos; White, Martin; Darcy, Joe; Sifniotis, Maria; Petridis, Panos; Basu, Anirban; Lister, Paul F et al.: Web3D and augmented reality to support engineering education. World transactions on engineering and technology education, 3(1):11–14, 2004.

[Pa17] de Paiva Guimarães, Marcelo; Alves, Bruno; Martins, Valéria Farinazzo; dos Santos Baglie, Luiz Soares; Brega, José Remo; Dias, Diego Colombo: Embedding augmented reality applications into learning management systems. In: International Conference on Computational Science and Its Applications. Springer, pp. 585–594, 2017.

[Pa18] de Paiva Guimarães, Marcelo; Alves, Bruno Carvalho; Durelli, Rafael Serapilha; de FR Guimarães, Rita; Dias, Diego Colombo: An Approach to Developing Learning Objects with Augmented Reality Content. In: International Conference on Computational Science and Its Applications. Springer, pp. 757–774, 2018.

[PDG20] Persia, Fabio; D'Auria, Daniela; Ge, Mouzhi: Improving Learning System Performance with Multimedia Semantics. In: 2020 IEEE 14th International Conference on Semantic Computing (ICSC). IEEE, pp. 238–241, 2020.

[Un20] Unity Technologies: Unity game engine. https://unity.com/, 2020. Accessed: 13.04.2020.

Presenters in Virtual Reality in Slideshow Presentations

Robin Horst[1] Linda Rau[1] Lars Dieter[1,2] Manuel Feller[1,2] Jonas Gaida[1,2] Andreas Leipe[1,2] Julian Eversheim[1,2] Julia Wirth[1,2] Jörn Bachmeier[1,2] Julius Müller[1,2] Maik Melcher[1,2] Ralf Dörner[1]

Abstract: Slideshow presentations have become ubiquitous in our everyday life, and are used for communicating information of different kinds. In this paper, we consider two different concepts that include both slides and VR technology in one presentation, *mixed presentations*, and *virtual presentations*, and examine the role of the presenter in these concepts. We conducted three user studies which indicate that it is not necessary that presentations need to be held completely in VR as both virtual and mixed presentations were accepted by our participants, and that our participants preferred immersed presenter integrations.

Keywords: Short Virtual Reality Experiences; Slideshow Presentations; Game Engine Integration; E-Learning

1 Introduction

Presentation software, such as PowerPoint, has become a standard tool in different environments of everyday life, such as work, home or education, and supports communicating information. Such software already supports different established resources, such as text, images, sound, and video. Virtual Reality (VR) is not among these established means, even though head-mounted displays (HMDs) for VR are becoming affordable and applicable concerning the costs and the ease of use. Therefore, VR is no longer reserved for expert use and becomes more and more a part of daily life (e.g., within E-Learning approaches). However, there exist challenges that must be considered before using VR among other resources in slideshow presentations. While the audience takes the active part of a VR-mediated presentation and uses HMDs to experience the virtual content, presenters still need basic control over the presentation procedure (e.g., switch to next/previous slides) to comply with prescribed constraints, such as time limitations. Another challenge relates to the technical integration of VR technology in common presentation software. How can a switch from a common PowerPoint slide to a VR experience be realized?

In this paper, we make the following contributions: We investigate how presenters can be integrated into *mixed presentations*, where a regular slide presentation switches to

[1] RheinMain University of Applied Sciences, Kurt-Schumacher-Ring 18, 65197 Wiesbaden, Germany, firstName.lastName@hs-rm.de
[2] These authors contributed equally.

doi:10.18420/inf2020_89

and from VR applications, and *virtual presentations*, where slides are adopted within a virtual environment. We implemented three prototypes to explore the user role of the presenter and show how short game-engine-based VR experiences can be integrated within established presentation software. We discuss aspects of our implementations, show how VR applications can be triggered from a PowerPoint presentation, and use our prototypes to draw conclusions on how the audience perceived the attendance and influence of presenters.

2 Related Work

In the literature, there exist examples for VR applications where one user has to guide another user through the virtual world. A remote instructor guides persons to repair complex machinery [Od15], or a researcher conducts a virtual demo [HD18, HD19].

Fuhrmann et al. [Fu01] contribute technical work about presentation systems that use VR technology. They adopt the slide concept of slideshow presentations in their system and transfer it into a VR environment. In a frontal presentation or a combined setting, they enable presenters to show 3D content to the audience. The presenter takes over the active part of a presentation and the audience can see both the presentation and the presenter in both of their settings. It remains open how the audience could take the active part and make use of the interactive VR technology. Their evaluation is based on technical feasibility.

Steed et al. [St02] suggest the user role of a virtual presenter in their 'ante-room'. It is a virtual representation of the experimenter during a study or a demo and can give participants instructions. The presenter is visualized by a virtual puppet that is controlled from a desktop PC. A study shows that the users' sense of traversal could be reinforced when a transition was provided for the VR users that also included visually transitioning the presenter from the physical to the virtual world.

Price [Pr08] proposes UnrealPowerPoint as a new learning and teaching methodology. The author describes the usage of common PowerPoint slides within a computer game. Both learners and teachers can participate within the presentation by using a desktop PC interface and both user groups are represented by humanoid avatars. Single slides are not presented separately but can also be visualized simultaneously. According to the authors, the Unreal slides also have additional functionalities that can go beyond common slides, which makes their concept more specific. Learners are granted the freedom to explore these slides in a non-linear way, which supports their educational purposes. Since the paper is oriented towards educational sciences, technical details are not considered here.

3 Presenter and Virtual Reality Integration

Both mixed and virtual presentations can integrate presenters in different ways using available devices. While the audience uses HMD-based VR to benefit from the immersive

3D technology, presenters may (1) not interact with the virtual content at all and let the audience explore the VR scene, (2) interact through a desktop interface (*asymmetric integration*) or (3) be fully immersed using HMD-technology (*immersive integration*), as well. The connection between VR and presentation software can serve as a foundation for these integrations. We implemented three prototypes:

- PowerPoint integration (*PP*) – This prototype serves as a foundation for connecting presentation software and VR technology. As a use-case, it implements a presentation about forestry and forests for both a complete virtual and a mixed presentation.
- Asymmetric presenter (*AP*) – The AP prototype provides an asymmetric interface for presenters to interact with the virtual world and VR users. It uses a presentation about the solar system.
- Immersive presenter (*IP*) – The IP implementation represents an immersive interface for presenters. This prototype presents content about different sights on the world (e.g., the Eiffel Tower and the Chichén Itzá)

The PP implementation connects PowerPoint with a Unity game engine application. This feature is fundamental for mixed presentations, as game-engine-based VR is to be inserted at specific points during the slideshow. Based on using an existing PowerPoint presentation and enriching it with VR experiences, we identified three possibilities for realization. First, the Microsoft Office Interop interface provides the possibility to react on the advancing of a slide from a running C# application, so that a running Unity application can process these events for example to switch to certain Unity scenes. But an identifier would be needed to support using different VR experiences at different points within the presentation. Secondly, APIs such as OpenXML provide functionalities to search slides with a specific layout for keywords. This mechanism can be used to trigger certain Unity events when a keyword is found on the currently active slide. However, this option would restrict authors of the slides to comply to a necessary slide-layout/structure. Finally, Interop allows to query the number of the current slide independent of the layout. This information enables a running Unity application to start the VR or turn it off when a specific slide of the presentation is set as an active slide. Before using this connection in a presentation, a mapping from VR experiences to PowerPoint slides must be performed during the authoring process. Therefore, the slideshow should be finalized before the mapping process, as inserting or removing slides would destroy the intended flow of the transitions. We implemented this third method for our prototype. Since PowerPoint is a software that can be opened multiple times, or even outside our use case on the same machine, we use a controller script that knows both the path of the .pptx-file and the corresponding Unity application. This script establishes a connection between them and can be used to start the applications as well. This ensures that VR is only triggered by the correct PowerPoint instance.

The PP prototype also included usual PowerPoint slides within the virtual environment itself to implement a virtual presentation. We identified two methods to integrate slides

within a Unity application with low additional effort for presenters/authors. At first, a digital screencast or a webcam that records the physical presentation can be used to stream the slide content on a texture within the virtual environment. This requires additional soft- or hardware and can be difficult to set up, especially for webcam technology. As a result, visual quality may suffer and is directly dependent on the additional components. Secondly, each slide can be integrated within the VR as a separate image. These can be exported automatically with the mentioned Interop API. The images are used as textures on the virtual projection plane. They can be loaded by a running Unity application when they are exported to the 'Resources' folder within the assets. The drawback of this method is the absence of animations. Finally, we chose to implement an image-based workflow within our prototype. Necessary animations could be included in this method by using multiple slides to approximate the visual animation. We designed a separate virtual room with a projection plane for all slide adoptions (Fig. 1 top left) and changed the position of the VR users when a switch from a slide adoption to a VR experience (Fig. 1 bottom right and left) was intended. Events to change the positions or swap the slide images can be implemented using the Interop events as described. Presenters only have to interact with the original PowerPoint software in this prototype. All virtual rooms were implemented in one scene in this case (Fig. 1 top right). A simple webcam stream of the presenter was visualized during the slide adoptions to reflect the experience of a common slide presentation, where the presenter can be seen by the audience. Both the following AP and IP implementations share features described for the PP prototype and will therefore not be mentioned again.

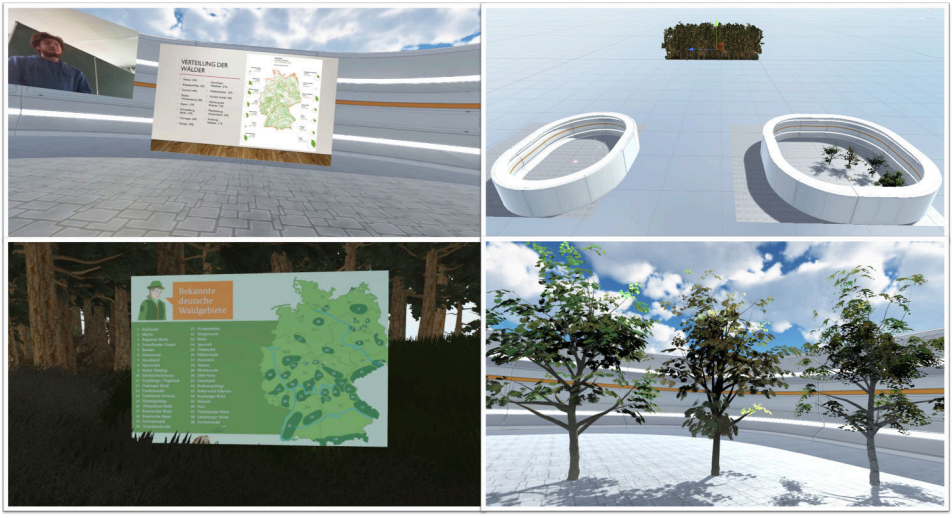


Fig. 1: Screenshots from the PP prototype presentation.

The AP prototype provides presenters a desktop PC interface to interact with the virtual environment during the VR experiences between the slides (Fig. 2). We provide presenters a top-down view on the virtual scene of the VR users. This view includes buttons for

adjusting the participants' position, resizing predefined objects, and system interactions (e.g., switching to next slides/VR experiences). The interface was implemented using Unity UI components and an orthographic camera.

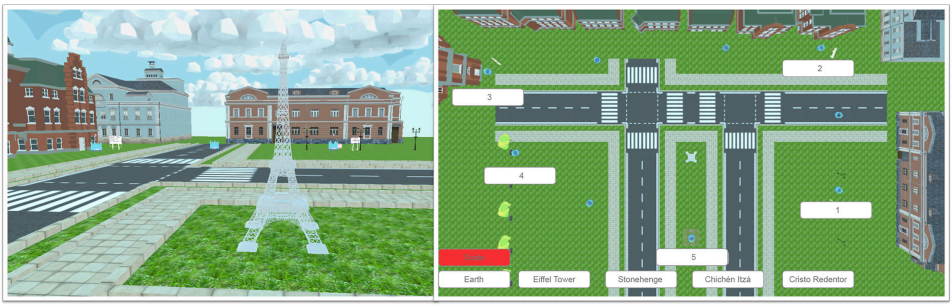


Fig. 2: Screenshots from the AP prototype presentation.

The IP implementation provides presenters an immersive interface to the VR experiences of the VR users (Fig. 3). Presenters are integrated within the virtual environment with a first-person view. It provides them with similar interactions as the AP prototype, with the additional functionality to invite VR users to a quiz about the slide content and to rate slides. It also enables them to point with a laser pointer within the scene to guide the VR users through the scene. VR users are represented with a minimalistic humanoid avatar. The same representation is used for presenters, with the difference that they wear a crown to indicate that they have capabilities beyond the ones for VR users (Fig. 3 left). Avatars and interactions are implemented using Steam VR components.

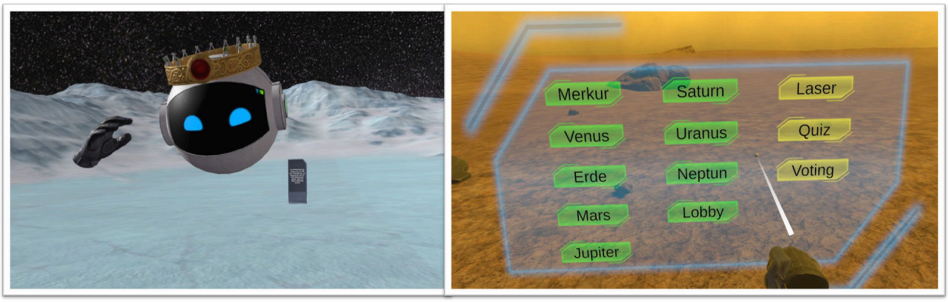


Fig. 3: Screenshots from the IP prototype presentation.

4 Evaluation

We evaluated the proposed presentation techniques and presenter integrations in three distinct user studies – one with each prototype (PP, AP, and IP). We call the corresponding studies *PP study*, *AP study*, and *IP study* respectively. Each prototype incorporated different

content but with comparable quality of the assets (e.g., 3D models and textures). Overall, the user studies involved 35 unpaid, voluntary, and experienced participants. Their VR experience was captured on a 0-3-point scale, where 0 means they had never used VR technologies and 3 means they regularly use VR. PP study involved 11 participants (Ø 26,7 years, 3 females) with Ø 2.0 experience. AP study involved 14 participants (4 female, Ø ~21,5 years) with Ø 2.0 experience. IP study involved 10 participants (2 female, Ø 23,5 years) with Ø 1.9 experience. The procedure of each study took place as follows: The participants were briefly introduced to the user interface of the prototype and the VR hardware. Then an experimenter took the role of the presenter and gave a presentation to the participants using the distinct prototype for each study. In the AP study, we divided a longer presentation into two experiments which were executed in a randomized order: AP-1) Participants experienced the VR parts of the presentations by themselves and AP-2) the participants were guided by the experimenter using the desktop interface. Similarly, we divided the PP study into two experiments with randomized order: PP-1) The presentation was held using a mixed presentation methodology and PP-2) the presentation was held using a fully virtual presentation.

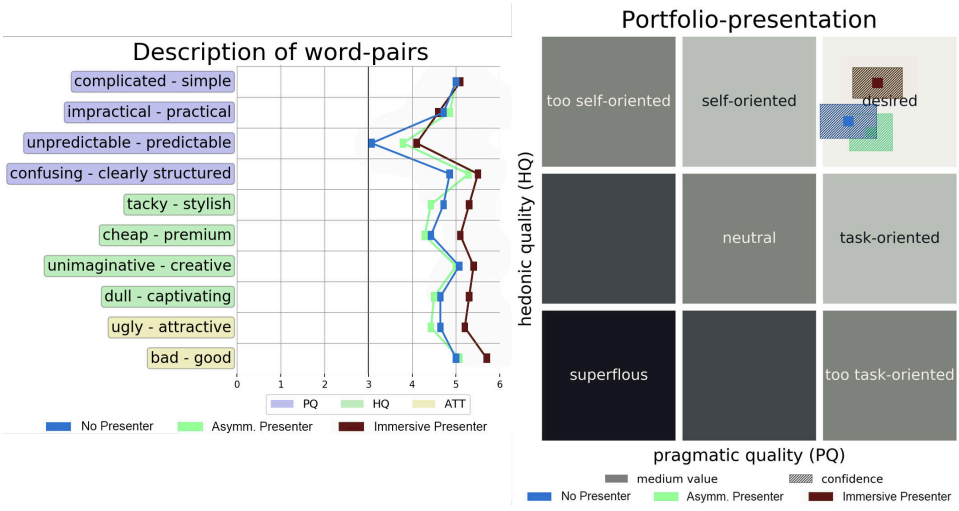


Fig. 4: AttrakDiff analysis [HBK03] that compares an immersive presenter interface (IP), an asymmetric interface (AP-2), and an interface that does not offer direct interactions with the virtual scene (AP-1). Left: Description of word pairs. Right: Portfolio presentation.

After each experiment/presentation, the participants were asked to fill out a questionnaire, which was translated into their native language. The AP and the IP questionnaire consisted of an abbreviated version of the AttrakDiff questionnaire [HBK03]. The questionnaire for the PP study included eight questions that utilized a 7-point semantic differential scale: Q1: Would you like to stay longer in the virtual world? Q2: Did the virtual rooms help in understanding the content of the presentation? Q3: Were the texts, drawings, graphics easily

recognizable? Q4: Would you like more interaction with the presentation? Q5: Did you find the HMD unpleasant? Q6: Did you find your way around the VR well? Q7: How did you feel about the different VR rooms? Q8: Would you recommend this type of presentation to others?

The results of the AttrakDiff questionnaires for the AP and IP studies are illustrated in Fig. 4. The charts show that all three presentations (no presenter, asymmetric presenter, and immersive presenter) were perceived positively by our participants. Presentations without a presenter or with an asymmetrical presenter integration performed similarly well, with the difference that the asymmetric presenter was perceived more task-oriented, whereas the scores for the absence of a presenter tended towards self-orientation (Fig. 4 right). The immersive version was perceived best regarding hedonic and pragmatic qualities. Only one of its items did not get the highest score ('impractical-practical', Fig. 4 left).

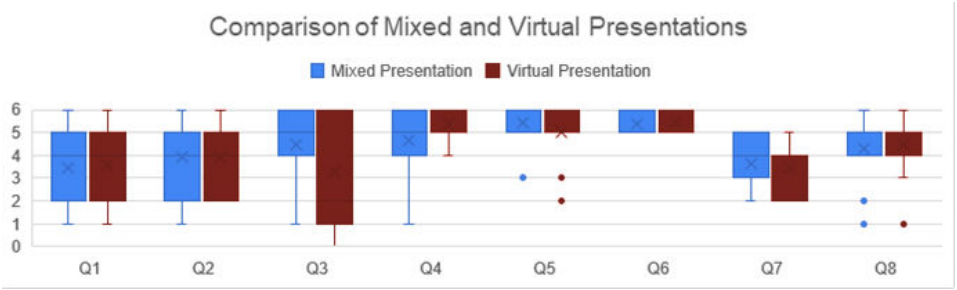


Fig. 5: Box- whisker plots comparing mixed (PP-1, blue) and virtual presentations (PP2, orange).

The chart in Fig. 5 illustrates differences between mixed and virtual presentations drawn from the PP study. The bar chart shows similar scores and distributions among the two modes (PP-1 and PP-2) for Q1, Q2, Q5, Q6, Q7, and Q8. We performed Wilcoxon Signed-Rank tests [WW64] on the items Q3 and Q4 with a threshold for statistical significance of 5% to analyze further differences between mixed and virtual methodology. The tests could not confirm a statistically significant difference between the two conditions. The absolute differences for Q3 indicate that our participants preferred to view common slides in the physical world as they stated to recognize drawings, graphics, and texts more easily there. Even though both presentations were perceived similarly positive, the scores for Q4 indicate that our participants expressed the desire to have more interaction possibilities with virtual slides than it would be possible with usual slides in the physical world.

5 Conclusions and Future Work

In this paper, we explored possibilities to integrate presenters when VR is used within a slideshow presentation. As a technical foundation, we have shown how game-engine-based VR technology can be used to implement these concepts and how game engine VR can be

connected to established slideshow presentation software, such as PowerPoint. Our user studies indicate that both virtual and mixed presentations were accepted by our participants and that an immersed presenter was preferred.

Finally, we will explore transitioning techniques between physical slideshows and short VR experiences within mixed presentations. Current work targets rather extensive and complex transitioning to VR to improve the experience of VR users. As participants of mixed presentations may put on and take off VR HMDs frequently within a single presentation, such elaborate transitions could be disproportionate concerning our VR experiences in-between slides. This will be addressed in future research directions. Furthermore, we evaluated the presenter integration from a VR user's perspective, but expert presenters must be included in future studies, too, in order to create the best possible experiences for both presenters and the audience of VR-enriched presentations.

Acknowledgments
The work is supported by the Federal Ministry of Education and Research of Germany in the project Innovative Hochschule (funding number: 03IHS071).

Bibliography

[Fu01] Fuhrmann, Anton L; Prikryl, Jan; Tobler, Robert F; Purgathofer, Werner: Interactive content for presentations in virtual reality. In: Proceedings of the ACM symposium on Virtual reality software and technology. ACM, pp. 183–189, 2001.

[HBK03] Hassenzahl, Marc; Burmester, Michael; Koller, Franz: AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität. In: Mensch & computer 2003, pp. 187–196. Springer, 2003.

[HD18] Horst, Robin; Dörner, Ralf: Opportunities for Virtual and Mixed Reality Knowledge Demonstration. In: 2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct). IEEE, pp. 381–385, 2018.

[HD19] Horst, Robin; Dörner, Ralf: Integration of Bite-Sized Virtual Reality Applications into Pattern-Based Knowledge Demonstration. In: Proceedings of the 16th Workshop Virtual and Augmented Reality of the GI Group VR/AR. Gesellschaft für Informatik, Shaker Verlag, pp. 137–148, 2019.

[Od15] Oda, Ohan; Elvezio, Carmine; Sukan, Mengu; Feiner, Steven; Tversky, Barbara: Virtual Replicas for Remote Assistance in Virtual and Augmented Reality. In: Proceedings of the 28th Annual ACM Symposium on User Interface Software Technology. ACM, 2015.

[Pr08] Price, Colin B: Unreal PowerPoint™: Immersing PowerPoint presentations in a virtual computer game engine world. Computers in human behavior, 24(6):2486–2495, 2008.

[St02] Steed, Anthony; Benford, Steve; Dalton, Nick; Greenhalgh, Chris; MacColl, Ian; Randell, Cliff; Schnädelbach, Holger: Mixed-reality interfaces to immersive projection systems. In: Immersive projection technology workshop. 2002.

[WW64] Wilcoxon, Frank; Wilcox, Roberta A: Some rapid approximate statistical procedures. Lederle Laboratories, 1964.

A Discussion on Current Augmented Reality Concepts Which Help Users to Better Understand and Manipulate Robot Behavior

Kai Groetenhardt[1]

Abstract:

For a safer, more trustful, and more dynamic collaboration, humans should understand and be able to manipulate the behavior of robots they are interacting with. Therefore, a way for a meaningful communication has to be established that takes place in a common perceptual space. One way to accomplish that is to use augmented reality (AR) in which the robot is able to display information for the human in 3D space, and the human can send commands to the robot using interaction methods provided by AR devices. In this work, a brief overview of AR concepts is given and discussed. They are divided into three categories: (1) understanding the movement of robots, (2) understanding the internal states of robots, and (3) manipulating robot behavior. Whereas (1) and (2) already show a number of promising approaches, and (3) is still in need for more innovative ideas.

Keywords: robot behavior; human-robot interaction; augmented reality

1 Introduction

Looking at active research in robotics, it is imaginable that robots will increasingly find their way into private households. Therefore, it seems important that humans without technical background are able to understand the behavior of robots and can control them. Breazeal et al. [Br01] identified an overlapping perceptual space as a key requirement for effective human-robot interaction, and Collett et al. [CM06] further explains that the perceptual space differs in input and output, which is illustrated in Fig. 1. Consequently, one obstacle is that humans and robots have different perceptual spaces, which just partially overlap. Not every action a human can perform is within the perceptual space of a robot, and a robot cannot reach every form of perception that is available to a human. Furthermore, humans and robots have differing conceptual models of the world; robots use several sensors to collect data of their surroundings and use various routines to interpret it. Sometimes they also need knowledge about the real world, which is provided by a knowledge framework. Those different perceptual spaces and different conceptual models of the world are reasons why understanding robot behavior can be difficult for humans.

[1] RheinMain University of Applied Sciences, Faculty of Design - Computer Science - Media, Unter den Eichen 5, 65195 Wiesbaden, Germany, kai.groetenhardt@hs-rm.de

doi:10.18420/inf2020_90

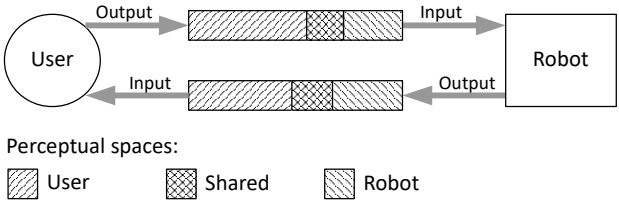


Fig. 1: An illustration showing the concept of the perceptual spaces of a robot and a human overlapping wherein a meaningful communication is possible. Source: Self-made rework of Figure 1 in [CM06].

Augmented reality (AR) can be a tool to widen the overlapping parts of the perceptual spaces of humans and robots, and is therefore able to make understanding robots easier for humans. Robots can send information to AR devices, which then can be visualized for the humans to perceive, and humans can use the interaction possibilities of those devices to send commands to the robot.

The aim of this paper is to give a brief overview of current work regarding human-robot interaction (HRI) with the focus on understanding and manipulating robot behavior using AR, and discuss possible future work. This paper starts with the categorization, presentation and discussion of the state of the art in Section 2. In Section 3 follows a conclusion summarizing what was learned, and suggesting ideas for future work to hopefully enhance the topic of using AR for HRI further.

2 Discussing State of the Art

AR in combination with robotics is not a new topic. In fact, the literature review of Geen et al. [Gr08] about human-robot collaboration AR approaches contains papers dating back to the early 2000s addressing that topic. In the past, efforts using AR were hampered by limitations in the available AR head-mounted displays (HMDs), which often were custom-built. With a new batch of AR HMDs like the *Microsoft HoloLens*[2] and the *Magic Leap 1*[3] some limitations of the past were reduced or eliminated and enabled researchers to conceptualize and implement new AR concepts for HRI.

The following papers, describing the usage of AR for HRI, are divided into the three categories (1) understanding the movement of robots, (2) understanding the internal states of robots, and (3) manipulating robot behavior. Each category is limited to a maximum of two papers to not exceed the scope of this work. If a category contains two papers, they are chosen to be similar for a better comparability, but they are different regarding their goals and approaches. Those papers are not depicted to the full extent, instead their presentment is limited to the most relevant parts. To every paper the motivation is stated, followed by the

[2] https://www.microsoft.com/en-us/hololens
[3] https://www.magicleap.com/en-us

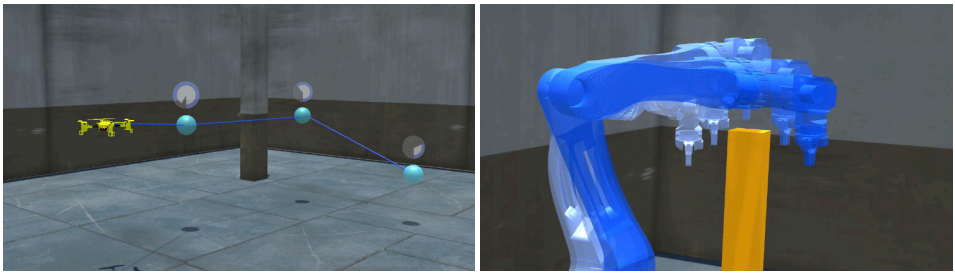
description of its AR concept, and concluded with the results of a user study, if available. At the end of each category, the concepts are being discussed.

2.1 Understanding the Movements of Robots

One defining characteristic of robots is their ability to move within the real world, sometimes in collaborative work spaces attended by humans. Unfortunately, robots do not necessarily have the ability to communicate their motion through gestures, gaze, or other social cues like humans. Here, two papers are presented showing possibilities to help the user understand the movement of robots with the help of AR.

Walker et al. [Wa18] argue that there are difficulties identifying when, where, and how a robot will move, which represents a primary challenge towards achieving safe and usable robotic systems. To tackle that problem, they introduce four concepts to indicate future movements of a flying robot using the AR HMD *Microsoft HoloLens*. The first concept, called *NavPoints* (shown in Fig. 2 (A)), adds virtual navigation points displayed as spheres into the 3D space. The spheres are connected through lines, which indicate in what order the robot will pass them. Above the spheres two radial timers are displayed, which show when the robot will arrive and when it will leave that position. The second concept, which is called *Arrow*, is similar, but a more minimal approach. An animated arrow shows the route the robot will take a few seconds in the future. As the arrow moves it leaves a line behind showing the path it was taking. The third concept is called *Gaze*, which augments the robot by a 3D model of an eye that is looking in the direction of travel. The fourth and last concept they presented is named *Utilities*. It is a 2D circular radar displayed at a corner of the user's perceptual space that shows the robots position relative to the user. Eventually, they compared the concepts by conducting a user study to see, among other things, how the displayed virtual imagery affected participant understanding of robot movement intent. The test showed that *NavPoints* ranked best followed by *Arrow*, *Gaze* and *Utilities*.

Rosen et al. [Ro20] indicate that a robot's movement intent can be shown on a 2D screen, but this requires the human to take their attention away from the robot's physical space to observe the display, which could be dangerous. Additionally, a 2D projection of a 3D motion can take time for a human to understand, requiring interaction to inspect different points of view. As a test scenario they chose a robot arm that performs a programmed movement with some objects nearby. The task for the human is to check if the robot will hit the objects before it even starts moving. To make that possible, a virtual 3D model of the robot arm is displayed multiple times along the planned path in 3D space visible through the AR HMD *Microsoft HoloLens*, shown in Fig. 2 (B). To have a reference, they implemented that same concept for a 2D screen with the possibility to move the virtual camera via mouse and keyboard. They compared both concepts and found that their AR system reduced the completion time of the task and increased the average accuracy of collision predictions.



(A) (B)

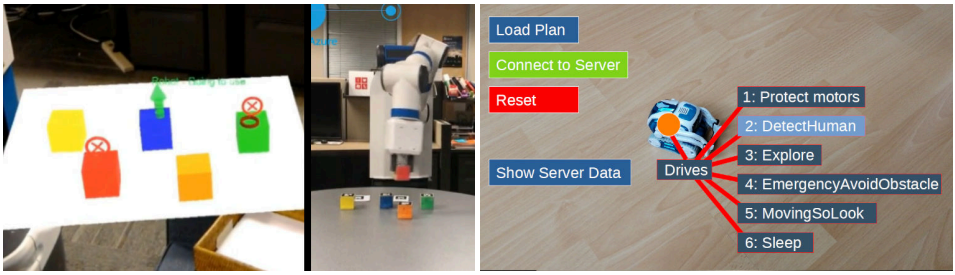
Fig. 2: Two different concepts communicating robot movement intent to humans. (A) shows the concept of view through the *Microsoft HoloLens* displaying the *NavPoints* concept from Walker et al. [Wa18] in which connected waypoints, the arrival, and the departure time of a robot can be seen. Source: Self-made rework of Figure 1A in [Wa18]. In (B) the concept of Rosen et al. [Ro20] is shown in which several steps of the planned movement are displayed in full size. Source: Self-made a rework of Figure 1 in [Ro20].

Comparing the concepts of Walker et al. and Rosen et al., it is apparent that they both show the robots motion intents, but target different scenarios. Walker et al. show where the robot will be located for users to adapt their own behavior towards the robot. Rosen et al. show the robot's future movement for the human to be able to intervene in the robot's behavior. It is imaginable to combine both concepts, but divide them into a planning and execution mode, which can be switched by the user. For the planning mode, the concept of Rosen et al. could be used to see detailed movements and to identify collisions. In execution mode, the concept of Walker et al. would show the path of the robot and when it will reach waypoints.

2.2 Understanding Internal States of Robots

To not only understand the movement of robots, but also the robots' decision-making process that leads to movements or other actions, an interface to the humans' perceptual space needs to be established. In this section, two papers are discussed showing a robot's plan of action via AR.

Chakraborti et al. [Ch18] cite the *Roadmap for U.S. Robotics* [Ch09] by saying "humans must be able to read and recognize robot activities in order to interpret the robot's understanding". They argue that attempts were made to accomplish the idea with natural language, but the state of the art limits the scope of such interactions, especially where precise instructions are required. To show an alternative, they communicate the intentions of a robot using AR to a collaborating human. Their setup consists of a robot that is tasked to stack colored boxes and a human who is equipped with a *Microsoft HoloLens* and has the ability to claim boxes through an AR interface. A virtual 3D model of boxes mirroring the boxes that are positioned in front of the robot is displayed for the human in 3D space, as can be seen in



(A) (B)

Fig. 3: Two different concepts to communicate a robot's intent to a human. (A) shows the concept of Chakraborti et al. [Ch18]. On the left hand side, the view through the *Microsoft HoloLens* and the mirrored model of the boxes in front of the robot can be seen. The virtual boxes are annotated with symbols that indicate the robot's plan. On the right hand side, the robot is displayed while executing its task of stacking boxes. Source: A frame taken from a video linked in [CSKK18] with kind permission of Tathagata Chakraborti. In (B), the concept of the *Android* AR app of Rotsidis et al. [Ro19] can be seen that shows the robot annotated with its current plan represented as a hierarchical graph in which the active task "DetectHuman" is highlighted. Source: Self-made rework of Figure 2 in [Ro19].

Fig. 3 (A). Those virtual boxes can be annotated by the robot to show what its intentions regarding those boxes are. The robot marks a box with a green upward pointing arrow to communicate that this box is the next one the robot is going to pick up. Also, boxes the robot intends to use in the future are marked with a circled red cross. The human on the other hand, has the ability to claim boxes for themselves even if the robot already has indicated to use them. In that case, the robot removes the mark at the corresponding box and chooses another one to complete its task. Unfortunately, there is not a user study yet, but Chakraborti et al. announced their intention to conduct one.

Rotsidis et al. [Ro19] state that it's important for end-users to have a mental model of their robot that contains the capabilities and awareness of its limitations in order to trust it. Subsequently, through transparent decision-making of the robot it is possible for the users to adjust their expectations and forecast certain actions of the robot. The authors' attempt to tackle that challenge is based on an AR application running on an *Android* handheld device that shows the plan of the robot in form of an hierarchical graph. If the app detects the robot, it displays the graph next to it. The graph shows tasks the robot is able to perform and highlights the task the robot is currently executing, which can be seen in Fig. 3 (B). The user has the possibility to interact with the graph to see more or less information. They conducted a user study that showed the robot is perceived more alive, livelier, and friendlier with the app than without it.

Interestingly, Chakraborti et al. chose to show a virtual copy of the real objects and annotated them instead of annotating the real objects directly in AR. The authors did not disclose why the went this way, but it would be interesting to find out if direct annotations could improve the usability. Comparing the concepts of Chakraborti et al. and Rotsidis et al., it is clear that

both show the robots' plan, but, like in the previous Section 2.2, one is more detailed in its approach. Rotsidis et al. only show what task is being executed, whereas Chakraborti et al. also show how the current task is being executed. Additionally, Chakraborti et al. developed a specific vocabulary to communicate the robots plan in form of annotations. In contrast, Rotsidis et al. communicate the plan via text arranged in a graph. Of course, a more detailed approach is not always the better choice since too much information could lead to problems of its own, for example by overloading the user or showing unwanted information. It needs to be determined in what scenario one concept is more suited than the other, or if a scalable solution combining the two concepts would be the better approach.

2.3 Manipulating Robot Behavior

If a robot needs to be taught how to execute a new task or change its behavior, a typical way to achieve this is to reprogram it by using text-based or even visual programming languages. But there are also other approaches like programming by demonstration (PbD), which is a field of research of its own that also includes AR solutions (e. g. [OK18], [Qu18]), or rather unconventional approaches like knowledge patching used in the following paper.

Liu et al. [Li18] point out that machine learning methods have reached a remarkable level of effectiveness in specific tasks, but still have their limitations. For example, they lack interpretability of the knowledge representation, especially about how and why a decision is made, which plays a vital role in the scenarios where robots work alongside humans. In their system, they use interpretable knowledge represented by an And-Or-Graph (AOG) instead. Their setup consists of a robot with two arms and a *Microsoft HoloLens*, which can, among other things, display a 2D interface for the AOG in 3D space in front of the robot. The task to be solved is for the robot to open a medicine bottle with a lid that does not only have to be twisted but also pushed. The user starts with an AOG that describes how to open a normal bottle. The robot needs to be taught a push movement using PbD within the AR environment for it to be patched into the AOG by the human. To do the patching, the human can interact with the graph using hand gestures, comparable to mouse clicks and mouse drags, to remove and add nodes. The interface can be seen in Fig. 4, also a video [In] showing the whole process is available.

One could argue that the described part of Liu's et al. concept is a movement snippet manager that allows the user to combine little movements into a more complex motion, which is an interesting approach to not overburden the user. In their paper, they changed the process of opening a bottle globally, which means even non-medical bottles get opened using that push and twist movement. It would be interesting to see this concept combined with some sort of a teachable object recognition system to be able to chose the opening process in a more targeted manner. Consequently, that would need to be implemented into the AR interface, which would present a new challenge. Researching robot behavior modeling via AR, unfortunately, revealed very little approaches outside the PbD field, which

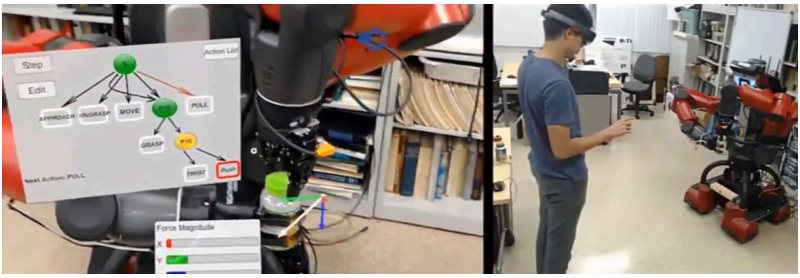


Fig. 4: The concept of Liu et al. [Li18] from two different perspectives. On the left hand side, the view trough the *Microsoft HoloLens* is displayed in which the human can see the AOG representing the knowledge to open a medicine bottle. On the right hand side, a third person's view without virtual elements is displayed. Source: A frame taken from a video [In] linked in [Li18] with kind permission of Hangxin Liu.

resulted in only one paper in this section without something comparable, which leaves that topic open for more innovative ideas.

3 Conclusion and Future Work

In this work, a brief overview of AR concepts dealing with robot behavior was given and discussed. The state of the art shows that there are already several concepts proven to be helpful in understanding robot behavior. Others look promising, but their effectiveness needs to be tested. The presented papers differ in their aims and level of detail in a way that makes them prone to be combined. Combinations of the described concepts could lead to improvements that would be interesting to see in future work. All things considered, using AR to understand robots seems to be a viable approach to further pursue.

In contrast, more accessible interfaces to manipulate robot behavior in AR seem to be a difficult endeavor. After a thorough research, only one paper could be found that chooses an approach (at least partly) deviant to PbD. More ideas need to be developed and tested to see if AR is the right tool to manipulate robot behavior.

During the discussion, some suggestions for improvements were made, which could be conceptualized in more detail in future work. Especially the concept of Chakraborti et al. [Ch18] is an interesting candidate to pursue further to see if annotating real objects instead of the virtual copies of them feels more natural for the users.

Bibliography

[Br01] Breazeal, Cynthia; Edsinger, Aaron; Fitzpatrick, Paul; Scassellati, Brian: Active vision for sociable robots. IEEE Transactions on systems, man, and cybernetics-part A: Systems and Humans, 31(5):443–453, 2001.

[Ch09] Christensen, Henrik I; Batzinger, T; Bekris, K; Bohringer, K; Bordogna, J; Bradski, G; Brock, O; Burnstein, J; Fuhlbrigge, T; Eastman, R et al.: A roadmap for US robotics: from internet to robotics. Computing Community Consortium, 44, 2009.

[Ch18] Chakraborti, Tathagata; Sreedharan, Sarath; Kulkarni, Anagha; Kambhampati, Subbarao: Projection-aware task planning and execution for human-in-the-loop operation of robots in a mixed-reality workspace. In: 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, pp. 4476–4482, 2018.

[CM06] Collett, T. H. J.; MacDonald, B. A.: Developer Oriented Visualisation of a Robot Program. In: Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-robot Interaction. HRI '06, ACM, New York, NY, USA, pp. 49–56, 2006.

[Gr08] Green, Scott A; Billinghurst, Mark; Chen, XiaoQi; Chase, J Geoffrey: Human-robot collaboration: A literature review and augmented reality approach in design. International journal of advanced robotic systems, 5(1):1, 2008.

[In] Interactive Robot Knowledge Patching Using Augmented Reality - YouTube. https://www.youtube.com/watch?v=AqjmIhKGGus. accessed: 07/04/2020.

[Li18] Liu, Hangxin; Zhang, Yaofang; Si, Wenwen; Xie, Xu; Zhu, Yixin; Zhu, Song-Chun: Interactive robot knowledge patching using augmented reality. In: 2018 IEEE International Conference on Robotics and Automation (ICRA). IEEE, pp. 1947–1954, 2018.

[OK18] Ostanin, M; Klimchik, A: Interactive robot programing using mixed reality. IFAC-PapersOnLine, 51(22):50–55, 2018.

[Qu18] Quintero, Camilo Perez; Li, Sarah; Pan, Matthew KXJ; Chan, Wesley P; Van der Loos, HF Machiel; Croft, Elizabeth: Robot programming through augmented trajectories in augmented reality. In: 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, pp. 1838–1844, 2018.

[Ro19] Rotsidis, Alexandros; Theodorou, Andreas; Bryson, Joanna J; Wortham, Robert H: Improving robot transparency: An investigation with mobile augmented reality. In: 2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN). IEEE, pp. 1–8, 2019.

[Ro20] Rosen, Eric; Whitney, David; Phillips, Elizabeth; Chien, Gary; Tompkin, James; Konidaris, George; Tellex, Stefanie: Communicating robot arm motion intent through mixed reality head-mounted displays. In: Robotics Research, pp. 301–316. Springer, 2020.

[Wa18] Walker, Michael; Hedayati, Hooman; Lee, Jennifer; Szafir, Daniel: Communicating Robot Motion Intent with Augmented Reality. In: Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction. HRI '18, Association for Computing Machinery, New York, NY, USA, p. 316–324, 2018.

R. Reussner, A. Koziolek, R. Heinrich (Hrsg.): INFORMATIK 2020,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2021 979

Requirements and Mechanisms for Smart Home Updates

Peter Zdankin[1] Oskar Carl[2] Marian Waltereit[3] Viktor Matkovic[4] Torben Weis[5]

Abstract: The interconnection of sensors and actuators of smart home devices creates dependencies that allow for ubiquitous services. These devices can be subject to transformative changes through software updates that might lead to unintended consequences. Users have no tools to predict the negative consequences caused by updating their smart home. In this paper, we address this problem and propose mechanisms that enable organized update planning in a smart home. We compare self-description standard approaches that allow reasoning about resulting functionality before updates are installed. Updating devices to their latest versions is not necessarily the best way to update smart homes, therefore we discuss multi-objective optimization in the update process. Finally, outsourcing functionality to external providers might reduce the complexity of certain tasks, but can also pose threats if the wrong tasks are offloaded.

Keywords: Smart Home; Longevity; Self-Description; Update Configuration; Edge Computing

1 Introduction

A smart home is composed of multiples devices from various vendors. Each vendor is producing software updates on his own, which leaves it to the user to select the best set of updates. Vendors cannot always consider all possible setups, and therefore faults can occur when a device is updated. Furthermore, downgrading software versions is not always possible, which means that a single unfortunate update can permanently impact the functionality of a smart home. Currently, users have no tools to master the update problem without risking damage to the system. Thus, it is important to investigate the update process and to propose solutions to compute the optimal update configuration automatically. In Section 3 we discuss how self-description of devices and services can principally solve the update problem. Then we discuss possible definitions of optimality of such a solution in Section 4. Finally, in Section 5 we compare solutions to the update problem relying on centralized services in the cloud with solutions that work locally in the smart home and discuss their impacts on ease of use and autonomy.

[1] University of Duisburg-Essen, Distributed Systems Group, Duisburg, Germany, peter.zdankin@uni-due.de
[2] University of Duisburg-Essen, Distributed Systems Group, Duisburg, Germany, oskar.carl@uni-due.de
[3] University of Duisburg-Essen, Distributed Systems Group, Duisburg, Germany, marian.waltereit@uni-due.de
[4] University of Duisburg-Essen, Distributed Systems Group, Duisburg, Germany, viktor.matkovic@uni-due.de
[5] University of Duisburg-Essen, Distributed Systems Group, Duisburg, Germany, torben.weis@uni-due.de

doi:10.18420/inf2020_91

2 System Model

The architecture of smart home systems may vary considerably between different implementations. Some systems are designed and deployed during the construction of a house and permanently deployed e.g. via bus systems in walls. While this is an interesting option, most homes require expensive renovation to become smart this way. An alternative to this is the usage of modular subsystems, for example, standalone lighting or heating systems, which can be bought individually and may offer their own separate platform, or integrate into bigger smart home platforms such as Amazon Alexa, Google Home, Samsung SmartThings or HomeKit. The benefit of modular smart home systems is a lower entry price for consumers, as they might start off with light bulbs or thermometers that can be controlled through already available smartphones or affordable hubs. In this paper, we focus on the latter, as the distributed and heterogeneous nature of these systems is more likely to break at some point during its lifetime, in comparison to a permanent solution installed during house construction. A smart home configuration can be described through the following system model:
*A **smart home** has a set of **devices** that are connected through a **platform**. Each device has a certain **software version** and a set of available **updates**. Each software version has a set of available **predefined services**. Devices can use services of other devices which creates **dependencies**. An **update configuration** is one of the finite states of the nondeterministic finite automaton NFA that can be constructed by using the configurations as states and connecting them using individual updates as transitions.*

During the lifetime of devices in a smart home, security issues or new features might require new software, which might alter the functionality of a device through for example a modified public API. These alterations might be intended, as part of a necessary change, or accidental because certain side effects were not considered. In any way, if such an update is installed, it will likely break an existing dependency. Currently, users cannot predict the changes that will be imposed on a smart home if a subset of its devices is updated simultaneously. Naive approaches that only consider individual updates for each update step, miss the broad picture because dependency problems can manifest after a certain set of devices have been updated already and rolling them back to previous versions may not always be possible.

3 Self-Description of Smart Home Devices

To enable interoperability, devices can describe their services to other devices, as was proposed in an architecture by Barbas et al. [Ve12]. Devices must be able to list all their currently available services, either through self-awareness or information included in the updates. By comparing the currently available services with the available services of an update, the differences can be computed. That way a system can predict possible faults in dependencies before they occur. However, comparing service definitions cannot capture unintended incompatibilities due to programming errors. Furthermore, most service

definitions are only syntactic, e.g. via an API definition, but they do not specify behaviour. Thus, an update might cause a change of behaviour without a change of its service definition. Two widely used approaches for service definition are *descriptive* or *prescriptive* standards.

3.1 Descriptive vs Prescriptive

A standard for service definitions that only regulates how a device can describe itself is a *descriptive standard*, because it allows vendors to describe their devices in their own terms. A recently finalized descriptive standard is the Web Of Things [Ka20; Ko20]. If devices need to communicate across different descriptive standards, a translation must be considered. However, McCool et al. have stated security concerns about purely descriptive approaches, such as scanning for door locks with known physical weaknesses [MR18].

A *prescriptive* standard does not only regulate the *how* but also the *what* of self-description. These standards prescribe how devices and services should be defined and vendors can use these terms for unambiguity. By using the same terms to define services, translation is not required anymore. A prescriptive standard defines all possible devices and all their services they could implement. However, a device can implement a subset of these services only. This restriction has benefits as well because it ensures compatibility across device vendors. Widely used examples are the smart home standards of Amazon, Apple, and Samsung[6].

3.2 Descriptive Standards Only

In a descriptive standard, each vendor can create its own namespace and definition for a certain type of device. This open approach is flexible and allows vendors to act quickly and independently. It also allows vendors to reinvent service definitions, as they are incentivized to invent proprietary definitions in order to support new features. This can lead to multiple definitions of the same device type, thus creating redundancy as no incentive is given to find an agreement between vendors. As multiple definitions can coexist, communicating devices might need to translate between the definitions. Alternatively, a middleware could be introduced to translate between incompatible but equivalent descriptions. Over time, device vendors could update their own definitions in ways that impair interoperability, even to the point of defective device functionality. Therefore, we assume that descriptive standards alone are not suited for the evolving smart home ecosystem.

[6] Amazon Alexa Skill API Documentation: `https://developer.amazon.com/de/docs/smarthome/understand-the-smart-home-skill-api.html` (accessed May 6, 2020),
Samsung SmartThings: `https://smartthings.developer.samsung.com` (accessed May 6, 2020),
Apple HomeKit Accessory Protocol Specification: `https://developer.apple.com/support/homekit-accessory-protocol/` (accessed May 6, 2020)

3.3 One Prescriptive Standard

In a single prescriptive standard, each device type and service would be defined free of redundancy. While this allows for optimal interoperability, many practical problems require consideration. First of all, a regulatory body needs to be decided on to define this standard. This might cause long delays in the process of reaching consensus between the stakeholders, given the current interest in smart home applications.

If only devices that are defined in this standard can function in such a smart home, innovative new products that are not yet defined must also first pass the committee before being able to operate in a smart home. This does not just create obstacles for new products, it also enables competitors in the market to start production of these devices while they are being standardized. Small vendors that invent new smart home products might have problems to compete against bigger ones, due to the time loss introduced by the certification process. Due to this, we assume that a single prescriptive standard is too restrictive since the domain is evolving rapidly.

3.4 Hybrid Standards

Based on the above conclusions, we propose a hybrid solution between descriptive and prescriptive standards. Smart home devices are usually connected to a single platform and not part of multiple platforms simultaneously. Hence, a prescriptive standard per platform could have the desired flexibility, because platform vendors can innovate independently. To improve interoperability, this can be combined with a descriptive standard between all platforms. All devices – regardless of their respective vendors and platforms – would self-describe using that descriptive standard but adhere to the rules of the prescriptive standard given by the respective platform vendor. This way, each platform maintains optimal interoperability via prescriptive definitions. Since all prescriptive definitions are instances of a single descriptive language, it is still possible to translate between multiple platforms as discussed above. Furthermore, it is more likely that multiple prescriptive standards converge (at least partially), since they use the same descriptive language. Thus, they only have to agree on terms, but not on the syntax used to define services and devices.

4 Optimality of Update Configurations

When smart home systems are updated, the objective of current systems is to install the latest software on all devices, regardless of dependencies or other objectives. While this is one possible approach, there exist others such as dependency robustness or flexibility of the smart home system. These different goals are much harder to achieve because additional constraints must be upheld. Single-objective optimization might be able to yield acceptable

results to some degree. Users might want to hold onto their existing automations, which in some cases could be damaged by the latest software policy.

If the goal is only to ensure that no functionality is broken, the solution is to simply pick the latest versions that do not remove any functionality currently in use. However, as soon as an update removes functionality, the problem is expanded to multiple objectives: It requires weighting between the benefits provided by the latest software version and the convenience of keeping all functionality unchanged. Determining an optimal solution becomes even more complicated when new functionality is introduced at the cost of another one. Zitzler et al. have compared evolutionary algorithms as a means to search for possible solutions while considering conflicting optimization goals [ZT99]. Solving multiple-objective optimization is a subject that has been explored for a long time, and a large number of possible solutions have been found [MA04; ZT99]. In this problem space the aim is not perfect optimization, which is commonly impossible, but for *Pareto optimality*. It is an approximation in which multiple configurations might be considered equal. This requires a choice between these solutions to be made, which can be implemented in the form of user choice between policies like *feature stability* and *security*. Equivalent results according to optimality can be presented to the user, who is then required to choose *a posteriori* [Br08].

However, the devices available in a smart home are usually also constrained in terms of performance or power. This can render such approaches unviable in the smart home context. To amend this situation, the user choice should be made *a priori* [MA04].

Self-description of services allows considering dependencies in a smart home analytically to find optimal update configurations for new objectives before the updates are installed. As it is possible to know which changes will happen once a certain update configuration is chosen, configurations that disrupt dependencies can be discarded. Static analysis over the available services and possible dependencies can be enhanced through a dynamic approach that tracks which services are actually used in a smart home. This way, services that are actually used can be considered during the update process, while services that are not used can be completely ignored in the search for an optimal configuration.

5 Autonomy of Smart Homes

User management, device communication, update planning, and automated tasks of a smart home might not occur locally but on remote servers. If the autonomy of a smart home is constrained by outsourcing functionality to external providers, the smart home depends on the availability of these providers. This availability cannot be guaranteed for the lifetime of smart homes. Despite this, it represents a common mode of operation for smart home systems currently in use, such as in the solutions of Amazon or Samsung.

External services (located in the cloud) can offer resources to solve computationally intensive tasks, manage authentication and other security-critical necessities, and offer a

gateway for remote access. These benefits provide a large incentive to waive autonomy in a smart home and assume that external services and internet connection are always available. An external service can even use approaches like testing updates of devices against their specification to reason about the correct functionality. It might also perform updates on specific configurations under laboratory settings. Nevertheless, if smart homes target lifetimes of at least 10 years and multiple vendors are involved, it becomes likely that some external services are shut down. Possible threats against the longevity of smart homes exist and have happened before [Zd20a; Zd20b].

5.1 Remote Update Planning

Giving away autonomy can be dangerous if update planning is performed remotely. To find an optimal update configuration, a smart home system must transmit information about all devices in the smart home, their installed software, and used dependencies to the remote service. Transferring information about usage habits in the form of dependencies and usage patterns is privacy-invasive. Specifics on the software installed on devices can disclose vulnerabilities currently open for exploitation at a location. By abusing the knowledge of vulnerable software on smart devices, access to various parts of the smart home could be gained and used to invade the privacy of users or even risk the security of the entire local network. Remote update planning can also pose additional dangers, as devices can be advised to update to a vulnerable software version, which might open up an attack vector. Waiving autonomy in a smart home must be considered carefully, as the impact depends on the task performed remotely.

5.2 Local Update Planning

Autonomy in a smart home requires local resources to solve problems that would otherwise be resolved with the help of a centralized external service. Smart home platforms like OpenHAB[7] strive to be autonomous, at the cost of much higher complexity. The higher complexity is a burden for non-technical users who feel overwhelmed by the amount of work necessary to configure and maintain a completely autonomous system.

To find the optimal update configuration autonomously, a suitable device must be available in the smart home. We will refer to this as the *central device*. This device needs to find out which other devices are part of the smart home, how they are connected, and which dependencies exist. Furthermore, the central device should monitor which functionality is actually being used in the smart home installation. Thus, the central device must query devices in the network or it must query local hubs, at least one for each platform in use. The practical problem of this approach is that some platforms provide no API to query this

[7] OpenHAB Documentation: `https://www.openhab.org/docs/` (accessed May 6, 2020)

information. While it is usually possible to enumerate all devices connected to a hub, it is often not possible to query which functionality is being used, or how these devices are connected among each other.

As smart home devices are created by numerous vendors using various platforms, a single database for available updates does not exist in general. The central device must therefore either query all device vendors for updates, or it must rely on the user to make updates available locally. Automatically querying device vendors implies that an external service is being used again. This implies that the smart home is leaking information about the devices deployed and the software versions installed to a wide range of device vendors. From a privacy perspective, this might be even more questionable than transmitting this information to a platform vendor like Apple, Amazon or Google, because these are at least known to the user, while users are usually not able to judge the trustworthiness of an oversea device manufacturer.

Once the central device has information about all possible updates including the self-description for all updates, it can compute which services are added or removed if a specific update is installed. Through optimality criteria, many of these update configurations can be discarded and the most advantageous configurations can be obtained. As stated before, multiple optimal configurations might exist. In the worst case, the number of possible options is $O(2^n)$ where n is the number of updates because this is the count of all possible subsets of updates. Obviously, it is not reasonable to present these options to the user.

Therefore, we propose a policy selection like *latest version*, *conservative*, or *feature set* to capture the intent of the user. This policy can be used to further filter the set of Pareto optimal configurations. Finally, the central device performs an update path that is Pareto optimal and this complies best to the chosen policy. Further research is required to actually develop and evaluate such a system to gain insight into the feasibility of this approach.

6 Conclusion

We analyzed the update problem for smart homes. The currently dominant approach is potentially dangerous, as not enough measures are taken to prevent harmful updates from being installed in a live system. Furthermore, the update process can potentially leak critical information, which violates privacy and can pose security risks for the entire network, because it might disclose attack vectors. We have shown that service definitions are required for the update planning and discussed descriptive and prescreptive approaches and their practicality. While autonomy is a desireable property for smart homes, local update planning is more complex than update planning that relies on cloud-based services of platform vendors. Finally, we discussed what optimality means for update planning and concluded that optimality alone is not sufficient to select an update path, since multiple optimal configurations can exist. We proposed policies to capture the user intent and to finally select one optimal update path.

Optimality criteria are not exclusive to smart homes. Dependency management of software projects can encounter similar issues. Hence, research in one of these domains might benefit both. As devices become more powerful, it might be possible to pull services running in the cloud into the home network and to deploy them in containers. Thus, smart homes could use external services for convenience, but lack no features if the cloud becomes unavailable or the user does not want to use external services.

References

[Br08] Branke, J.; Miettinen, K.; Deb, K.; Sowiski, R.: Multiobjective Optimization, Vol. 5252 of Lecture Notes in Computer Science. Multiobjective Optimization 5252/, pp. 1–8, 2008.

[Ka20] Kaebisch, S.; Kamiya, T.; McCool, M.; Charpenay, V.; Kovatsch, M.: Web of Things (WoT) Thing Description, first Edition of a Recommendation, https://www.w3.org/TR/wot-thing-description/, W3C, Apr. 2020.

[Ko20] Kovatsch, M.; Matsukura, R.; Lagally, M.; Kawaguchi, T.; Toumura, K.; Kajimoto, K.: Web of Things (WoT) Architecture, first Edition of a Recommendation, https://www.w3.org/TR/wot-architecture/, W3C, Apr. 2020.

[MA04] Marler, R. T.; Arora, J. S.: Survey of multi-objective optimization methods for engineering. Structural and multidisciplinary optimization 26/6, pp. 369–395, 2004.

[MR18] Mccool, M.; Reshetova, E.: Distributed Security Risks and Opportunities in the W3C Web of Things. In. Jan. 2018.

[Ve12] Vega-Barbas, M.; Casado-Mansilla, D.; Valero, M. A.; López-de-Ipiña, D.; Bravo, J.; Flórez, F.: Smart Spaces and Smart Objects Interoperability Architecture (S3OiA). In: 2012 Sixth International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing. Pp. 725–730, 2012.

[Zd20a] Zdankin, P.: Longevity of Smart Homes. In: PerCom PhD Forum 2020: 18th Annual IEEE International Conference on Pervasive Computing and Communications PhD Forum (PerCom PhD Forum 2020). Austin, USA, Mar. 2020.

[Zd20b] Zdankin, P.; Waltereit, M.; Matkovic, V.; Weis, T.: Towards Longevity of Smart Home Systems. In: PerIoT 2020: 4th International Workshop on Mobile and Pervasive Internet of Things (PerIoT 2020). Austin, USA, Mar. 2020.

[ZT99] Zitzler, E.; Thiele, L.: Multiobjective evolutionary algorithms: a comparative case study and the strength Pareto approach. IEEE Transactions on Evolutionary Computation 3/4, pp. 257–271, 1999.

R. Reussner, A. Koziolek, R. Heinrich (Hrsg.): INFORMATIK 2020,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2021 987

Complexity Analysis of Task Dependencies in an Artificial Hormone System

Eric Hutter[1], Mathias Pacher[1], Uwe Brinkschulte[1]

Abstract: The Artificial Hormone System (AHS) is a self-organizing tool able to allocate tasks in a distributed system. We extend the AHS in this paper by negator hormones to enable conditional task structures and provide a thorough complexity analysis of the resulting system. The analysis shows that the problem to decide if a given task A is instantiated at all respecting the negators is NP-complete.

Keywords: Artificial Hormone System; negators; conditional task execution; complexity analysis

1 Introduction

We describe and analyze the decision problem NEGATOR-SAT occurring when using an *Artificial Hormone System* (AHS) [BP12] in this paper. The AHS is able to allocate tasks on a set of distributed processors without using a central instance and offers a high dependability of the task allocation. While the original AHS assumes a task model of independent tasks, we extend it here by assuming conditional dependencies between the tasks: E.g. a task T_1 can only be executed when another task T_2 is *not* executed. This allows to enable alternative task structures within the AHS.

Our contribution in this paper is twofold: (1) We shortly describe our extension of the AHS including the *negator hormones*. Their purpose is to enable conditional task structures. (2) Conditional task dependencies induced by negators make it hard to determine if a given task A can be instantiated at all. We call this decision problem NEGATOR-SAT and prove its NP-completeness. We end the paper by providing a transformation example of a satisfiable propositional formula to a task set using negators allowing to instantiate task A.

The paper is structured as follows: Section 2 presents the State of the Art in self-organizing systems. Section 3 gives an introduction to the original AHS while section 4 briefly explains our negator implementation. The complexity analysis of NEGATOR-SAT as well as an example are provided in section 5. Section 6 concludes the paper and describes future work.

2 State of the Art

IBM's *Autonomic Computing* initiative [LMD13] introduced so-called *self-x* properties such as self-configuration, self-optimization and self-healing. The MAPE-K loop was

[1] Goethe University, Frankfurt am Main, Germany, {hutter, pacher, brinks}@es.cs.uni-frankfurt.de

doi:10.18420/inf2020_92

established to realize monitoring and analyzing of a system's behavior and to plan and execute actions controlling its behavior according to a knowledge base and user-defined goals. This loop has recently been adopted to establish self-explainable systems by using a MAB-EX loop (monitor, analyze, build, explain), see [Bl19]. The above mentioned self-x properties are also central to systems realized using *Organic Computing* concepts [TSM17]: Here, computer systems and embedded systems are constructed by incorporating concepts inspired by biological systems and their organization principles. This approach allows systems to dynamically adapt to changing operational conditions, realizing self-x properties like self-configuration or self-healing at run-time.

3 The Artificial Hormone System

The AHS' main purpose is to allocate tasks in a distributed system of processors, called *processing elements* or *PEs*. It is completely decentralized and has no single point of failure. In addition, it provides self-x properties such as self-configuration, self-optimization and self-healing and guarantees real-time bounds [BP12].

The AHS uses different kinds of hormones (which are short messages) to allocate the tasks. The main hormone types are eager values, suppressors and accelerators. Eager values indicate the suitability of a PE to take a task. As soon as a PE takes a task it sends suppressors for it. In this way, it tells the other PEs that it has taken the task: This is a life-sign on the one hand and it saturates the hormone balance on the other hand, thus limiting the number of allocated instances of this task. Accelerators are used to locate related tasks (i.e. tasks with communication relations or access to the same sensors or actors) nearby each other.

The core of the AHS is the hormone loop. Each PE iterates the loop, computing the hormone balance for each task. The duration needed by one hormone loop iteration is called a *hormone cycle*. In the *receive stage*, the hormones for each task are received. In the *compute and decision stage*, the suppressors received for a task are subtracted from its local eager value and the accelerators for this task are added. The result is the modified eager value which indicates the PE's current suitability to take this task. This computation is performed for each task. A PE's AHS instance then decides for a *single* task allocation per hormone loop iteration in order to allow the suppressors and accelerators to unfold their effect. In the following *send stage*, the PEs send eager values for all tasks (with the exception of an eager value that is 0) as well as suppressors and accelerators for all tasks they are currently executing. In this paper, we want to express conditional task relationships using the self-organizing AHS. A conditional task relationship means that a task T_1 can only be executed when another task T_2 is *not* executed. The concept is realized by special hormones called *negators* and allows to use alternative task structures in the AHS. This may be useful in a heterogeneous system of PEs. Details on the negators are described in section 4. Figure 2 gives a sketch of the hormone loop (already including the negators).

4 Conception

As mentioned before, our goal was to enhance the AHS by introducing the possibility to model conditional dependencies between tasks. To be precise, we introduced so-called *negator* relationships between tasks as visualized in Figure 1: Here, task T_j negates task T_i, meaning that task T_i cannot be assigned to any PE if T_j is assigned to a PE. Thinking in terms of a directed graph, this relationship can be expressed as the tuple (T_j, T_i).

Fig. 1: Negator relationship between tasks T_j and T_i: If T_j is assigned, T_i must not be assigned

This allows to express task dependencies: Suppose one PE_α can execute a task T realizing some functionality. Multiple other PEs cannot execute T but rather a set of tasks that realize the same functionality as T but with some kind of degradation, e.g. loss of precision. If PE_α is running, T's negator relationships to the other tasks prevent them from being instantiated. If PE_α fails, T is no longer available but the other tasks can be instantiated, keeping the functionality available.

4.1 Implementation

We implemented our concept of negator relationships that realize conditional inter-task dependencies by modifying the AHS' hormone loop as shown in Figure 2 (cf. [Br13] for

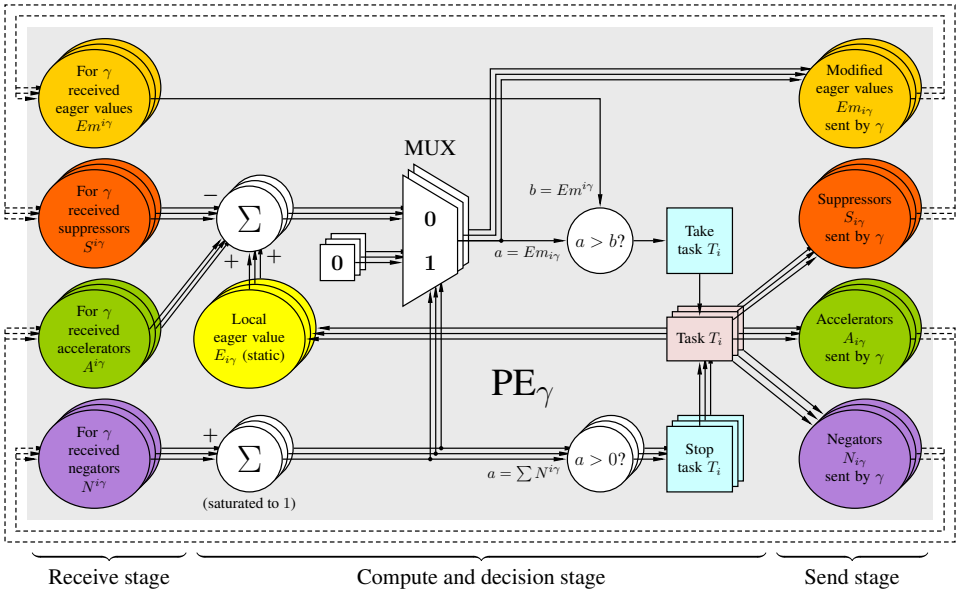


Fig. 2: Hormone loop with negators, running on PE_γ

information on the AHS' original hormone loop): We added an additional type of hormone, the so-called *negators*. If some task T_j is running and a negator relationship (T_j, T_i) exists, T_j will send a negator hormone to T_i during the hormone loop's send stage.

The received negator hormones are counted for each task. If at least one negator was received for some task T_i, two things happen: (1) T_i is stopped if it is running. (2) T_i gets blocked by forcing its modified eager value to 0, regardless of any suppressors or accelerators received for T_i. This prevents T_i from being assigned. If the negating task T_j is no longer assigned to any PE (e.g. because the PE it was running on failed), it won't send a negator for T_i any longer. This allows T_i's modified eager value to rise above 0 again, allowing T_i to be assigned again.

5 Theoretical Analysis

As has been seen in the previous section, the introduction of negators allows to model task dependencies, e.g. alternate sets of tasks to realize some functionality. However, negators also introduce new kinds of possible mistakes a designer can make during a system's design. Consider the following problem:

Definition 1 (NEGATOR-SAT)**.** Let $\mathcal{T}$ be a finite set of tasks and $\mathcal{N} \subseteq \mathcal{T} \times \mathcal{T}$ a set of negator relationships among those tasks.

The decision problem NEGATOR-SAT is now stated as follows: Given a task $A \in \mathcal{T}$, does a set of assigned tasks $\mathcal{V} \subseteq \mathcal{T}$ exist (with $T \in \mathcal{V}$ iff T is assigned to a PE) so that the following conditions are all satisfied:

(1) There is no task $T \in (\mathcal{T} \setminus \mathcal{V})$ that could be assigned to a PE even if all PEs had infinite computational resources,
(2) $\mathcal{V}$ is a stable task assignment, i.e. all negator relationships among tasks from $\mathcal{V}$ are respected,
(3) $A \in \mathcal{V}$, i.e. task A is assigned to some PE.

In simple terms, NEGATOR-SAT asks whether a stable task assignment exists so that A can be assigned to a PE. Condition (1) prevents the system's computational capacities from imposing any limits on such task assignment. Clearly, it can be regarded a design mistake if some task cannot be assigned to a PE at all. Thus, it should be checked if each task is assignable. However, it turns out that this seemingly simple problem is difficult to solve algorithmically:

Theorem 2. NEGATOR-SAT is NP-hard.

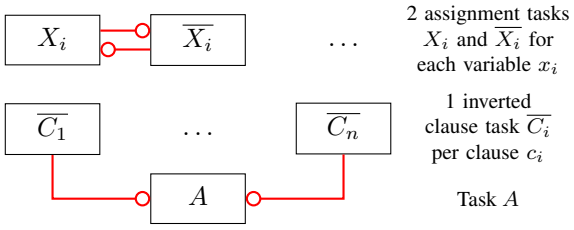
Proof. By reduction of 3-SAT to NEGATOR-SAT: We will show 3-SAT $\leq_p$ NEGATOR-SAT where $\leq_p$ denotes a polynomial-time reduction. If 3-SAT is reducible to NEGATOR-SAT in polynomial time, we can deduce that NEGATOR-SAT is at least as hard as 3-SAT. With 3-SAT being NP-complete, the NP-hardness of NEGATOR-SAT follows.

We will thus describe a transformation τ so that

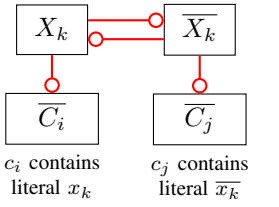
(i) τ can be computed in polynomial time w.r.t. the input length and
(ii) $f \in$ 3-SAT $\iff \tau(f) \in$ NEGATOR-SAT.

Let $f := \bigwedge_{i=1}^{n} c_i$ with $c_i := (l_{i,1} \vee l_{i,2} \vee l_{i,3})$ be a 3-SAT formula with $l_{i,j} \in \{x_k, \overline{x_k}\}$ for some k.[2] We will transform f into a task set $\mathcal{T}$ with negator relationships $\mathcal{N}$ so that the transformed input $\tau(f) := (\mathcal{T}, \mathcal{N}, A)$ is an instance of NEGATOR-SAT for the task $A \in \mathcal{T}$.

Construction of τ. The basic construction principle of this transformation is shown in Figure 3a. For each variable x_k occurring in f, we create two *assignment tasks* X_k and $\overline{X_k}$ that negate each other. This ensures only one of them can be assigned in a stable system, representing x_k's interpretation. Condition (1) ensures at least one of those assignment tasks is assigned per variable while condition (2) ensures that both cannot be assigned simultaneously.



(a) Basic construction principle



(b) Negator relationships between assignment and inverted clause tasks

Fig. 3: Construction principle of transformation from 3-SAT to NEGATOR-SAT

Additionally, we introduce one *inverted clause task* $\overline{C_i}$ per clause c_i. Thus, the resulting task set is

$$\mathcal{T} = \bigcup_{x_k \in \text{Variables}(f)} \{X_k, \overline{X_k}\} \cup \bigcup_{i=1}^{n} \{\overline{C_i}\} \cup \{A\}$$

where n is the number of clauses in f.

We will ensure that each inverted clause task $\overline{C_i}$ can (and per condition (1) will) be assigned iff f's interpretation does *not* satisfy the corresponding clause c_i by introducing negator relationships as follows (cf. Figure 3b):

- $(X_k, \overline{C_i}) \in \mathcal{N} \iff c_i$ contains the literal x_k and
- $(\overline{X_k}, \overline{C_i}) \in \mathcal{N} \iff c_i$ contains the literal $\overline{x_k}$.

Finally, the following negator relationships ensure that task A can (and, again, per condition (1) will) be assigned iff no inverted clause task $\overline{C_i}$ is assigned (and thus, all corresponding clauses c_i are satisfied): $(\overline{C_i}, A) \in \mathcal{N}$ for all $1 \le i \le n$.

[2] For simplicity, we require that each clause in f consists of exactly three literals. Note that this restriction does not change the problem's complexity as a clause can be padded to exactly three literals by repeating one of its literals.

See section 5.1 for an example of this construction.

τ is a polynomial-time reduction. We now need to show that above claims (i) and (ii) hold for τ, i.e. that τ is indeed a reduction of 3-SAT to NEGATOR-SAT.

(i): Polynomial time: It is easy to see that a formula f with n clauses and v different variables (with $v \leq 3n$) can be transformed in polynomial time w.r.t. f's length: We only need to construct $2v$ assignment tasks, n inverted clause tasks and the task A. This sums to $2v + n + 1 \leq 7n + 1$ tasks which is polynomial in the input formula's length.

Additionally, we construct $2v$ negator relationships for mutual exclusion of X_k and $\overline{X_k}$, $3n$ negator relationships between X_k resp. $\overline{X_k}$ and $\overline{C_i}$ and n negator relationships between $\overline{C_i}$ and A. This totals at $2v + 4n \leq 10n$ relationships which is also polynomial in the input formula's length.

(ii), part 1: $f \in$ 3-SAT $\Rightarrow \tau(f) \in$ NEGATOR-SAT:

Proof. Since $f \in$ 3-SAT, there must exist a satisfying interpretation $\mathcal{I}$: Variables$(f) \to \{0, 1\}$. Thus, consider the set $\mathcal{V} \subseteq \mathcal{T}$ of assigned tasks given as follows:

- $A \in \mathcal{V}$,
- for each inverted clause task $\overline{C_i}$: $\overline{C_i} \notin \mathcal{V}$,
- $X_k \in \mathcal{V} \iff \mathcal{I}(x_k) = 1$ and $\overline{X_k} \in \mathcal{V} \iff \mathcal{I}(\overline{x_k}) = 1$.

Due to τ's construction, it is easy to see that $\mathcal{V}$ satisfies conditions (1) to (3) as given by Definition 1:[3]

(1) All tasks in $(\mathcal{T} \setminus \mathcal{V})$ have an inbound negator link coming from an assigned task, thus no additional task can be instantiated.
(2) All negator relationships are respected: No two tasks from $\mathcal{V}$ share a negator relationship.
(3) A is assigned. ◇

(ii), part 2: $\tau(f) \in$ NEGATOR-SAT $\Rightarrow f \in$ 3-SAT:

Proof. Let $\mathcal{V} \subseteq \mathcal{T}$ be a set of assigned tasks so that conditions (1) to (3) as given by Definition 1 are satisfied. Thus, task A must be assigned. Therefore, per condition (2), no inverted clause task $\overline{C_i}$ can be assigned. Thus, at least one assignment task per inverted clause task $\overline{C_i}$ must be assigned (else, $\overline{C_i}$ would have to be assigned per condition (1)). Additionally, per condition (2), for each assignment task X_k resp. $\overline{X_k}$, the inverse assignment task $\overline{X_k}$ resp. X_k cannot be assigned.

This allows to construct an interpretation $\mathcal{I}$ for f so that $\mathcal{I}(x_k) = 1 \iff X_k \in \mathcal{V}$ and

[3] Note that—since $\mathcal{I}$ satisfies f—at least one literal is satisfied for each clause c_i, thus the corresponding inverted clause task $\overline{C_i}$ is not assigned. Since all inverted clause tasks are *not* assigned, A can (and per condition (1) will) be assigned to some PE.

$\mathcal{I}(\overline{x_k}) = 0 \iff \overline{X_k} \in \mathcal{V}$. In addition, $\mathcal{I}$ must satisfy f: Suppose $\mathcal{I}$ would not satisfy f. Then, there must be a clause c_i in f so that $\mathcal{I}$ does not satisfy any of its literals $l_{i,j}$. Due to τ's construction, this would mean that the inverse clause task $\overline{C_i}$ must be assigned per condition (1) which forbids A's assignment per condition (2). ⋄

Final remarks. Since τ can be constructed in polynomial time w.r.t. the input length, it follows that τ is indeed a polynomial-time reduction from 3-SAT to NEGATOR-SAT. Thus, NEGATOR-SAT is at least as hard as 3-SAT. Since 3-SAT is NP-complete, the NP-hardness of NEGATOR-SAT follows. □

However, NEGATOR-SAT is not only NP-hard, but also complete for NP:

Theorem 3. NEGATOR-SAT is NP-complete.

Proof. Let $(\mathcal{T}, \mathcal{N}, A)$ be an input consisting of a set of task $\mathcal{T}$, a set of negator relationships $\mathcal{N}$ and a task $A \in \mathcal{T}$. A nondeterministic Turing machine can now nondeterministically select a subset $\mathcal{V} \subseteq \mathcal{T}$ of assigned tasks and then deterministically check that conditions (1) to (3) from Definition 1 are all satisfied:

(1) This condition is satisfied if, for each $T \in (\mathcal{T} \setminus \mathcal{V})$, there is a negator relationship $(T', T) \in \mathcal{N}$ so that $T' \in \mathcal{V}$. However, this can be checked in $O(\text{poly}(|\mathcal{T}|, |\mathcal{N}|))$ time.
(2) This condition is satisfied if, for each $T \in \mathcal{V}$, there is *no* negator relationship $(T', T) \in \mathcal{N}$ so that $T' \in \mathcal{V}$. This can also be checked in $O(\text{poly}(|\mathcal{T}|, |\mathcal{N}|))$ time.
(3) This condition is satisfied if $A \in \mathcal{V}$ which can be checked in $O(\text{poly}(|\mathcal{V}|))$ time.

The Turing machine shall accept the input iff all three conditions are satisfied.

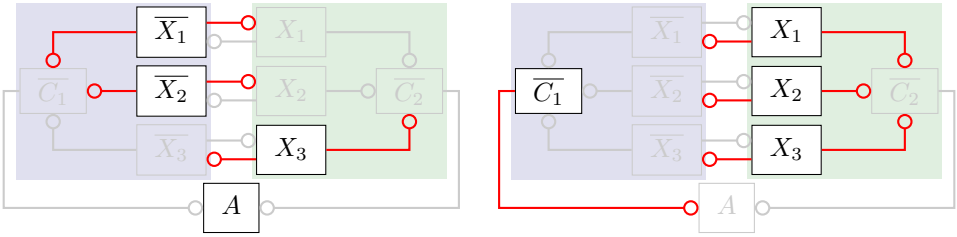
Since, after nondeterministically guessing $\mathcal{V}$, the verification can be performed in polynomial time w.r.t. the input length, it follows that NEGATOR-SAT ∈ NP. Together with Theorem 2, it follows that NEGATOR-SAT is NP-complete. □

This shows the power introduced by negators: Unless P = NP holds, it is not possible to decide in deterministic polynomial time whether a given task A can be assigned to a PE at all (when requiring a stable task assignment in which no additional tasks can be assigned).

5.1 Example of Construction

Figure 4 shows the construction result $\tau(f)$ for the formula $f = (\overline{x_1} \vee \overline{x_2} \vee \overline{x_3}) \wedge (x_1 \vee x_2 \vee x_3)$. Note that f is satisfiable and hence $f \in$ 3-SAT. It is easy to see that assigning either X_i or $\overline{X_i}$ for $i \in \{1, 2, 3\}$ allows assignment of A iff the assignment corresponds to a satisfying interpretation of f.

Additional examples of this construction and further considerations on problems involving negators can be found in [HPB20].



(a) With task assignment corresponding to satisfying interpretation of f

(b) With task assignment corresponding to unsatisfying interpretation of f

Fig. 4: NEGATOR-SAT instance constructed from 3-SAT instance $f = (\overline{x_1} \vee \overline{x_2} \vee \overline{x_3}) \wedge (x_1 \vee x_2 \vee x_3)$

6 Conclusion

We presented an negator extension for the AHS middleware in this paper. Negators enable conditional task execution in the AHS which is useful in a heterogeneous processor system. The use of negators impose the problem to determine if a given task A can be instantiated in the context of a stable task allocation in the overall system. We called the problem NEGATOR-SAT and proved its NP-completeness. Future work will consider the negators' impact on the AHS' real-time bounds and the stability of task allocations. This is important as it is simple to see that negators can generate oscillating task allocations.

Bibliography

[Bl19] Blumreiter, Mathias; Greenyer, Joel; Garcia, Francisco Javier Chiyah; Klös, Verena; Schwammberger, Maike; Sommer, Christoph; Vogelsang, Andreas; Wortmann, Andreas: Towards Self-Explainable Cyber-Physical Systems. In: 22nd ACM/IEEE International Conference on Model Driven Engineering Languages and Systems Companion, MODELS Companion 2019, Munich, Germany, September 15-20, 2019. pp. 543–548, 2019.

[BP12] Brinkschulte, Uwe; Pacher, Mathias: An Agressive Strategy for an Artificial Hormone System to Minimize the Task Allocation Time. In: 15th IEEE International Symposium on Object/Component/Service-Oriented Real-Time Distributed Computing Workshops, ISORC Workshops 2012, Shenzhen, China, April 11, 2012. pp. 188–195, 2012.

[Br13] Brinkschulte, Uwe; Pacher, Mathias; von Renteln, Alexander; Betting, Benjamin: Organic Real-Time Middleware. In (Higuera-Toledano, M. Teresa; Brinkschulte, Uwe; Rettberg, Achim, eds): Self-Organization in Embedded Real-Time Systems, pp. 179–208. Springer New York, New York, NY, 2013.

[HPB20] Hutter, Eric; Pacher, Mathias; Brinkschulte, Uwe: On the Hardness of Problems Involving Negator Relationships in an Artificial Hormone System. arXiv:2006.08958 [cs], June 2020.

[LMD13] Lalanda, Philippe; McCann, Julie A.; Diaconescu, Ada: Autonomic Computing - Principles, Design and Implementation. Undergraduate Topics in Computer Science. Springer, 2013.

[TSM17] Tomforde, Sven; Sick, Bernhard; Müller-Schloer, Christian: Organic Computing in the Spotlight. arXiv:1701.08125 [cs], January 2017.

Unified Approach to Static and Runtime Verification

Olga Thoss[1] Andreas Werner[1] Robert Kaiser[1] Reinhold Kroeger[1]

Abstract: Smart living environments are increasingly based on embedded information and communication technology. Generally, users are no technical experts and rely on the correct functioning of the system. Formal verification of a system's functional and non-functional properties is often regarded as the ultimate way to achieve the highest levels of trust as demanded for today's dependable systems. However, static verification, though sound in theory, is often impractical given the ever-increasing complexity of software and the non-deterministic nature of some mechanisms of the underlying hardware architecture. We argue that by supplementing static verification with runtime verification, a high level of trust can be achieved. In this paper, we report on an ongoing effort for tool-supported verification of functional and non-functional properties by combining static and runtime verification techniques.

Keywords: static verification; runtime verification; OS microkernel; SPARK; WCET; AQUAS

1 Introduction

Today's systems become more and more complex, and even domain experts are sometimes in doubt regarding their correct behaviour in rare and non-standard situations. Especially embedded systems incorporate increasing functionality and have to deal with a wide spectrum of sensors and actors interacting with the environment. Real-time properties requiring a guaranteed reaction of the system within a given limited time window are often associated as well, and safety of the users has to be ensured by law. Furthermore, these critical systems often have to face uncertainty which may originate from unknown device configurations at design time or unforeseen changes of the environment during operation. Uncertainty may also exist in control algorithms. For example, to guarantee a safe behaviour of trained AI algorithms in previously unseen situations is inherently difficult, if not impossible.

Under these conditions it is a complex and highly responsible task for developers to deliver a high level of trust in the developed software. This is commonly achieved through certification. To certify software for a given Safety Integrity Level (SIL), or ASIL level in the automotive systems context, it has to be thoroughly tested, specific models and analysis methods have to be used up to a formal, mathematical verification of required system properties. Today, all this has to happen before the system is actually used.

Due to the described complexity of current and future systems we do not believe that a full static verification and validation at design time is possible any longer to deliver the necessary trust. Instead, we have started to work on a methodology which distinguishes between

[1] RheinMain University of Applied Sciences, firstname.lastname@hs-rm.de

doi:10.18420/inf2020_93

verification activities carried out at design time and those at runtime. In summary, at design time static verification takes place, i.e. specified functional as well as non-functional or timing system properties are formally proven to the highest possible degree for a reasonable maximal effort. For properties which cannot be proven statically, sufficiently strong monitors are generated which are executed at runtime to monitor correct system behaviour. In the undesired case of detecting a property violation at runtime, the underlying system architecture is prepared to reconfigure the application to ensure acceptable behaviour. This adaptivity has to be supported by the application design. In total, a trusted self-adapting application seems to be reachable.

In the following Section 2, the functional properties, their verification and the approach for adaptation are considered. In Section 3, timing is taken into account as this is the most important non-functional property regarding real-time behaviour. In both sections, the current status of work is described. The paper closes with a summary and outlook. The considered use case and examples are taken from the AQUAS EU project [20b] to which we contribute.

2 Functional Verification

2.1 Static Verification

Typically, in the design phase a system model is constructed which specifies the overall system and its functional and non-functional properties and constraints. To simplify the designer's work, rather than using the industry standards SysML and OCL directly, we propose to use a DSL (Domain Specific Language) with an expressiveness to target the class of applications in mind. Such standard models shall then be generated from the DSL.

This generated system model will then be semi-automatically transformed into a set of views by exporting and transforming the model. Each view provides information concerning a specific aspect of the system. For the functional view and its verification, the SysML model is transformed into a SPARK interface definition, and functional OCL constraints into SPARK contracts with pre- and post-conditions (see Fig. 1). In addition to specified application constraints, other conformance rules, e.g. standards, company rules etc., can be taken into account as well, resulting in additional contracts or contract restrictions.

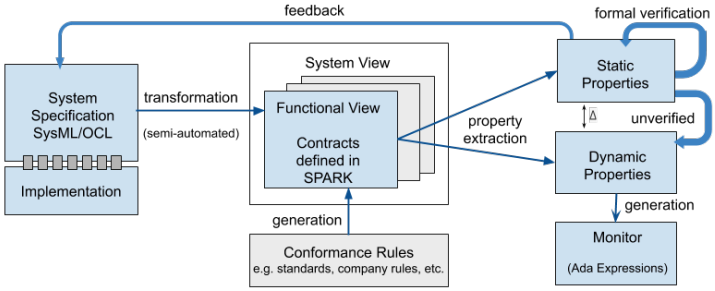


Fig. 1: Functional Verification During the Design Phase

During the design phase, the verification of selected functional properties can be carried

out solely at the contract or API level by taking advantage of a hierarchical system structure, deducing higher-level contracts from lower layers and assuming elementary contracts as facts [LNR80]. Thus, it is possible to provide early feedback regarding correctness of the system specification. Later, the assumed facts have to be proven by verification of the corresponding implementation.

2.2 Runtime Verification

Runtime verification is regarded as the discipline of computer science dealing with techniques to monitor systems during runtime in order to detect violations of given correctness properties [LS09]. More recent research, however, also considers controlling the system via feedback as belonging to runtime verification as well [LS09; Ru16]. Augmenting a functional system with a corresponding management or control system allows for autonomous behaviour of the system during operation.

As previously explained, not all contracts can be statically verified at design time. Our methodology extends the verification activities to the system runtime. The amount of possible static analysis and needed dynamic verification depends on the specific system and the complexity of its constraints. During the design phase, our methodology aims to separate as many properties or partial predicates as possible that can be verified statically and to automatically derive the complementary properties or predicates that have to be ensured and verified at runtime. Concerning the application level, necessary monitors will be generated from the unverified SPARK contracts during the design phase and executed by the runtime architecture.

In principle, monitors may be associated with all critical parts of the system which may be a source of uncertainty, like the application itself, the operating system and the hardware, but also the environment, especially when considering embedded systems. If a monitor assertion fails, an event is signalled to the runtime system (see Fig. 2).

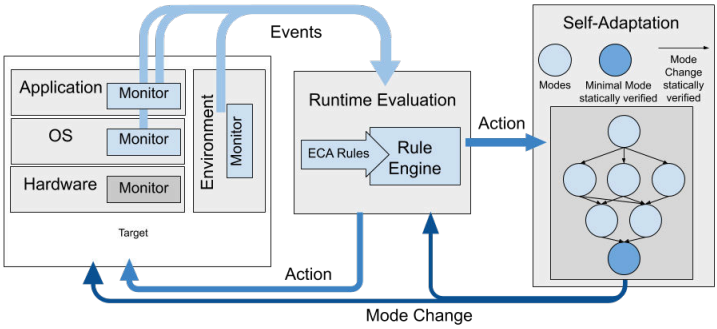


Fig. 2: Architecture for Runtime Verification

Thus, by extending the verification process to runtime, the methodology supports finding a manageable approach to verification of complex systems also covering uncertainty. This is especially important when static verification is based on pre-conditions, whose outcomes

can only be evaluated during runtime, e.g. if the pre-conditions are depending on user input or environmental factors.

As a disadvantage, incomplete static verification of system properties may result in property violations at runtime. Thus, provisions for events originating from failing runtime verification may be required. This is considered in the following section.

2.3 Adaptation Architecture

In the runtime system, signalled events will be received by an Event Condition Action (ECA) rule engine. Based on an application-dependent statically defined rule set, appropriate actions can be taken for detected runtime violations. Such actions can lead to a simple reconfiguration, like changing control parameters, or may require a more complex adaptation of the system.

As a basis for adaptation, the well-known concept of operational modes will be used. At each point in time, the system runs in a certain mode determining the provided functionality. At design time, a mode change graph will be developed, defining the set of operational modes and allowed changes. Also, the mode change algorithm must be statically verified for correctness. At runtime, mode changes may take place in response to signalled events. Thus, the system is guaranteed to be able to enter a well-defined state in case of a runtime verification failure, thus supporting graceful degradation.

In case of safety-critical systems, not every mode must ensure safety. A so-called "Minimal Mode"providing a safe state for the system is assumed to exist, whose functionality is statically verified. Due to the verified mode change algorithm, the minimal element can be reached in any case. Thus, a minimal level of service is ensured under all conditions.

2.4 Current Status

For initial evaluation, the functional part of the methodology was applied to the design and implementation of a queueing system and its use inside the scheduler of our microkernel Marron. Marron was developed by our group to serve as a template for a future verified microkernel. Besides scheduling, it features strict separation between user and kernel space, interrupt handling and inter-task communication. The queues where designed directly in SysML and constraints were defined in OCL. An appropriate DSL will be specified later, when more experience has been made. The use of SysML or OCL features was manually restricted to fit SPARK 2014 capabilities, and the transformation was done manually for now. The specification was verified based on the SPARK API. Necessary facts that have to be verified by the implementation were indicated by the `assume` pragma. Finally, the implementation was verified separately.

Based on the verified queueing system, a graduate student developed and verified the Marron scheduler in SPARK. The student had no prior experience with Ada, SPARK or formal verification in general. The goal of this experiment was also to evaluate the efficiency of our approach by measuring the effort needed to develop a verified operating system component.

The student spent a total effort of 450h over a course of six months, including literature work, project management, documentation, etc. He was able to verify 216 out of 224 verification conditions (VCs). The measured efforts spent on implementation and verification, as well as the relative effort in minutes per line of code are shown in Table 1.

	hours	**loc**	**ratio (min/loc)**
implementation	62.25	296	4.7
verification	95.25	330	19.31

Tab. 1: Effort analysis for implementation and verification of an OS scheduler in SPARK.

Regarding runtime verification of the remaining eight statically unverified VCs, the pragma `Assertion_Policy` was simply used to execute all contracts as assertions during runtime, but no exhaustive tests were yet carried out, nor have the predicates for runtime verification been optimized. The adaptation architecture has not been implemented yet.

3 Non-Functional Verification

Non-functional properties in the context of this work refer to the timing behaviour of a system. In order to reason about the timing of a system, a notion of time, a specification of timing properties and constraints and also a verification environment are needed. To formally verify timing behaviours means to find a mathematical proof showing that, under all conditions, the system will behave temporally as specified.

3.1 Modelling and Verification of Timing Behaviour

There is a long history of formal languages that can be applied to specify diverse aspects of timing behaviours [Wa04]. The classical event-oriented temporal logics such as Linear-Time Propositional Temporal Logic (LPTL) or Computation Tree Logic (CTL) only model the temporal order of events (e.g. before, after, always, never, eventually, ...), but do not provide a notion of real, physical time, as is needed to model real-time systems. One possible language to start with is called Timed CTL* (TCTL*). It is the foundation of the UPPAAL verification framework [20d], which is based on model checking techniques. However, such an approach often leads to state explosion or undecidability problems, making it impractical for complex real systems. Our method wants to avoid these limitations by keeping the human in charge as director for the proof, aided by semi-automated theorem provers like Coq [MT18].

3.2 WCET Estimation

In order to check whether a real-time program temporally behaves as specified in a model, it is necessary to know the actual execution times of relevant program sections, and to associate states in the model with program states.

The actual execution time of any piece of code can vary each time the code is executed. In real-time systems, the *Worst Case Execution Time* (WCET) is a commonly used concept to abstract from these variations. A good overview of the classification and techniques for WCET calculation can be found in [Ca19]. The paper differentiates between static, measurement-based or a combined approach to determine the WCET, each in a deterministic

or probabilistic variant. The static deterministic approach, called Static Deterministic Timing Analysis (SDTA), uses symbolic execution on an accurate model of the hardware. This approach is only practical for systems which are amenable to modelling, but it is well trusted and well established in industry. However, as today's multicore hardware achitectures frequently do not fulfil the assumptions made for their modelling, newer methods combine these approaches with probabilistic ones such as Extreme Value Theory (EVT). These are subject of ongoing research [Ca19].

3.3 Current Status

We evaluated the AbsInt tool for SDTA named aiT [20a] with a small application from the AQUAS space usecase and compared the WCET bounds with real measurements. The application cyclically receives a message from a serial communication interface, encrypts the message and sends it out over another serial communication interface. This application was executed on top of two different operating systems: (1) the library-based RTEMS kernel [20c] designed for microcontrollers and mostly used in avionic and space systems, and (2) our own microkernel Marron equipped with a small RTEMS adaptation layer which currently only implements the interface subset needed by the application. As target hardware, we use the TI TMS570 microcontroller with two ARM Cortex-R4 in lock-step mode. Marron was originally designed to run on Cortex-A multicore processors, but these more complex processors are not supported by aiT. The ARM Cortex-R4 was the smallest processor on which our system runs without modification.

	AIS2 code	**Infeasible routines**	**Analysis times**	**aiT WCET**	**Max Exec. Time**
RTEMS	659 loc	36	40 s	0.659 ms	0.321 ms
Marron	476 loc	23	4 s	0.580 ms	0.247 ms

Tab. 2: WCET and Measurement Results

The results of the WCET analysis are presented in Table 2. Annotations of the source code using the AbsInt AIS2 language are instructions to the AbsInt tools directing the static analysis. Declaring routines as infeasible means that the developer is sure they are never executed by the analysed code (e.g. POSIX and kernel error handling) and thus, they do not contribute to the estimated WCET.

We also compared the WCET estimations with real measurements of the execution times of the application. The execution time measurement starts upon reception of the first byte and ends when the last byte is sent out, thus being the same code sequence as for the WCET analysis. Both versions were compiled for ARM Thumb Code and with optimisation set to -Og (i.e. weak, "debug-friendly" optimisation). For the measurements we use the ARM Performance Monitoring Unit which measures the number of elapsed CPU cycles. All measured data was buffered in SRAM, as SRAM accesses are deterministic on the used platform. The buffering overhead was determined in a separate measurement and subtracted for compensation. For the static analysis, the buffering overhead was excluded by appropriate annotations. The measurement overhead itself was measured to be 60 CPU cycles based on 10 x 1000 single measurements. The aiT tool estimated the WCET for one

measurement to be 77 CPU cycles. The density function of the measured execution times for

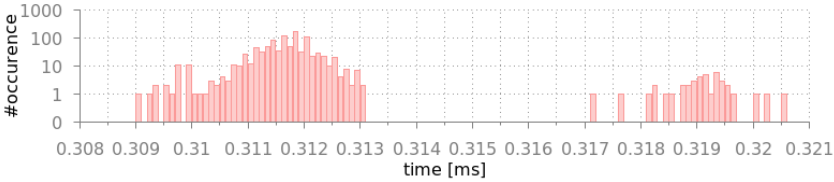


Fig. 3: Measurement Results for RTEMS

the RTEMS version is shown in Fig. 3, for the Marron version in Fig. 4. RTEMS execution times below 0.3131ms are caused by the communication through a software queue between the application and the receiver interrupt. The execution times above 0.317ms are the result of the same communication delayed by the system timer interrupt. This interference does not appear in the WCET analysis, as the tools do not model task communication or processor interrupts. The Marron RTEMS adaptation layer does not provide software queues for

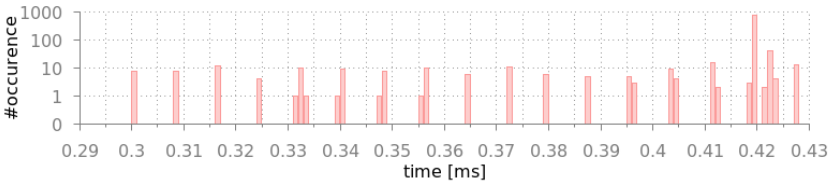


Fig. 4: Measurement Results for Marron

signalling, the only buffering mechanism used are the hardware FIFOs built into the serial interfaces. If no new data is available, the layer waits for the interrupt through a system call in user space. Execution times below 0.45ms result from rare cases of beneficial interference between the application and the system timer interrupt: If the application is interrupted before the hardware FIFO is checked, it is possible that the execution time will be shorter because new data arrived while the system timer interrupt was being processed. In this case, the data is ready upon exit from the timer interrupt and the receiving task does not need to wait for a receive interrupt.

Comparing the measurement results for RTEMS and Marron with the correspnding WCETs in Table 2 we can detect an overestimation of the WCET. This can be reduced up to a certain degree with many more annotations. Up to this point there were 120 hours spent in WCET analysis including a training period and meetings.

4 Conclusion and Outlook

In this paper, we presented the early stage of a method aiming at tool-supported verification to complex embedded systems, considering both functional properties and timing as a non-functional property. For functional properties, simple examples were used to evaluate parts of the methodology with promising results. However, more complex, realistic problems need to be considered and the methodology needs to be developed further. For timing properties, static WCET estimations were compared against measured execution times. It

turns out that static methods are difficult to apply to today's increasingly complex multicore hardware architectures, especially when these were not designed for determinism. In order to model worst-case behaviour for such architectures, very pessimistic assumptions need to be made, correspondingly leading to pessimistic WCET estimations.

To deal with these problems, monitors observing execution times could be generated from the timing specification and serve as a basis for supplementing static WCET estimation with runtime verification. This would lead to a similar approach as presented above for functional verification. Such a method would belong to the class of measurement-based probabilistic timing analysis methods in the sense of [Ca19]. A unified approach for the verification of functional as well as timing properties, supplementing static verification with runtime verification such that even complex systems remain controllable seems to be feasible.

Acknowledgements: This project has received funding from the Electronic Component Systems for European Leadership Joint Undertaking under grant agreement No 737475. This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation programme and Spain, France, United Kingdom, Austria, Italy, Czech Republic, Germany. This project has also received funding from the Federal Ministry of Education and Research (BMBF) under agreement No 16ESE0157. We would like to give special thanks to the people from AbsInt Angewandte Informatik GmbH for their support and Thales Alenia Space for the usecase application.

References

[20a] AbsInt aiT, Apr. 2020, URL: https://www.absint.com/ait/.

[20b] AQUAS EU Project, 2020, URL: https://aquas-project.eu.

[20c] RTEMS Real Time Operating System, Feb. 2020, URL: https://rtems.org/.

[20d] UPPAAL, 2020, URL: http://www.uppaal.org/.

[Ca19] Cazorla, F. J.; Kosmidis, L.; Mezzetti, E.; Hernandez, C.; Abella, J.; Vardanega, T.: Probabilistic Worst-Case Timing Analysis: Taxonomy and Comprehensive Survey. ACM Comput. Surv. 52/1, Feb. 2019.

[LNR80] Levitt, K. N.; Neumann, P. G.; Robinson, L.: The SRI Hierarchical Development Methodology (HDM) and its Application to the Development of Secure Software. In: Report 500-67. SRI International, Menlo Park, NBS, 1980.

[LS09] Leucker, M.; Schallhart, C.: A Brief Account of Runtime Verification. Journal of Logic and Algebraic Programming 78/5, pp. 293–303, May 2009.

[MT18] Mahboubi, A.; Tassi, E.: Mathematical Components, Creative Commons License, 2018, URL: https://math-comp.github.io/mcb/.

[Ru16] Rufino, J.: Towards integration of adaptability and non-intrusive runtime verification in avionic systems. ACM SIGBED Review 13/, pp. 60–65, Mar. 2016.

[Wa04] Wang, F.: Formal verification of timed systems: a survey and perspective. Proceedings of the IEEE 92/8, pp. 1283–1305, Aug. 2004.

Künstliche Intelligenz in der Umweltinformatik

R. Reussner, A. Koziolek, R. Heinrich (Hrsg.): INFORMATIK 2020,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2021 1005

1. Workshop Künstliche Intelligenz in der Umweltinformatik

Andreas Abecker[1], Julian Bruns[2], Stefan Naumann[3]

Abstract: In den letzten Jahrzehnten hat die Umweltforschung begonnen, eine datengesteuerte Perspektive einzunehmen, die durch riesige Sensornetze, satellitengestützte Erdbeobachtung und einen fast allgegenwärtigen Internetzugang ermöglicht wird. Von einigen dieser datengestützten Ansätze wird erwartet, dass sie Visionen einer nachhaltigen Zukunft umsetzen können. Zum Beispiel, indem sie es ermöglichen, in nachhaltigen intelligenten Städten zu leben oder die Welt mit „Smarter Landwirtschaft" zu ernähren. Oder indem man die Umweltverschmutzung oder die globale Entwaldung mit besserer Erdbeobachtung bekämpft. Es besteht jedoch eine Kluft zwischen einigen der derzeitigen Erwartungen, die in datengesteuerte Techniken gesetzt werden, und der Reife im Bereich des (räumlichen) Maschinellen Lernens und der Künstlichen Intelligenz. Im Workshop werden offene Forschungsfragen adressiert und Anwendungsbeispiele aus den Schnittfeldern von KI und Umweltanwendungen diskutiert.

Keywords: Umweltinformatik; Künstliche Intelligenz; Maschinelles Lernen; Umweltschutz; Nachhaltigkeit

1 Ziele und Motivation des Workshops „Künstliche Intelligenz und Umweltinformatik"

Die Umweltinformatik befasst sich interdisziplinär mit der Analyse und Bewertung von Umweltsachverhalten. Aus informationstechnologischer Sicht spielen dabei beispielsweise Simulationen komplexer Systeme, Geographische Informationssysteme (GIS) und räumliche Datenanalyse, Messnetze und Sensordatenverarbeitung sowie Fernerkundung und Bildverarbeitung eine große Rolle. Außer in der Wissenschaft findet die Umweltinformatik ihre wichtigsten Anwendungen in der öffentlichen Verwaltung (Natur- und Umweltschutz, Umweltdatenportale, Katastrophenschutz, Verbraucherschutz, Wassermanagement etc.), aber auch in der Wirtschaft (betriebliche Umweltinformationssysteme). Enge Bezüge und teilweise Überlappungen ergeben sich auch zur Agrar-, zur Hydro-, zur Energie- und zur Gesundheitsinformatik sowie im Bereich Green IT.

Die Umweltinformatik betrachtet in aller Regel sehr komplexe Prozesse in Ökosystemen, deren Verhalten (noch) nicht vollständig bekannt und verstanden ist, deren Verhalten nur approximativ oder vereinfachend modelliert, simuliert oder vorhergesagt werden kann

[1] Disy Informationssysteme GmbH, Ludwig-Erhard-Allee 6, 76131 Karlsruhe, andreas.abecker@disy.net
[2] Disy Informationssysteme GmbH, Ludwig-Erhard-Allee 6, 76131 Karlsruhe, julian.bruns@disy.net
[3] Hochschule Trier, Umwelt-Campus Birkenfeld, Institut für Softwaresysteme, Postfach 1380, 55761 Birkenfeld, s.naumann@umwelt-campus.de

doi:10.18420/inf2020_94

und bei deren Beobachtung häufig relevante Größen unbekannt sind oder nur geschätzt werden können. Zusätzlich müssen bei der Betrachtung von Ökosystemen noch weitere Systeme, die aus Informatiksicht nicht einfach zu behandeln sind, wie z.B. Wetter, betrachtet werden, da diese ebenfalls direkte und indirekte Einflüsse ausüben. Entscheidungen in Anwendungsfällen (wie z.B. Planungsverfahren, Notfallmanagement, Politikgestaltung) sind in der Regel schwierige Abwägungen und erfordern Fach- und Erfahrungswissen. Die entsprechenden Fragestellungen haben vielfältige Wechselwirkungen zu hoch aktuellen und enorm wichtigen Themen wie z.B. Klimawandel, Energiewende, Biodiversität und Nachhaltigkeit, aber auch zu großen Technologietrends wie Smart Cities, Smart Agriculture oder Smart Grids.

Die Anwendung von Methoden und Technologien der KI drängt sich also auf. Für den Workshop KIU-2020 haben wir Forscher, Entwickler und Anwender eingeladen, gemeinsam ihre Fragestellungen, Lösungsansätze und Ergebnisse zu intelligenten IT-Ansätzen für Umweltanwendungen zu diskutieren.

Aus technischer Sicht sollte die Umwelt-KI die gesamte Breite intelligenter Software-Ansätze, also aus symbolischer und subsymbolischer KI, intelligenter Datenanalyse, maschinellem Lernen usw. betrachten.

2 Thematische Schwerpunkte

Der Workshop war bewusst breit angelegt und umfasste einerseits KI-Technologien für Umwelt und Umweltschutz und andererseits Anwendungen aus dem Bereich Umwelt, die durch KI unterstützt werden können.

Mögliche KI-Technologien waren (nicht ausschließliche Liste):

- Big / Smart / Linked / Open Data
- Bildverarbeitung und Fernerkundung
- Data Mining, Machine Learning, Deep Learning
- ELSI-Aspekte zur Umwelt-KI, Responsible AI
- Explainable AI
- Human-Centered AI
- Multiagentensysteme
- Multimodale Interfaces, AR, VR und KI
- Robotik
- Semantische Technologien

- Soft Computing / Computational Intelligence
- Spatial Data Mining, Spatio-Temporal Data Analytics
- Sprachverarbeitung
- Unsicherheit und Vagheit
- Wissensbasierte Systeme

Mögliche Umweltanwendungen mit KI-Unterstützung umfassen:

- Betriebliche und behördliche Umweltinformationssysteme
- Biodiversität
- Erneuerbare Energien und Energiewende
- Green IT und Energiemanagement
- Katastrophenschutz und -management (aus Umweltsicht)
- Klimawandel
- Natur- und Umweltschutz
- Ressourcenschutz und Landmanagement
- Smart Agriculture (aus Umweltsicht)
- Smart City (Umweltaspekte)
- Umweltbildung
- Verbraucherschutz (Umweltaspekte)
- Wasser 4.0

Auch übergreifende Aspekte wie die Energie- und Ressourcenverbräuche durch KI-Training und -Anwendung sind durch die Workshop-Themen angesprochen.

3 Eingereichte Fachbeiträge

Insgesamt wurden für den Workshop 8 Fachbeiträge eingereicht, die aufgrund ihrer Qualität auch sämtlich angenommen wurden. Die thematische Bandbreite reicht dabei von Vorgehensmodellen und Anwendungsszenarien von KI im industriellen Bereich über die Frage der Vorhersage der Vegetationsentwicklung nach Naturkatastrophen wie Waldbränden sowie der generellen Behandlung von Zielkonflikten in der Landnutzung bis hin zu der Frage der Überwachung von Chemikalien hinsichtlich ihrer Wirkung und Ausbreitung sowie der

allgemeinen Prognose von Wasserverbräuchen. Methodisch wurde hier überwiegend auf Maschinelles Lernen gesetzt, was aufgrund der aktuellen Attraktivität und Verbreitung nachvollziehbar ist. Inwieweit symbolische Verfahren zukünftig eine stärkere Rolle spielen, bleibt abzuwarten. Neben den Fachbeiträgen konnte Prof. Dr. Martin Wagner von der Technischen Universität München als Keynote-Speaker für den Beitrag „Computational Challenges for Artificial Intelligence and Machine Learning in Environmental Research“ gewonnen werden.

4 Programmkommitee

- Dr. Ansgar Bernardi; Deutsches Forschungszentrum für Künstliche Intelligenz, Kaiserslautern; https://www.dfki.de/
- Dr. Matthias Budde; Karlsruhe Institut für Technologie, Karlsruhe; http://www.kit.edu/
- Prof. Dr. Frank Fuchs-Kittowski; Hochschule für Technik und Wirtschaft, Berlin; https://www.htw-berlin.de/
- Dr. Desirée Hilbring; Fraunhofer IOSB, Karlsruhe; https://www.iosb.fraunhofer.de/ Julian Huber ; FZI Forschungszentrum Informatik, Karlsruhe; https://www.fzi.de/
- Dr. Christian Jolk; Ruhr-Universität Bochum, Bochum; https://www.ruhr-uni-bochum.de
- Prof. Dr. Gerlinde Knetsch; Umweltbundesamt, Dessau-Roßlau und HTW, Berlin; https://www.umweltbundesamt.de
- Dr. Sven Lautenbach; Universität Heidelberg, Heidelberg; https://www.geog.uni-heidelberg.de/
- Dr. Tanja Liesch; Karlsruhe Institut für Technologie, Karlsruhe; http://www.kit.edu/
- Dr. Martin Memmel; Deutsches Forschungszentrum für Künstliche Intelligenz, Kaiserslautern ; https://www.dfki.de/
- Prof. Dr. Jens Nimis; Hochschule Karlsruhe Technik und Wirtschaft, Karlsruhe; https://www.hs-karlsruhe.de/
- Dr. Steffen Thoma; FZI Forschungszentrum Informatik, Karlsruhe; https://www.fzi.de/

R. Reussner, A. Koziolek, R. Heinrich (Hrsg.): INFORMATIK 2020,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2021 1009

Computational Challenges for Artificial Intelligence and Machine Learning in Environmental Research

Martin Werner[1], Gabriel Dax[2], Moritz Laass[3]

Abstract: In the last decades, environmental research has started to adopt a data-driven perspective enabled by huge sensor networks, satellite-based Earth observation, and almost ubiquitous Internet access. Some of these data-driven approaches are expected to make visions of a sustainable future come true. For example, by enabling societies to live in sustainable smart cities, or to feed the world with precision agriculture. Or by fighting environmental pollution or global deforestation with increased observational power. However, there is a serious gap between some of the current expectations put into data-driven techniques and the maturity of the field of spatial machine learning and artificial intelligence or computer science in general. We give a few examples of open research issues that computer science has to solve in order to make data-driven approaches to environmental sciences successful.

Keywords: Environmental Sciences; Computer Sciences

1 Introduction

We live in a world that is in severe danger due to issues with the environment. The human population has been abusing the planet for a very long time, taking more than giving back to the ecosystem. But in the last decades, our understanding of the impact that humankind has on the planet has increased and we are on the way trying to mitigate some of these effects. For example, the world has been able to agree on certain goals and time frames for concrete reductions in the amount of CO2 emission. In a more general way, we see a trend that the broad public starts taking care of environmental questions more than ever before. Not only international movements like "Fridays For Future" or small countries proceedings against industry countries, as well global political institutions like the United Nations have clearly identified the problem of sustainable development.

The United Nations Sustainable Development Goals identify 17 areas depicted in Figure 1 in which action needs to be taken for a "shared blueprint for peace and prosperity for people and the planet, now and into the future." [Na].

[1] Technical University of Munich, Professorship of Big Geospatial Data Management, TUM Faculty Aerospace and Geodesy, Ottobrunn, Germany, martin.werner@tum.de
[2] Technical University of Munich, Professorship of Big Geospatial Data Management, TUM Faculty Aerospace and Geodesy, Ottobrunn, Germany, gabriel.dax@tum.de
[3] Technical University of Munich, Professorship of Big Geospatial Data Management, TUM Faculty Aerospace and Geodesy, Ottobrunn, Germany, moritz.laass@tum.de

doi:10.18420/inf2020_95

Fig. 1: U.N. Sustainable Development Goals

Many of these United Nations Sustainability Goals are related to data. For example, the goal "Zero Hunger" is of course related to global food production and sharing, which is a complex process whose efficiency is largely depending on data. Similarly, it is expected that sustainable cities and communities will need data-intensive services to solve some issues related to density in urban regions like smart transport or air pollution. As a final example, the goal "Climate Action", spelled out "Take urgent action to combat climate change and its impacts", will depend largely on big data provided by Earth observation satellites in order to quantify effects related to climate. In this context, the European Space Agency (ESA) is running research and development in the climate change initiative[4] worth exploring. It currently contains global data products related to important climate variables such as aerosols, biomass, clouds, fire, glaciers, ice shields, land cover, land surface temperature, sea level, sea ice, soil moisture and much more. Without going into details, one realizes that these global big data products are important to the environment and should be exploited as much as possible in data-intensive systems. This links environmental research to computer science, statistics, and machine learning or more general to data science.

For this paper, data science is the culmination of knowledge from three important aspects: mathematics and statistics, computer science, and domain knowledge [WF17]. In this paper, we want to discuss some computer science issues that must be addressed in the context of environmental research for an overall increasing ability to understand, control, manage, and solve issues related to the environment.

[4] see http://cci.esa.int/

2 Data Acquisition, Representation & Compression

In order to come up with an overview of how computer science contributes to environmental research, we first discuss two different types of data because they pose very different challenges to the computer science community. The first type captures all types of measured or scientific data, for example, collected by professional controlled surveying methods with reliable and high-quality sensors. The second type of data subsumes all sorts of user-generated data including volunteered geographic information (VGI) as well as social media or web sources. We will shortly characterize these two types of data in this section.

2.1 Measured Data and Scientific Data

Due to the development of cheap sensors and computing, all domains of engineering generate increasing amounts of data. And in addition, the domains of engineering that design and build sensors are increasing both accuracy and quantity of information gathered from sensors every year. One domain very relevant to environmental research is the domain of remote sensing, especially from space. In remote sensing from space, satellites sense information about the Earth and this information can be used to quantitatively measure on the Earth surface. One very important type of satellites employs multispectral cameras to capture images of the Earth surface. These satellites are conceptually similar to RGB cameras, except that they can capture more different spectral bands (e.g., colors) and that advanced techniques are needed to georeference images taken from space and to mitigate atmospheric distortions including cloud cover and cloud shadows. These satellites typically orbit around Earth with slowly rotating orbits and are, therefore, capable of taking images everywhere on the planet from time to time. Similarly, synthetic aperture radar (SAR) satellites are routinely used in Earth observation. They have advantages such as that they are not affected by clouds and generate more continuous observations and they are capable of surprising measurements using a technique called interferometry in which changes in small scale on the Earth surface can be observed. In both cases, engineers have made sure and are continuously working on high-quality data processing including proper estimation and calibration of all system parameters. The resulting data is huge, but highly accurate.

Another important domain generating measured and scientific data is the area of autonomous driving currently under development. In this area, cars are employed with high-quality, certified sensors to allow vehicles to navigate unknown environments. These sensors include cameras, radar systems, and laser scanners. Together, these systems are assumed to enable safe vehicle operation in unknown environments. However, the amount of data that current development systems capture is extremely huge and complicated as well.

A third domain generating huge amounts of data are our communities and businesses. For example, cities are routinely measuring traffic information, bus locations, taxi data, electricity information, ranging up to large installations of video surveillance in some cases.

In addition, mobile networks are capturing presence and mobility information of mobile devices almost everywhere in our complex cities.

In general, these data sources are well-designed, well-understood and pose questions mainly related to ethics, data size, and data ownership.

3 Personal, Human-generated and Societal Data

Complementary to such measured sources of big data for environmental science are sources originating more in human and societal contexts. This includes news streams, social media messages, human-curated knowledge such as OpenStreetMap and Wikipedia, opinionated data sources such as blog posts from certain platforms, or blind web scale data collections such as common crawl. This type of data has significant limitations with respect to its quality measured in aspects such as bias, risk of abusive publication of information (fake news, etc.), data quality, and data completeness. Though the Internet and social media act as a mirror of aspects of our societies due to discussions taking place on major platforms such as Facebook, Twitter and major news platforms, one is tempted to forget that the Internet does only represent a small fraction of our society and that people behave and communicate differently from real life.

4 The Evolution of Big Data for Environmental Research

With these preparations on two different types of data relevant to environmental research, we focus on computer science aspects in this chapter and first match the difference between these two types of data with concepts of big data. Big data is a family of techniques evolving in order to solve the challenges posed by current data collections on computational infrastructures. One widely accepted definition of big data relies on three aspects Volume, Variety, and Velocity (3V). Though there are many additional Vs such as veracity, value or vagueness being used (even in standards), the advantage of the 3V definition is that all three aspects can be quantitatively measured to some extent. With volume, one refers to the sheer amount of data. Big data is a situation in which traditional computational techniques (relational database management systems, file systems, personal computers) fail due to the amount of data being too large. With velocity, one refers to the situation in which data arrives or needs to be moved faster than feasible in a single computer. Volume and velocity are tightly interwoven as processing high-volume data sets in distributed computing leads to high volumes of data being communicated in short time, that is high velocity. And vice versa, storing high velocity data leads to high volume collections. The third aspect of big data, variety, is represented by the high number of ways, data can be organized, modeled, and stored. This includes aspects such as file formats or data representation as well as organizational aspects such as data ordering and data distribution. Some general aspects of big data are that pre-processing is usually considered impossible or ineffective. That is, most big data systems deal with “raw” data in their “natural” ordering and format.

In environmental research, big data issues are currently limiting the exploitation of data by scientists. For example, even the open access satellite data published by ESA (e.g., Sentinel satellites) or U.S.G.S. (e.g., Landsat satellites) is difficult to acquire, store and use in practice due to the size of the data. This practically limits very large (e.g., global analysis) to well-funded research agencies and companies.

Similarly, the data generated by mobile sensors in the public (cameras, cars, etc.) is very difficult to share in a research community due to legal issues related to privacy. In practice, this means that mainly car manufacturing companies or large research institutes can do their own data acquisitions to base their research on. There are some ongoing activities to share benchmark data and lately even car manufacturers support research with publishing data, but the availability of data remains limited. Due to recent abuse such as during democratic elections all over the world, social media APIs are limited. That is, the whole source of data related to social media is difficult to access without cooperating with social network companies. And there are significant ethical and legal challenges to keep in mind when working with or on social media data. Because even if you think that a given data collection is ethically sound for a certain research purpose, one must make sure that the data collection is not being used in another way now or in the future. In addition, one must think of mechanisms to curate such datasets with updated information on the user's intent. For example, when observing a social media message on one day, one should not use it anymore as soon as the original user has deleted the post. In fact, one should delete the message as well.

5 Open Computer Science Challenges

In the context of challenges related to Big Data, there are several computer science aspects that need further research in order to increase the adoption and reduce the barriers inherent to the amount of data.

The first domain of research is **compression**. Currently, the most valuable data for environmental sciences is generated by high-quality sensors and the data distribution mechanisms do their best not to compromise any of this quality. For example, satellite images are published with lossless compression, GPS data is published in non-simplified form, car sensor information logs are published without any form of processing in order to not sacrifice value. However, the authors argue that compromising volume and data quality is essential to sustainable application and open science. Research is needed on how to compress data such that (1) the amount of researchers that are financially able to study data at large scale is increased and (2) that the application of computing to the data is not wasting too much electrical energy for results that would have been possible from highly-compressed data extracts as well.

Data compression is coupled with **information theory** and information theory can be used as a tool to integrate compression and algorithms and there is some research already going

on. One starting point is possibly the area of compression distance and feature-free data mining [CV05] or some work on spatial data representation with probabilistic sketches [We15; We19].

Another direction of computer science research for environmental research and beyond is to research **efficient algorithms**. By reducing the computational complexity of algorithms, one decreases the overall need for computational capacity and can consume larger amounts of data in shorter time either reducing the amount of parallelization or increasing on the limits of a given infrastructure.

Furthermore, a subdomain of database research is research on **reasoning and managing data under uncertainty**. Uncertain data is very common in big data and it is a pity that though theoretical models of uncertain data management have been researched for a long time, the amount of practicable techniques from this field is small. More research needs to be done with respect to handling uncertainty in a scalable way.

As many of the current environmental challenges are rooted in human behavior, it is essential that we are observing human activities. However, collecting precise data about human behavior limits the freedom of individuals and is – for good reasons – not allowed in democratic countries. One has to make sure that the individual rights of people being observed stay protected as much as possible and that the environmental research application's societal value is higher than the privacy loss of individuals. It is a very difficult ethical area and one contribution from computer science is to design **privacy-preserving algorithms**. These are distributed multi-party algorithms in which the data contribution of individuals (e.g., people, sensors, etc.) is protected while sufficient aggregate results or results of computation is still possible. One widely accepted framework is the framework of differential privacy in which data is perturbed when it is produced (e.g., at a sensor) in a way that is statistically cancelling in aggregates.

It is clear that we will be using more and more measured data in decision-making processes as a consequence of "digitization". But we need to be very careful as it is unclear how resilient against fake data such decision-making processes are. This opens a research area of **resilient distributed algorithms** in which algorithms are required to work even in case of attackers. This includes topics ranging from **resilient computer vision to the cyber security** of individual nodes and components in a future Internet of Things.

Tightly coupled with resilient distributed algorithms is the domain of **correctness and verification**. How can we organize all data processes in a way such that the behavior of a system adheres to a given specification. In fact, assessing the correctness of distributed algorithms is a very challenging topic and also less strict aspects such as debugging and testing are complicated due to the huge amount of possible situations the system might face at runtime.

6 The Evolution of Machine Learning and AI for Environmental Research

Big data forms the foundation of machine learning and artificial intelligence and complex methodologies from these fields usually consume a lot of data leading to a strong link between big data and modern machine learning. The current state of artificial intelligence leads to extreme expectations in general as it has been solving hard problems in narrow domains quite successfully. However, it is unclear whether these expectations can be fulfilled outside a few narrow domains. Deep learning and modern artificial intelligence has unquestionably had significant impact on computer vision and image recognition, on text understanding and speech recognition, on dialogue systems and in handwriting recognition, and in playing games. It has as well been applied to subdomains of environmental science such as in remote sensing and cartographic applications [Sa20]. However, traditional methods such as decision trees or support vector machines still dominate these fields for the reason that the amount of available training data is very small compared to the variety and complexity of real-world classification tasks in these data domains. Therefore, a lot of research on combating overfitting and reducing the information-theoretic complexity of input data (data simplification, dimensionality reduction) is still needed in order to see significant adoption of deep learning in these complicated real-world tasks.

This is where the workshop on "Künstliche Intelligenz in der Umweltinformatik" (artificial intelligence for environmental sciences) will put its emphasis: with which contributions from the computer science domain can we enable environmental scientists to apply the unquestionable power of deep learning and artificial intelligence in a promising, scientifically rigid, valuable and sustainable way? Personally, the authors fear that the discrepancy between narrow AI and research papers based on narrow AI in fields related to environmental sciences and real-world deployability and generalization capabilities of such models might lead to the next AI depression in which people realize that artificial intelligence is not a magic tool unless you understand and model your data and your tasks rigorously and come up with techniques related to solving the most pressing problems of applying deep learning in the spatial domain.

7 Open Computer Science Challenges

Let us again list some fields of computer science that will play an essential role in the further development of artificial intelligence for environmental sciences. One important aspect holding back the wide adoption of artificial intelligence is the fear that something goes wrong. As long as we cannot assess the decisions made in artificial intelligence, we can never be sure to base decisions and political policies in such results. That is, without **explainable artificial intelligence**, adoption of artificial intelligence in environmental sciences remains an academic exercise. Related to this explainability is the question of dependability, which is also linked to the computer science challenge of verification. A

dependable artificial intelligence system is a system where one can safely base decisions in. Though dependability and explainability are interrelated concepts, they differ very much. While explainability just asks for human-understandable reasons for decisions, the decisions might be wrong or might be based on tampering. With dependable AI, one goes one step further that the system makes sure that the data is real data and not the result of tampering with an AI system.

The most difficult and most important area of research for a wide adoption of deep learning in environmental sciences and spatial information sciences in general is, however, more general the question of whether and how spatial processes can be captured in deep learning settings. This amounts to understanding, estimating and taking into account all sorts of **spatial autocorrelation** and also mitigate issues related to structures introduced by space subdivisions such as **gerrymandering**. These are very hard problems where only some initial success is known from spatial statistics and where there are quite a few negative results related to that current machine learning approaches are not able to learn spatial analysis tasks unless they have trivial autocorrelation structure. Given this hardness, this is an exciting area of research. Solving or mitigating some of these issues is really a major step forward for data-driven environmental science and beyond.

But even today, machine learning can be applied to spatial data and in order to reduce the risks of domain scientists in claiming wrong results, we – as computer scientists – should come up with software, techniques and analyses of best practices: how can you use machine learning in a spatial domain? How can you avoid many of the pitfalls of transferring from the narrow domains of image recognition or language analysis to the wide and complicated domain of modelling human behavior and the Earth system? Best practices are needed as well as some textbooks and major reference works to spread the word about how to correctly and carefully exploit the advances of machine learning in environmental science. And as the risks of doing something wrong or suboptimal when dealing with spatial data are so high, we call for a culture of reproducible research. We should not anymore accept any papers that just claim a certain achievement without a stringent and open scientific debate based in data and software. And for this, we should concentrate on topics such as compression or privacy-preserving data mining as enablers of open science in computational environmental science.

8 Conclusion

Machine Learning and Artificial Intelligence has shown great advances in certain domains in the last decade. Unfortunately, this leads to expectations in other domains with much more involved or complicated situations. In summary, this paper tried to remark that many of these complications are related to computer science problems and that significant contributions of computer scientists are needed aside domain expert knowledge and engineering in order to allow a responsible and sustainable exploitation of machine learning and artificial intelligence for environmental sciences.

As a conclusion, we see computer science research needs working on **(1) reducing the data footprint, (2) reducing the computational complexity, and (3) reducing abusive or wrong adoption of techniques** leading to claims that are too optimistic or even just wrong. To some extent, the computer science community is responsible for coming up with solutions and guidance in these three directions as spelled out in more detail in the chapters.

We are convinced, that artificial intelligence and machine learning are integral parts of solving the environmental challenges we are facing by understanding and monitoring the changes to the Earth system as well as by allowing more sustainable behavior in large scales through ideas such as smart cities. Therefore, we hope that this first iteration of the workshop "Künstliche Intelligenz in der Umweltinformatik" (artificial intelligence for environmental science) will serve as a focal point of research and development with respect to the computer science contributions needed for environmental sciences and, thereby, for a sustainable future.

References

[CV05] Cilibrasi, R.; Vitányi, P. M.: Clustering by compression. IEEE Transactions on Information theory 51/4, pp. 1523–1545, 2005.

[Na] Nations, U.: U.N. Sustainable Development Goals, URL: https://sustainabledevelopment.un.org/?menu=1300.

[Sa20] Salcedo-Sanz, S.; Ghamisi, P.; Piles, M.; Werner, M.; Cuadra, L.; Moreno-Martínez, A.; Izquierdo-Verdiguier, E.; Muñoz-Marí, J.; Mosavi, A.; Camps-Valls, G.: Machine learning information fusion in Earth observation: A comprehensive review of methods, applications and data sources. Information Fusion 63/, pp. 256–272, 2020.

[We15] Werner, M.: BACR: Set Similarities with Lower Bounds and Application to Spatial Trajectories. In: 23rd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (ACM SIGSPATIAL 2015). 2015.

[We19] Werner, M.: GloBiMaps - A Probabilistic Data Structure for In-Memory Processing of Global Raster Datasets. In: 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (SIGSPATIAL '19). 2019.

[WF17] Werner, M.; Feld, S.: Successful Data Science Is a Communication Challenge. In: DIGITAL MARKETPLACES UNLEASHED. 2017.

R. Reussner, A. Koziolek, R. Heinrich (Hrsg.): INFORMATIK 2020,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2021 1019

Applying a deep learning-based approach for scaling vegetation dynamics to predict changing forest regimes under future climate and fire scenarios

Werner Rammer[1], Rupert Seidl

Abstract: The ability to anticipate future changes in terrestrial ecosystems is key for their management. New tools are required that bridge the gap between a high level of process understanding at fine spatial grain, and the increasing relevance for management at larger extents. Such a tool is SVD (Scaling Vegetation Dynamics), a scaling framework that specifically uses deep learning to learn the behavior of detailed vegetation models in response to different environmental factors. This trained deep neural network (DNN) is then applied within the framework on large spatial scales. In addition, SVD includes also explicitly modelled processes such as fire disturbances. Here we use the framework to simulate forest regime change in the 3 Mio. ha landscape of the Greater Yellowstone Ecosystem. We used four climate change scenarios and pre-defined fire events from statistical modelling, and analyzed whether prevailing forest types are able to regenerate after fire. Our results show that up to 60% of the area may undergo regime change until the end of the 21st century.

Keywords: SVD; vegetation dynamics; deep learning; Greater Yellowstone Ecosystem; fire; climate change

1 Introduction

Terrestrial vegetation is of crucial importance for human well-being and provide a wide variety of ecosystem services to society [As05]. However, vegetation is not static but changes dynamically, responding to drivers such as land-use change and climate change [Er18, Li10]. Thus, the ability to faithfully predict future trajectories of vegetation development is highly relevant for decision makers and society. Dynamic global vegetation models (DGVMs) are frequently used to simulate vegetation dynamics at large spatial scales. Such models increasingly include structural details (e.g., representing leaves or individual trees as entities of the simulation), but they typically assume that these structures represent the conditions of an entire grid cell (with a cell size usually between 10 and 250km). Therefore, biotic interactions such as seed distribution, mortality, or plant competition are neglected. Stand and landscape-level vegetation models – on the other hand - simulate vegetation demography on a detailed level and include biotic interactions (e.g., mortality and demography), as well as spatiotemporal controls (e.g., migration and legacies). However, they usually are limited to small spatial extents, which limits their application on larger scales. Since biotic interactions

[1] Ecosystem Dynamics and Forest Management Group, Technical University of Munich, 85354 Freising, Germany, werner.rammer@tum.de

doi:10.18420/inf2020_96

and the resulting demographic structures are crucial for many ecological questions such as carbon storage [Kö17], approaches that are able to capture the drivers of vegetation development at small scales and dynamically scale ecosystem dynamics across spatial domains are needed.

New methods in the field of artificial intelligence and machine learning excel at identifying structure in complex, nonlinear data, and generate accurate predictive models [GBC16]. Specifically, deep learning is an emerging machine learning technique at the core of recent breakthroughs in computer vision, speech synthesis, autonomous driving, and other fields [LBH15] and has been advocated for providing new opportunities in Earth Sciences [Re19] and Ecology [RS19a].

One approach for the scaling-up of detailed simulation models is the Scaling Vegetation Dynamics (SVD) framework [RS19b]. The modeling framework uses a deep neural network to learn the response of detailed vegetation simulation models to different environmental factors. This trained "meta model" can then be applied in a computationally efficient manner for projecting vegetation dynamics at regional to continental scale.

An important driver of future changes in forest ecosystems are disturbances, particularly as they are expected to increase in the future with a warming climate [Se14]. For example, the Greater Yellowstone Ecosystem (GYE) is an epitome of the complexities of climateand fire-driven vegetation changes. Situated in the Northern Rocky Mountains of the USA, it has received considerable attention due to 709,000 ha of wildfires affecting the system in 1988. Alarmingly, fire projections for the region suggest that the extreme event of 1988 could become the new normal at the end of the 21st century [We11]. It thus remains unclear whether changing climate and fire regimes will exceed the resilience of the GYE.

In this contribution we describe the deep learning-based modelling approach and present an example application. In this example, we assessed the probability of regime shifts in the forests of the Greater Yellowstone Ecosystem, given the expected future climate and fire regimes. Regime shifts were defined as changes in major forest types, with the inability of a current forest type to regenerate under future conditions used as an indicator for impending changes in forest type. Specifically, we asked how much of the currently prevailing vegetation experiences permanent regeneration failure.

2 Material and Methods

2.1 Scaling Vegetation Dynamics

Conceptually, the scaling vegetation dynamics (SVD) framework [RS19b] follows a state and transition approach, where vegetation is classified into discrete vegetation states and transitions between states are probabilistic. In SVD probabilities are conditional on environmental conditions and the local neighborhood of a cell and estimated by a deep

neural network [LBH15]. In this example, the DNN at the core of SVD was trained on data generated by the process-based model iLand [Se12], which was used to simulate regeneration success after fire disturbance over a wide range of environmental conditions. Note that for other applications, the DNN can be trained to encapsulate the full dynamics of the simulated vegetation [RS19b]. The application of SVD consists of three distinct phases: the first phase comprises the generation of training data, and the second the training of the DNN, yielding a condensed meta-model of the post-fire response of the process-based model. The third phase applies the trained DNN for the dynamic simulation of vegetation transitions for the whole GYE within the SVD model.

The spatial scale of SVD is 1 ha, and the time step is annual. The SVD model is extended by modules to encompass other drivers of vegetation transitions such as natural disturbances or ecosystem management. These modules provide additional pathways of vegetation transitions and can interact with each other. Here we use the fire disturbance module that simulates fire spread dynamically on the landscape and uses data on fire size, ignition point and time as input.

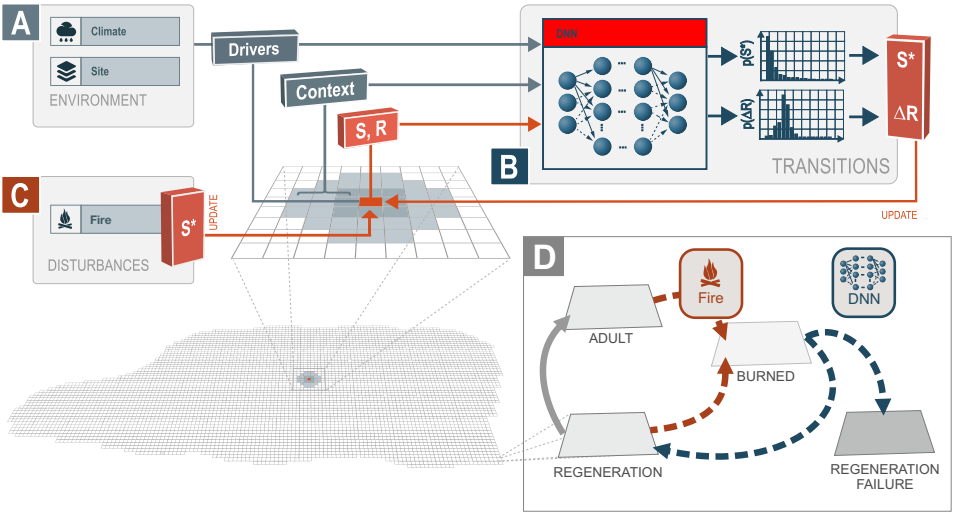


Fig. 1: Conceptual view of the scaling vegetation dynamics (SVD) framework (A-C), and the state and transition pathways (D) used in this study. In SVD, transitions on a single cell are predicted by a Deep Neural Network (B) and depend on environmental drivers (A), the current state (S) and residence time (R), as well as the spatial context (here: distance to seed source). Fire (C) as simulated by the SVD fire module adds an additional pathway of state change. States and transitions (D): Adult and regenerated cells transition to early seral states due to fire, and the regeneration success or failure is consequently determined by the DNN. The transition from regeneration to adult is deterministic.

Fig. 1 shows the conceptual pathways of vegetation transitions used in this study. Cells with vegetation of different forest type can be affected by fire. The regeneration success of recently burnt cells (“Early seral”) is determined by the DNN, considering environmental

conditions and distance to seed source within the dynamic simulation. Cells transition to a “Regeneration” state in case of success, or to the state “Regeneration failure” when regeneration fails to establish for more than 30 years. “Regeneration” cells become seed-producing “Adult” deterministically at a forest type specific maturation age.

SVD is a standalone software written in C++ that uses the C++ API of the deep learning library TensorFlow [Ab16] to perform DNN predictions (Fig. 2). By integrating the full TenorFlow framework into the model, the full functionality of TensorFlow can be used and the structure of the DNN can be tailored to the specific needs of each application. The DNN training itself facilitates the standard TensorFlow workflow and tools (e.g., Python, Keras, TensorBoard). When the training has finished, the full DNN (structure and weights) is saved and consequently used by the SVD model(Fig. 2) .

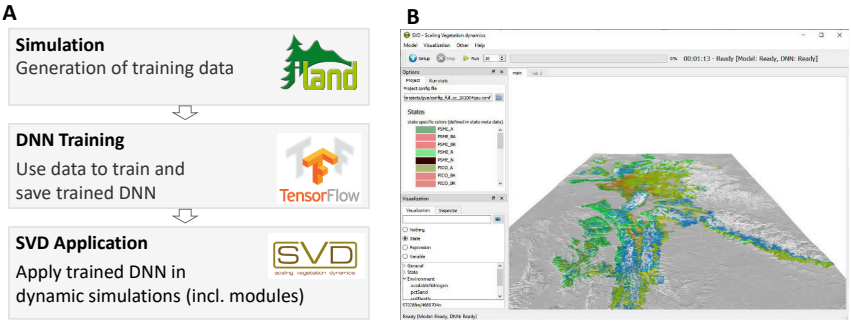


Fig. 2: Conceptual steps of a SVD application (A) and a screenshot from the SVD model (B). The training data for the DNN is generated with a detailed process -based model and then the DNN is trained with Te nsorFlow. The final model is saved and consequently used in the SVD model in the context of a dynamic simulation (including the fire module). DNN predictions are executed by TensorFlow which are triggered via the C++ API of the library.

2.2 Generation of training data

We followed the approach of [Ha18], who used the individual based forest landscape and disturbance model iLand [Se12] to analyze the conditions under which the most important forest types in Yellowstone National Park might fail to regenerate after stand replacing fire disturbance. iLand simulates individual trees within a stand and uses a hierarchical framework wherein broader-scale processes emerge dynamically from interactions among

individual trees. The model represents tree growth, mortality, and competition in response to canopy light interception, radiation, thermal conditions, soil water, and nutrient limitation. While climate and soil conditions are considered at spatially homogeneous within a stand (1 ha), variation in light is simulated at 2 m horizontal resolution based on overstory structure and composition. Climate data (temperature, precipitation, radiation, vapor pressure deficit) is considered at a daily temporal grain. The model has been well tested and extensively used in the western United States [SRS14] and Europe [TRS17], and has recently been parameterized and evaluated for the Yellowstone area [Br18]. The model explicitly simulates tree regeneration based on seed production, seed dispersal, and effects of temperature, light, and soil-moisture conditions on seedling establishment and survival.

For the generation of the training data we set up a factorial simulation experiment focusing on factors that are likely to change under future climate and fire regimes [We11], and strongly affect regeneration success by altering either the availability of seeds or the establishment success of tree plants. We considered in the experiment fire return interval (FRI), distance to seed source, and variation in climate, and applied all combinations for four forest types on 1,296 representative sites across the region. We used four levels of FRI (11, 20, 50, 100 years) that cover the range from shortened future fire intervals [We11] to the upper end of historically observed FRI. The availability of seeds was represented by setting the distance to the nearest seed source between 50 m and 1250 m (50 m stepwidth). Finally, unfavorable climate conditions can hamper seedling establishment, predominantly due to drought in post-fire years [HT19]. We therefore repeated the simulations assuming either warmer, or warmer and drier conditions consistent with our climate change scenarios for the 21st century.

2.3 Training the deep neural network

The data derived from the iLand simulation experiment was then used to train the deep neural network that is used in dynamic SVD simulations. The network learned to predict the success or failure of regeneration contingent on FRI, distance to seed source, site and climate conditions (Tab. 1). We experimented with different DNN architectures and hyper-parameter settings and monitored the predictive performance of the network using a fixed subset of the training data set aside for evaluation (11% of the examples). The best-performing network was consequently used in the SVD model for dynamic simulations across the whole region.

2.4 Simulation setup

We simulated the vegetation development in GYE under four different climate change scenarios from 2005 to 2100. We selected scenarios from the RCP (Representative Concentration Pathways) 4.5 and RCP 8.5 emission scenario families; RCP 4.5 is considered as an intermediate scenario where CO2 emissions peak in the 2040s and decline afterwards, while

Tab. 1: Description of the data used for the training of the DNN.

Variable	Description
Predictors (input data)	
CLIM	Climatic conditions (mean monthly temperature (°C) and precipitation (mm) in the next ten years from the current year, i.e. 2 x 12 x 10 = 240 values)
CSS	Site conditions (Nutrient supply (with plant available nitrogen as proxy, in kg N ha^{-1} yr^{-1}), soil depth (m), % sand (proxy for soil texture))
DSS	Distance to seed source (m)
R	Residence time (yrs), the number of years the stand is already in state S
S	Current state (-)
Response variables (labels)	
S^*	Predicted next state (may be equal to *S*)
ΔR	Time until state change (up to 10 years)

RCP 8.5 represents rising emissions throughout the 21st century [Me11]. Global climate models still do not agree on future precipitation trends in the northern Rocky Mountains and we integrated this uncertainty by selecting two diverging scenarios for each RCP. Predictions of fire activity (fire locations and fire sizes) were derived from statistical fire modelling [We11] and were available as 20 replicated time series of fire for each scenario. Data on the current vegetation as well as site and climate conditions were compiled from available data sources for the entire area.

3 Results

In order to generate the training data for the machine learning algorithm in SVD, we ran the factorial simulation experiment in iLand, yielding a total of 2.59 Mio individual stand trajectories each including a stand replacing fire and post fire vegetation development from which a total number of 7.7 Mio training examples were extracted (see Tab. 1).

The final network architecture (i.e., the type, shape, and number of layers) used for this study was a feedforward network with 674,420 trainable parameters (Fig. 3). A series of fully connected layers compressed the information content of the climate input variables (CLIM, N=240) to a vector with ten elements. An embedding layer [5] was used to transfer the numerical state identifier to an embedding vector with six dimensions. Climate and state information were then combined with the remaining input data (site variables (CSS), distance to seed source (DSS), residence time (R)). From the concatenation layer two branches of fully connected and dropout layers led to the final Softmax layers with the classification result for the next state S^* (23 classes), and residence time ΔR (10 classes). See [An16, GBC16, Me11, RS19a] for more details on network architecture and the different types of network layers.

DNN performance metrics obtained for the evaluation data set. Response variable was the regeneration success or failure over 30 years.

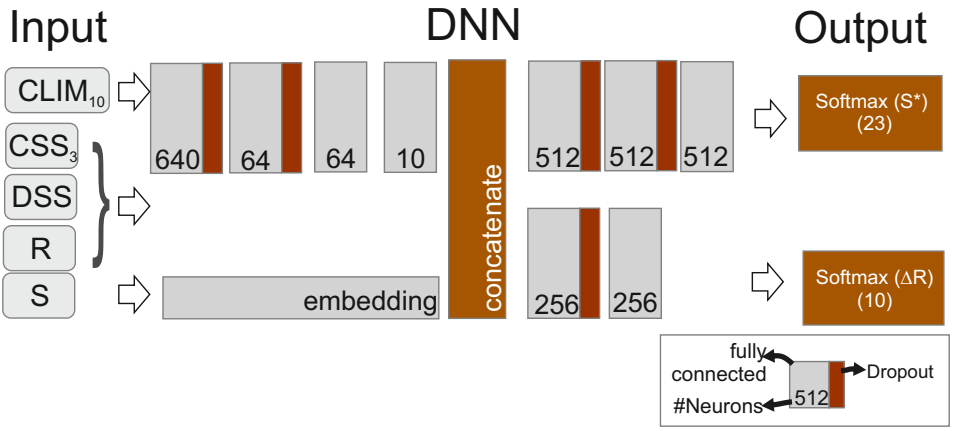


Fig. 3: Structure of the DNN.

Metric	Equation	Value
Accuracy	$\frac{tn+tp}{N}$	0.941
Precision	$\frac{tp}{tp+fp}$	0.948
Recall	$\frac{tp}{tp+fn}$	0.960
F1 Score	$\frac{2*precision*recall}{precision+recall}$	0.954
Conditional Kappa	$\frac{precision-\frac{tp+fn}{N}}{1-\frac{tp+fn}{N}}$	0.859
True skill statistic	$\frac{precision+tn}{tn+fn-1}$	0.878

Tab. 2: The DNN achieved an accuracy for predicting regeneration success or failure of 0.941 (F1 score 0.954) on an evaluation data set not used for network training (see Tab.2 for additional performance metrics).

The simulations with the SVD model for the GYE predicted a substantial proportion of area that failed to regenerate until the end of the century (Fig. 3). The proportion of the affected area was between 28 % under the moderate RCP 4.5 (wet) and 58 % under the RCP 8.5 scenario, indicating an increased probability of failure under hot and dry conditions. Forest types were also affected differently by fire. Historically well fireadapted forest types such as Lodgepole pine burned more frequently, but were also better able to regenerate compared to forest types less adapted to fire.

4 Discussion

Scaling has long been a central issue in ecology and remains a challenge for ecological modeling. The SVD framework combines the discretization of vegetation states with the predictive power of a deep neural network for estimating transition probabilities and

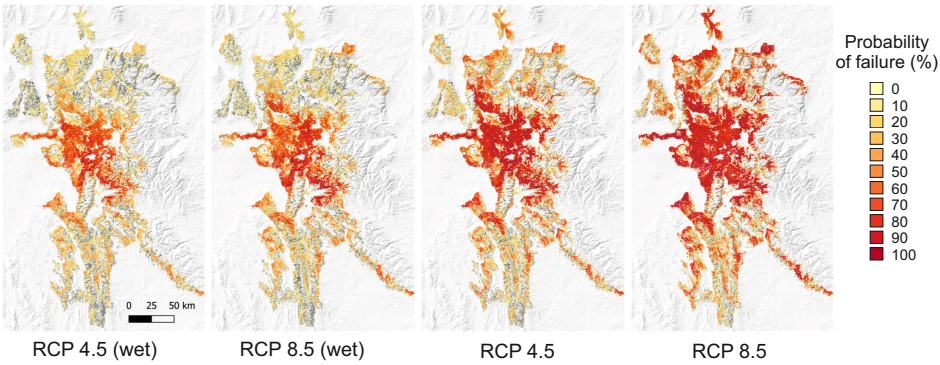


Fig. 4: Spatial distribution of the probability of regeneration failure under the four climate change scenarios for the year 2100. The probability is calculated as the average over the replicates per scenario.

pathways. We found DNNs to work well as the engine of such a meta-modeling approach. Specifically, the DNN accurately reproduced the complex responses of an underlying process-based model, yielding high prediction accuracies. The final DNN was well able to predict situations that were not included in the underlying training data, and thus showed high potential for generalization, which is an important ability in the context of upscaling [5, 9]. The application proved also the computational efficiency of SVD as simulation times were three to four orders of magnitude faster compared to the detailed process-based model. While simulations with the individual based model iLand are currently impractical for areas much larger than ~50,000 ha (with simulation times of hours), the whole GYE (2.9 Mio ha forested, 100 yrs simulation time) can be simulated with SVD in less than an hour on a standard PC (with a GPU to speed up DNN calculations).

The DNN that we used in our SVD simulations was the end point of many preparatory steps. Setting up a process-based model for simulations over a wide ecological gradient requires a considerable amount of model testing and evaluation. Here we were able to heavily build upon previous work [3, 6]. Furthermore, setting up and training a DNN require the modeler to make many design choices that can strongly affect the performance of the network. While the process of fine-tuning the structure of a network is potentially very time consuming, the availability of high-level abstractions with good default values (e.g., Keras) as well as powerful tools (e.g., TensorBoard) increasingly simplify this task.

In the current example we relied on synthetic data provided by the iLand model instead of empirical data as the “ground true” training data. A downside of this approach is that any biases present in the model are also transferred to the meta model. However, in many ecological domains the availability of empirical data is too sparse to allow an efficient training of notoriously data hungry DNNs . Moreover, process-based models are able to consistently consider also system responses under future “no analogous” conditions for which empirical data may not exist at all. A potential way to mitigate some of the problems

of synthetic data would be to use an ensemble of data generating models, thus reducing the uncertainty related to the formulation of a single underlying PBM.

The prevalence of early-seral forests will increase drastically throughout the Greater Yellowstone Ecosystem. At the end of the 21st century the share of the current forest area stocked with less than 50 trees per hectare reached values of 60% for RCP 8.5. This suggests that the GYE will transition from its signature densely forested landscapes to predominately open conditions. The increasing share of early-seral systems will not only alter the visual impression of the landscape for visitors but will also change its habitat quality. An increasing share of sparsely stocked stands will also reduce the C storage potential of ecosystem, with negative consequences for the climate system.

We here focused on the regeneration phase after a disturbance and assumed static forest types. Since we neglected the adaptation capacity of forests, e.g. due to species migration, our results should not be interpreted as forest loss but rather as areas with forest loss or altered forest types.

Bibliography

[Ab16] Abadi, Martín; Barham, Paul; Chen, Jianmin; Chen, Zhifeng; Davis, Andy; Dean, Jeffrey; Devin, Matthieu; Ghemawat, Sanjay; Irving, Geoffrey; Isard, Michael et al.: Tensorflow: A system for large-scale machine learning. In: 12th {USENIX} symposium on operating systems design and implementation ({OSDI} 16). pp. 265–283, 2016.

[An16] Angermueller, Christof; Pärnamaa, Tanel; Parts, Leopold; Stegle, Oliver: Deep learning for computational biology. Molecular systems biology, 12(7):878, 2016.

[As05] Assessment, Millennium Ecosystem: Ecosystems and human well-being: wetlands and water. World Resources Institute, 2005.

[Br18] Braziunas, Kristin H; Hansen, Winslow D; Seidl, Rupert; Rammer, Werner; Turner, Monica G: Looking beyond the mean: Drivers of variability in postfire stand development of conifers in Greater Yellowstone. Forest Ecology and Management, 430:460–471, 2018.

[Er18] Erb, Karl-Heinz; Kastner, Thomas; Plutzar, Christoph; Bais, Anna Liza S; Carvalhais, Nuno; Fetzel, Tamara; Gingrich, Simone; Haberl, Helmut; Lauk, Christian; Niedertscheider, Maria et al.: Unexpectedly large impact of forest management and grazing on global vegetation biomass. Nature, 553(7686):73–76, 2018.

[GBC16] Goodfellow, Ian; Bengio, Yoshua; Courville, Aaron: Deep learning. MIT press, 2016.

[Ha18] Hansen, Winslow D; Braziunas, Kristin H; Rammer, Werner; Seidl, Rupert; Turner, Monica G: It takes a few to tango: changing climate and fire regimes can cause regeneration failure of two subalpine conifers. Ecology, 99(4):966–977, 2018.

[HT19] Hansen, Winslow D; Turner, Monica G: Origins of abrupt change? Postfire subalpine conifer regeneration declines nonlinearly with warming and drying. Ecological Monographs, 89(1), 2019.

[Kö17] Körner, Christian: A matter of tree longevity. Science, 355(6321):130–131, 2017.

[LBH15] LeCun, Yann; Bengio, Yoshua; Hinton, Geoffrey: Deep learning. Nature, 521(7553):436–444, May 2015.

[Li10] Lindner, Marcus; Maroschek, Michael; Netherer, Sigrid; Kremer, Antoine; Barbati, Anna; Garcia-Gonzalo, Jordi; Seidl, Rupert; Delzon, Sylvain; Corona, Piermaria; Kolström, Marja; et al.: Climate change impacts, adaptive capacity, and vulnerability of European forest ecosystems. Forest Ecology and Management, 259(4):698–709, Feb 2010.

[Me11] Meinshausen, Malte; Smith, S. J.; Calvin, K.; Daniel, J. S.; Kainuma, M. L. T.; Lamarque, J-F.; Matsumoto, K.; Montzka, S. A.; Raper, S. C. B.; Riahi, K.; et al.: The RCP greenhouse gas concentrations and their extensions from 1765 to 2300. Climatic Change, 109(1–2):213–241, Aug 2011.

[Re19] Reichstein, Markus; Camps-Valls, Gustau; Stevens, Bjorn; Jung, Martin; Denzler, Joachim; Carvalhais, Nuno et al.: Deep learning and process understanding for data-driven Earth system science. Nature, 566(7743):195–204, 2019.

[RS19a] Rammer, Werner; Seidl, Rupert: Harnessing deep learning in ecology: An example predicting bark beetle outbreaks. Frontiers in plant science, 10:1327, 2019.

[RS19b] Rammer, Werner; Seidl, Rupert: A scalable model of vegetation transitions using deep neural networks. Methods in ecology and evolution, 10(6):879–890, 2019.

[Se12] Seidl, Rupert; Rammer, Werner; Scheller, Robert M.; Spies, Thomas A.: An individual-based process model to simulate landscape-scale forest ecosystem dynamics. Ecological Modelling, 231:87–100, Apr 2012.

[Se14] Seidl, Rupert; Schelhaas, Mart-Jan; Rammer, Werner; Verkerk, Pieter Johannes: Increasing forest disturbances in Europe and their impact on carbon storage. Nature Climate Change, 4(9):806–810, Aug 2014.

[SRS14] Seidl, Rupert; Rammer, Werner; Spies, Thomas A: Disturbance legacies increase the resilience of forest ecosystem structure, composition, and functioning. Ecological Applications, 24(8):2063–2077, 2014.

[TRS17] Thom, Dominik; Rammer, Werner; Seidl, Rupert: The impact of future forest dynamics on climate: interactive effects of changing vegetation and disturbance regimes. Ecological monographs, 87(4):665–684, 2017.

[We11] Westerling, Anthony L; Turner, Monica G; Smithwick, Erica AH; Romme, William H; Ryan, Michael G: Continued warming could transform Greater Yellowstone fire regimes by mid-21st century. Proceedings of the National Academy of Sciences, 108(32):13165–13170, 2011.

R. Reussner, A. Koziolek, R. Heinrich (Hrsg.): INFORMATIK 2020,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2021 1029

Einblicke in den Wasserverbrauch

Charakterisierung des Wasserverbrauchs mittels Machine Learning

Martin Wagner[1]

Abstract: Das Technologiezentrum Wasser (TZW) beschäftigt sich bereits seit Jahren mit den Themen Wasserverbrauch und Digitalisierung. Ein Aspekt der Digitalisierung ist die Analyse großer Datenmengen mit dem Ziel, neue Erkenntnisse daraus zu gewinnen und somit einen Mehrwert zu schaffen. Ein Anwendungsfall ist die Analyse des Wasserverbrauchs und der sorgsame Umgang mit Wasser als grundlegende Anforderung an den ordnungsgemäßen Betrieb der Wasserversorgung. Dies beinhaltet unter anderem die Ermittlung von Wasserverlusten. Im folgenden Kurzbeitrag wird ein Verfahren zur Eventdetektion in Wasserverbrauchsganglinien zum Zweck der Identifikation von Verbrauchsanomalien und Leckagen vorgestellt.

Keywords: Wasserverbrauch; Wasserverluste; Machine Learning; Nichtnegative Matrixfaktorisierung (NMF)

1 Einleitung

In der Trinkwasserversorgung ist die Minimierung von Wasserverlusten für einen nachhaltigen Betrieb von hoher Bedeutung. Wasserverluste sind auf Leckagen im Trinkwasserverteilungsnetz aber auch in Hochbehältern zurückzuführen. Je nach Netzgröße gibt es einen unvermeidlichen Hintergrundverlust, der selbst bei guter Instandhaltung unvermeidbar ist. Daher ist eine kontinuierliche Überwachung notwendig, um Veränderungen oder Anomalien im Wasserverbrauch zu erkennen und Leckagen frühzeitig zu identifizieren. Für die Wasserverluste ist in der Regel nicht die Größe der Leckage von Bedeutung, sondern deren Dauer und oftmals sind es kleine und langanhaltende Leckagen, die zu hohen Wasserverlusten führen.

In diesem Kurzbeitrag wird ein Verfahren vorgestellt, das eine Auswertung von zeitaufgelösten Wasserverbrauchsdaten vornimmt und sowohl das normale Verbrauchsverhalten als auch verschiedene Verbrauchsanomalien identifiziert.

Vergleichbare Ansätze zur Analyse von Verbrauchsdaten werden bereits von anderen Unternehmen benutzt [Ta19].

[1] TZW: DVGW-Technologiezentrum Wasser, Außenstelle Dresden, Wasserwerkstr. 2, 01326 Dresden, Deutschland, martin.wagner@tzw.de

doi:10.18420/inf2020_97

2 Methodik

2.1 Datengrundlage

Eingangsdaten sind Zeitreihen von Einspeise- bzw. Verbrauchsdaten einer Bilanzzone (abgeschlossenes Versorgungsgebiet) mit einer zeitlichen Auflösung von kleiner einem Tag, z. B. Minuten oder mindestens einer Stunde. Jede Zone wird mit einem durch das TZW entwickelten Verfahren analysiert. Es basiert auf Methoden des maschinellen Lernens und besteht im Wesentlichen aus zwei Analyseschritten:

- Bestimmung des typischen Verbrauchsverhaltens
- Bestimmung von Anomalien (Abweichungen vom normalen Verbrauchsverhalten)

2.2 Bestimmung des typischen Verbrauchsverhaltens

Im ersten Schritt wird durch ein Mustererkennungsverfahren das typische Verbrauchsverhalten ermittelt. Das Prinzip wird in Abb. 1 skizziert. Der Verbrauch eines Tages weist prinzipiell immer die gleiche Charakteristik auf. Nachts ist der Verbrauch gering, der Morgen und Abend ist jeweils durch eine Verbrauchsspitze gekennzeichnet, während sich tagsüber der Bedarf auf einem mittleren Niveau befindet. Dennoch sind nicht alle Tage gleich. So gibt es feine Unterschiede hinsichtlich der Höhe und der Uhrzeit der Bedarfsspitzen. Mit Hilfe einer affinen nicht-negativen Matrixfaktorisierung (affine NMF) werden normale Verbrauchstypen identifiziert. Bei der affinen NMF handelt es sich um eine Methode der Dimensionsreduktion, mit der eine Matrix $\mathbf{X}$ wie folgt in k Komponenten faktorisiert wird:

$$\mathbf{X} = \mathbf{AH} + \mathbf{a1}^T + \mathbf{E} \tag{1}$$

mit $\mathbf{X}$ als zu faktorisierende Matrix $\in \mathbb{R}^{m \times n}$, $\mathbf{A}$ als Matrix $\in \mathbb{R}^{m \times k}$, $\mathbf{H}$ als Matrix $\in \mathbb{R}^{k \times n}$, $\mathbf{a}$ als Scorevektor $\in \mathbb{R}^m$, $\mathbf{1}$ als Einsvektor $\in \mathbb{R}^n$ sowie $\mathbf{E}$ als Matrix der Residuen $\in \mathbb{R}^{m \times n}$.

Eine wichtige Nebenbedingung bei der NMF ist, dass alle Werte von $\mathbf{A}$, $\mathbf{a}$ und $\mathbf{H}$ positiv sind. Mehr zu dem Thema ist in [Ci09] zu finden. Das Verfahren wird beispielsweise in der Chemometrie für die Spektrenanalyse angewandt und bietet den Vorteil, dass die ermittelten, ausschließlich positiven, Komponenten in der Regel gut chemisch/physikalisch interpretierbar sind.

Im Beispiel in Abb. 1 wurden mittels der affinen NMF genau zwei verschiedene Typen ermittelt. Verbrauchstyp 1 entspricht einem typischen Werktag, Verbrauchstyp 2 einem typischen freien Tag (Wochenende, Ferien bzw. Feiertag). Beide Typen unterscheiden sich hinsichtlich des Zeitpunkts und der Höhe der Morgenspitze.

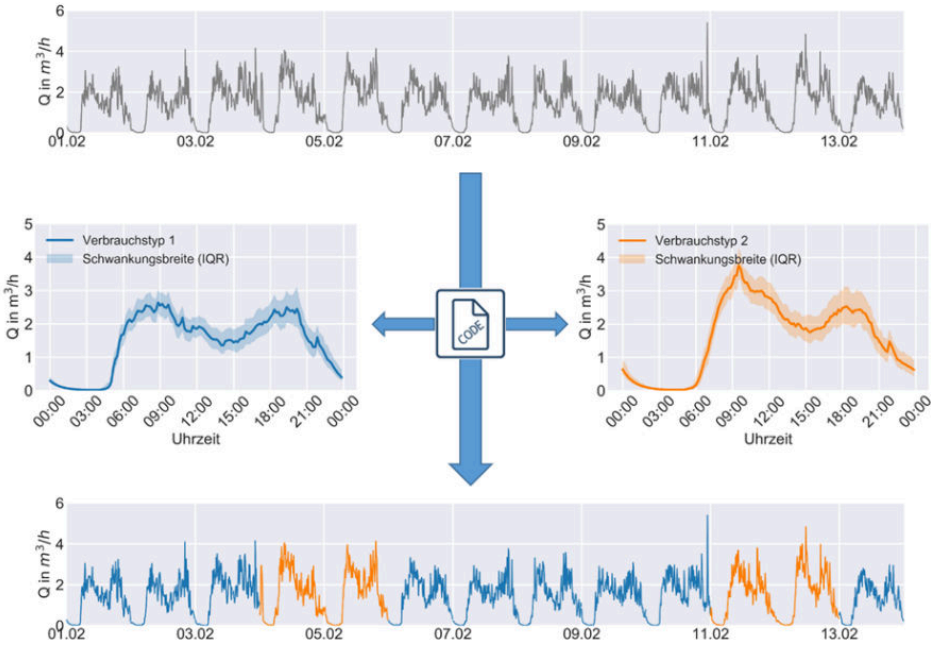


Abb. 1: Mittels NMF werden aus einer Zeitreihe typische Verbrauchsmuster identifiziert. Verbrauchstyp 1 entspricht einem typischen Werktag, Verbrauchsmuster 2 einem Wochenende.

2.3 Bestimmung von Anomalien

Ausgehend von dem Normalverhalten werden in einem zweiten Analyseschritt Abweichungen ermittelt, die als Verbrauchsanomalien ausgegeben werden und auf einfachen statistischen Methoden beruhen, wie die Ausreißer-Erkennung nach Tukey [Tu77]. Das Verfahren des TZW differenziert dabei zwischen vier verschiedenen Typen von Anomalien (Basislinien-/Tag-/Nacht-/Profilanomalien, siehe Abb. 2).

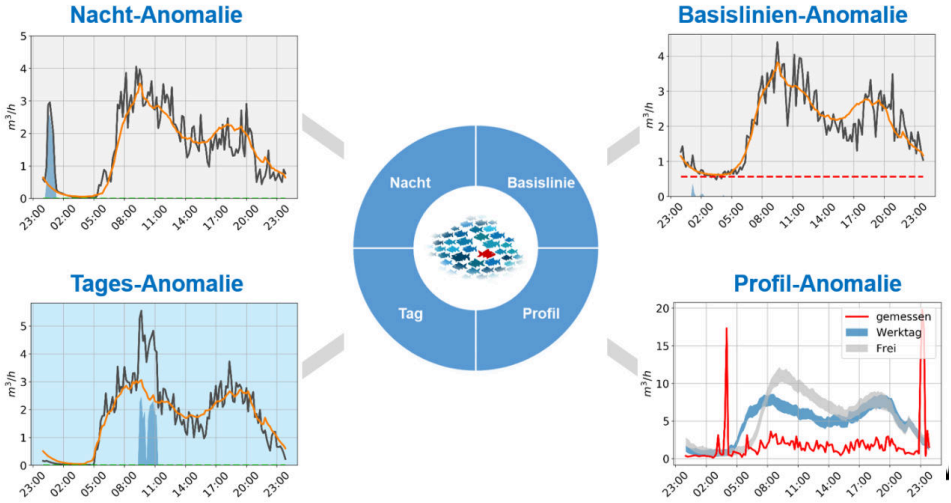


Abb. 2: Das Analyseverfahren des TZW differenziert zwischen vier verschiedenen Arten von Anomalien.

Eine Tages-Anomalie wird durch eine ungewöhnlich hohe Bedarfsspitze während des Tages charakterisiert. Eine Nacht-Anomalie wird analog durch eine ungewöhnlich hohe Bedarfsspitze während der Nachtstunden ermittelt. Sie ist nicht mit dem Nachtmindestverbrauch zu verwechseln. Eine Tag- oder Nachtanomalie kann beispielsweise durch betriebliche Tätigkeiten wie Netz- oder Behälterspülungen hervorgerufen werden oder aber auch durch Bewässerung oder Löschwasserbedarf.

Ungewöhnliche hohe Nachtmindestverbräuche werden als Basislinien-Anomalie bezeichnet. Hier weist das Verbrauchsprofil eine ganz normale Form auf, besitzt jedoch einen ungewöhnlich hohen Offset, der in Abb. 2 als rote gestrichelte Linie dargestellt ist.

Eine Profil-Anomalie ist durch ein Tagesprofil gekennzeichnet, das sich völlig von normalen Verbrauchsprofilen unterscheidet.

Neben der Ermittlung der Anomalie-Typen erfolgt gleichzeitig auch die Berechnung der täglichen Wassermengen, die durch den jeweiligen Typ hervorgerufen werden. Im Ergebnis können diese anschaulich in Form mehrerer Zeitreihen dargestellt werden. Dabei ist nicht jede Anomalie ist ein signifikantes Event. In einem weiteren Analyseschritt erfolgt daher die

automatische Segmentierung der Zeitreihen mittels einer Clusteranalyse, um echte Events identifizieren zu können. Das Ergebnis einer solchen Segmentierung ist in Abb. 3 dargestellt.

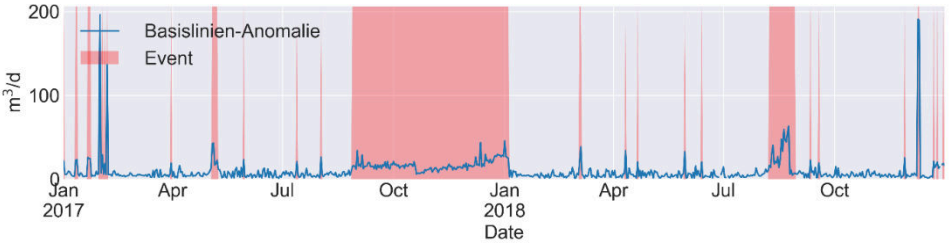


Abb. 3: Signifikante Events in der Zeitreihe der Basislinien-Anomalie.

In einem letzten Schritt erfolgt schließlich eine Plausibilisierung der ermittelten Events, die durch Gespräche mit dem Wasserversorgungsunternehmen erfolgt. Hier wird ermittelt, welche Events beispielsweise durch betriebliche Tätigkeiten erklärt werden können und welche nicht.

3 Zusammenfassung

In diesem Kurzbeitrag wurde ein vom TZW entwickeltes Verfahren für die detaillierte Analyse von Verbrauchsganglinien vorgestellt, welches sowohl eine umfassende Charakterisierung als auch die Identifizierung von Anomalien (hervorgerufen z. B. durch Wasserverluste, Sonderentnahmen, Wasserdiebstahl, etc.) ermöglicht.

Literaturverzeichnis

[Ci09] Cichocki, Andrzej; Zdunek, Rafal; Phan, Anh Huy; Amari, S.: Nonnegative Matrix and Tensor Factorizations - Applications to Exploratory Multi-way Data Analysis and Blind Source Separation. Wiley, 2009.

[Ta19] TaKaDu: , How predictive analytics curtails water loss and prevents collateral damage from hidden leaks. `https://www.takadu.com`, 2019. Accessed: 2020-07-07.

[Tu77] Tukey, John W.: Exploratory data analysis. 1977.

R. Reussner, A. Koziolek, R. Heinrich (Hrsg.): INFORMATIK 2020,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2021 1035

Feature-basiertes Clustering von Umweltzeitreihen mit Self-Organizing-Map-Ensembles

Short Paper

Andreas Wunsch [1], Tanja Liesch [2], Stefan Broda [3]

Abstract: Die Zeitreihenanalyse ist für Umweltwissenschaften ein wichtiges Werkzeug, um Systeme zu charakterisieren, da sich in den Zeitreihen Signale, welche von unterschiedlichen Einflussgrößen herrühren, wiederfinden lassen. Ein Clustering kann helfen ähnliche Dynamiken zu gruppieren um so entsprechende Einflussgrößen zu erkennen und deren Einflussbereich zu charakterisieren. Wir stellen einen unüberwachten Ensemble-Modellierungsansatz für das Clustering von Umweltzeitreihen auf der Grundlage ihrer Dynamik vor. Der Feature-basierte Ansatz erlaubt es auch heterogene Datensätze zu nutzen, das Clustering der Features erfolgt schließlich auf der Basis von Self-Organizing-Maps. Der Ensemble-Ansatz reduziert die Willkür bei der Featureauswahl und erhöht die Robustheit des Endergebnisses. Die Ergebnisse einer beispielhaften Anwendung im Grundwasserbereich zeigen, dass die vorgestellte Methodik adaptiv in der Lage ist, homogene Gruppen von Zeitreihen-Dynamiken zu identifizieren.

Keywords: Time Series Clustering; Environmental Time Series; SOM; Ensemble Modelling; Feature Clustering; Groundwater

1 Einführung

Umweltzeitreihen sind meist das Ergebnis eines komplexen Zusammenspiels einer Vielzahl unterschiedlichster Prozesse, oft zusätzlich überlagert von einem gewissen Maß an Rauschen. Daher ist die Zeitreihenanalyse für Umweltwissenschaften ein wichtiges Instrument um natürliche Prozesse und ihre Dynamik zu beschreiben und zu verstehen. Zeitreihen-Clustering kann hilfreich sein, um gemeinsame räumliche und zeitliche Dynamikmuster zu identifizieren und zwischen Signalen, die aus externen beeinflussenden Faktoren resultieren sowie Rauschen zu unterscheiden. Dies gilt insbesondere für größere Datensätze, für die manuelle Methoden schnell an ihre Grenzen stoßen. Auf diese Art und Weise können Gruppen mit gemeinsamen Dynamik-Mustern, also gemeinsamen Systemeigenschaften, identifiziert und

[1] Karlsruher Institut für Technologie (KIT), Inst. f. Angewandte Geowissenschaften, Abt. Hydrogeologie, Kaiserstr. 12, 76131 Karlsruhe, Germany, andreas.wunsch@kit.edu, https://orcid.org/0000-0002-0585-9549
[2] Karlsruher Institut für Technologie (KIT), Inst. f. Angewandte Geowissenschaften, Abt. Hydrogeologie, Kaiserstr. 12, 76131 Karlsruhe, Germany, tanja.liesch@kit.edu, https://orcid.org/0000-0001-8648-5333
[3] Bundesanstalt für Geowissenschaften und Rohstoffe (BGR), Wilhelmstr. 25-30, 13593 Berlin, Germany, stefan.broda@bgr.de, https://orcid.org/0000-0001-6858-6368

doi:10.18420/inf2020_98

so besser verstanden werden. Clustering kann darüber hinaus auch als Grundlage für weitere Analysen dienen, wie zum Beispiel der Berechnung von Prognosen und dem Aufbau von Szenarientools. Typische Anwendungen von Zeitreihen-Clustering im Umweltbereich finden sich unter anderem in der Hydrometeorologie [GL16, Wa18], Hydrologie [Mi19, TS17], Fernerkundung [GWW16] oder bei Luftschadstoffen [DDM15].

Neben den klassischen Ansätzen wie der Cluster-Analyse (CA) und der Hauptkomponentenanalyse (PCA), die beide häufig zum Clustering von Umweltzeitreihen angewendet werden, bieten Künstliche Neuronale Netze (ANN) alternative Lösungen für den Umgang mit größeren mehrdimensionalen Datensätzen, z.B. durch die Verwendung von Self-Organizing-Maps (SOM) für unüberwachtes Clustering. Üblicherweise verwenden die meisten Methoden die Zeitreihen direkt für das Clustering und sind in großem Maße von qualitativ hochwertigen Daten abhängig (Zeitreihen mit gleicher Länge oder Periode, keine Datenlücken usw.). Umweltzeitreihen zeichnen sich jedoch häufig durch eine signifikante Anzahl an Datenlücken und Ausreißern aus. Größere Datensätze sind weiterhin oft sehr heterogen bez. Zeitreihenlänge und zeitlicher Auflösung der Einzelzeitreihen. Hierdurch reduziert sich der Anteil der verwendbaren Daten in der Regel erheblich. Feature-basierte Ansätze können dieses Problem potenziell lösen, da auch lückenhafte Eingabedaten unterschiedlicher Länge verwendet werden können und sich somit die Abhängigkeit von der Datenqualität reduziert. Bei der Anwendung eines Feature-basierten Ansatzes auf Umweltdaten ist es wichtig, dass die gewählten Features die Besonderheiten der jeweiligen Anwendung bzw. des betrachteten Parameters berücksichtigen. Während einige allgemeine statistische Maße und Merkmale wahrscheinlich für die meisten (Umwelt-) Zeitreihen als Features geeignet sind, sind für aussagekräftige Clusterergebnisse meist zusätzliche, an die Dynamik des jeweiligen Zeitreihenparameters angepasste Features notwendig.

Ein Feature-basiertes Zeitreihen-Clustering unter Verwendung von SOM wurde für Umweltzeitreihen beispielsweise von [NAV15] im Bereich der Hydrologie und von [Di13] im Kontext einer geophysikalischen Fragestellung angewendet. Erstere nutzen Wavelet-basierte Features, letztere Features auf Basis von PLR (Piecewise Linear Representation).

In der vorliegenden Arbeit wird ein robuster und halbautomatisierter Ansatz für das Feature-basierte Clustering von Umweltzeitreihen skizziert.

2 Methodik

Die entwickelte Methode wird beispielhaft anhand eines Datensatzes von ca. 1200 Grundwasserstands-Zeitreihen in wöchentlicher zeitlicher Auflösung demonstriert. Aufgrund von üblicherweise heterogener Datenqualität in Datensätzen von Umweltzeitreihen (Zeitreihenlänge, Sampling Intervall, Datenlücken) wurde ein Feature-basierter Ansatz gewählt, bei dem die unterschiedlichen Aspekte der Zeitreihendynamik auch für Zeitreihen variabler Datenqualität beschrieben werden können. Neben Standard-Maßen der deskriptiven Statistik wurden dazu Features entwickelt, die sich speziell für die Beschreibung von

Dynamik-Aspekten in Grundwasserstands-Zeitreihen eignen, aber durch entsprechende Modifikation auch für andere Arten von Umweltdaten angepasst werden können. Darüber hinaus wurden auch Features aus der Literatur, wie z.B. aus der Zusammenstellung von [He19] verwendet. Tab. 1 zeigt die Auswahl der Features, welche sich nach visueller Überprüfung für den vorliegenden Datensatz als aussagekräftig erwiesen haben und erläutert sowohl deren Hintergrund als auch die Herkunft.

Tab. 1: Liste der verwendeten Features für das Clustering von Grundwasserstands-Zeitreihen

Feature Name (Abkz.)	Zweck/Beschreibung	Ref*
Jährliche Periodizität (P52)	Stärke des Jahresgangs, berechnet durch Korrelation (Pearson) der mittleren jährlichen (52 Wochen) Periodizität mit der vollständigen Zeitreihe	sd
High Pulse Duration (HPD)	Durchschnittliche Dauer der Abschnitte über dem 80. Perzentil der Nichtüberschreitung, siehe [Ri96], entnommen aus [He19]	lit
Jumps	Inhomogenitäten/Brüche, teilweise auch Variabilität, berechnet als absolute und standardisierte maximale Änderung des Mittelwertes von zwei aufeinanderfolgenden Jahren	sd
Longest Recession (LRec)	(unnatürlich) lang abfallende Grundwasserstände / längste Sequenz ohne steigende Werte	sd
Low Pulse Duration (LPD)	Durchschnittliche Dauer der Abschnitte unter dem 20. Perzentil der Nichtüberschreitung, siehe [Ri96], entnommen aus [He19]	lit
Median[0,1] (Med01)	Begrenztheit, Median nach Skalierung auf [0,1], statistisches Standardmaß, entnommen aus [He19]	ss/lit
Range Ratio (RR)	Detektion von überlagernden langperiodischen Signalen, auch ausreißerempfindlich, berechnet als Verhältnis der mittleren Jahresspannweite zur maximalen Spannweite	sd
Richards-Baker Index (RBI)	Flashiness, Häufigkeit und Schnelligkeit kurzfristiger Änderungen, für detaillierte Erklärung siehe [Ba04]	lit
SD_{diff}	Flashiness, Häufigkeit und Schnelligkeit von kurzfristigen Änderungen, berechnet als Standardabweichung aller ersten Ableitungen	sd
Seasonal Behaviour (SB)	Position des Maximums im Jahresgang, Übereinstimmung mit der erwarteten durchschnittlichen Saisonalität (Min im September, Max im März)	sd
Schiefe (Skew)	Begrenztheit, Inhomogenitäten, Ausreißer, Schiefe der Wahrscheinlichkeitsverteilung	ss
Standard Error of the Mean (SEM)	standardisierte Standardabweichung der Zeitreihe	ss
Jährliche Varianz (Y_{var})	Variabilität, Periodizität, berechnet als Median der jährlichen Varianz	sd

* lit: literature, sd: self-designed, ss: standard statistics

Das Clustering selbst erfolgt mit Self-Organizing Maps (SOM) in Verbindung mit dem DS2L-Algorithmus [CBF12], welcher speziell für prototypenbasiertes Clustering (z.B. SOM oder Neural-Gas) entwickelt wurde. Vorteil gegenüber etablierten Cluster-Algorithmen wie k-means und verschiedenen Arten von hierarchischen Methoden ist vor allem die

automatisierte Bestimmung der Clusteranzahl, aber auch die große Flexibilität bez. der möglichen Clustergrößen. So tendiert der DS2L-Algortihmus im Gegensatz zu k-means bspw. nicht zu ähnlich großen Clustern. Im Gesamtablauf der entwickelten Methodik werden zudem mehrere Ensembleansätze verfolgt. So dient ein erstes Ensemble zur Auswahl einer Feature-Kombination, mit deren Hilfe sich im mathematischen Sinne und mittels interner Clustervalidierungsindizes beurteilt, möglichst optimale Clusterergebnisse erzielen lassen. Im mathematischen Sinn werden demnach also möglichst kompakte und möglichst gut separierte Cluster gesucht, wodurch in diesem Schritt auch der Anteil an anwenderspezifischer Subjektivität im Feature-Auswahlprozess reduziert wird. Weiterhin erhöht ein nachgeschaltetes resampling-basiertes Ensemble die Robustheit des Cluster-Ergebnisses. Eine abschließende visuelle Beurteilung der Clusterqualität, also ob Cluster mit homogener Zeitreihendynamik gefunden und Zeitreihen mit charakteristischen Eigenschaften (z.B. Inhomogenitäten) zusammen gruppiert wurden, ist nötig um im Workflow korrigierend eingreifen zu können. So kann bei vorhandenem Vorwissen über das System und dessen Eigenschaften bspw. die Verwendung eines Features erzwungen werden, oder begrenzt Einfluss auf die Feinheit des Clusterings (mehr oder weniger Cluster) genommen werden. Zur Berechnung wurden Matlab 2019b sowie die SOM-Toolbox [Ve00] genutzt. Für eine detaillierte Beschreibung und Evaluierung der Methodik wird an dieser Stelle auf [WLB20, Wu20] verwiesen.

3 Ergebnisse und Ausblick

Die beispielhafte Anwendung der entwickelten Methodik auf 1196 Zeitreihen von Grundwasserständen aus Baden-Württemberg und Hessen im rechtsrheinischen Bereich des Oberrheingrabens zeigt generelle Dynamik-Unterschiede zwischen dem nördlichen und mittleren bzw. südlichen Oberrheingraben. Durch die reine räumliche Verteilung der Cluster, aber auch durch die Korrelation der Cluster mit räumlichen Daten zu mutmaßlich wichtigen Einflussfaktoren, können die Cluster zudem hydrogeologisch interpretiert werden. Mutmaßlich mit intensiverer Grundwasserbewirtschaftung und geringeren Grundwasserneubildungsraten in Verbindung stehend, finden sich im nördlichen Teil nur geringe ausgeprägte Jahresgänge und allgemein kleinere Zeitreihenvariabilität. Im mittleren und südlichen Oberrheingraben, einem Bereich mit höherer Grundwasserneubildung und weniger intensiver Grundwassernutzung treten hingehen vermehrt Ganglinien mit deutlich ausgeprägtem und eher gleichförmigem Jahresgang auf (Abb. 1a). Weiterhin lassen sich (Klein-) Gruppen differenzieren, die mutmaßlich stark durch externe Faktoren wie Zuflüssen vom Grabenrand aus Schwarzwald und Odenwald, durch den Rhein oder auch durch einzelne Staustufen bestimmt sind. Sowohl die Feature-Werteverteilungen der Cluster als auch die Ganglinien in den Clustern selbst zeigen, dass trotz Datenlücken und unterschiedlicher Zeitreihenlängen mit diesem Ansatz eine homogene Gruppierung erreicht werden konnte (Abb. 1a). Mittels bestimmter Features ist es darüber hinaus möglich, Ganglinien unterschiedlicher Dynamik, die jedoch durch andere gemeinsame Merkmale wie Ausreißer oder Sprünge charakterisiert sind, in einem Cluster zusammenzufassen (Abb. 1b).

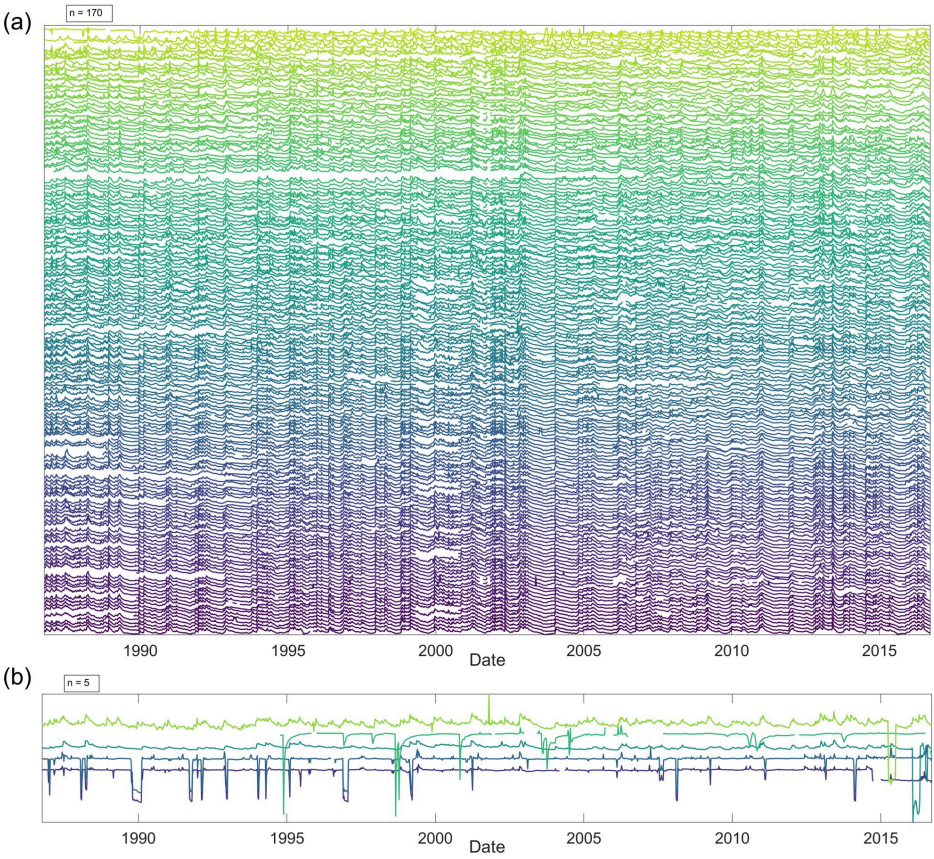


Abb. 1: Standardisierte, gestapelte Zeitreihen der Cluster 3 (a) charakterisiert durch stark korrelierte Verläufe sowie Cluster 21 (b) charakterisiert durch Ausreißer und Inhomogenitäten in der Zeitreihe.

Die vorgestellte Methode liefert damit ein solides Clustering-Framework für Umweltzeitreihen mit Vorteilen in Bezug auf (i) die Verwendung heterogener Daten, (ii) die vergleichsweise hoch automatisierte Arbeitsweise und die Möglichkeit, sich an bestimmte Datensatzmerkmale und Analyseziele anzupassen sowie (iii) robuste Ergebnisse. Neben dem verbesserten Systemverständnis und dem Erkennen lokaler und (über-) regionaler Zusammenhänge zwischen den betrachteten Zeitreihen, eignen sich die Cluster bei ausreichend guter Homogenität auch als Grundlage für weitergehende Analysen und Anwendungen. Hierzu zählen sowohl die Übertragung von z.B. Vorhersagen einer repräsentativen Zeitreihe des Clusters auf die übrigen Clusterzeitreihen, als auch das Schließen längerer Datenlücken oder eine Messwertplausibilisierung mit Hilfe hoch korrelierter Zeitreihen des gleichen Clusters. Auch die Umsetzung von Deep-Learning Ansätzen auf Zeitreihen mit für gewöhnlich zu wenigen Werten durch eine künstlich vergrößerte Datenbasis (z.B. Training auf Clusterdatensatz statt einzelne Zeitreihe) sind denkbar. Der Ansatz ist dabei auf andere Umweltzeitreihen übertragbar, wobei der Feature-Satz an die spezifische Dynamik des betrachteten Parameters angepasst werden sollten.

Literaturverzeichnis

[Ba04] Baker, David B.; Richards, R. Peter; Loftus, Timothy T.; Kramer, Jack W.: A New Flashiness Index: Characteristics and Applications to Midwestern Rivers and Streams. Journal of the American Water Resources Association, 40(2):503–522, April 2004.

[CBF12] Cabanes, Guénaël; Bennani, Younès; Fresneau, Dominique: Enriched Topological Learning for Cluster Detection and Visualization. Neural Networks, 32:186–195, August 2012.

[DDM15] D'Urso, Pierpaolo; DeGiovanni, Livia; Massari, Riccardo: Time Series Clustering by a Robust Autoregressive Metric with Application to Air Pollution. Chemometrics and Intelligent Laboratory Systems, 141:107–124, Februar 2015.

[Di13] Di Salvo, Roberto; Montalto, Placido; Nunnari, Giuseppe; Neri, Marco; Puglisi, Giuseppe: Multivariate Time Series Clustering on Geophysical Data Recorded at Mt. Etna from 1996 to 2003. Journal of Volcanology and Geothermal Research, 251:65–74, Februar 2013.

[GL16] Guo, Hongyue; Liu, Xiaodong: Dynamic Programming-Based Optimization for Segmentation and Clustering of Hydrometeorological Time Series. Stoch Environ Res Risk Assess, 30(7):1875–1887, Oktober 2016.

[GWW16] Gómez, Cristina; White, Joanne C.; Wulder, Michael A.: Optical Remotely Sensed Time Series Data for Land Cover Classification: A Review. ISPRS Journal of Photogrammetry and Remote Sensing, 116:55–72, Juni 2016.

[He19] Heudorfer, B.; Haaf, E.; Stahl, K.; Barthel, R.: Index-based Characterization and Quantification of Groundwater Dynamics. Water Resour. Res., 55(7):5575–5592, Mai 2019.

[Mi19] Mihailović, Dragutin T.; Nikolić-Đorić, Emilija; Malinović-Milićević, Slavica; Singh, Vijay P.; Mihailović, Anja; Stošić, Tatijana; Stošić, Borko; Drešković, Nusret: The Choice of an Appropriate Information Dissimilarity Measure for Hierarchical Clustering of River

Streamflow Time Series, Based on Calculated Lyapunov Exponent and Kolmogorov Measures. Entropy, 21(2):215, Februar 2019.

[NAV15] Nourani, Vahid; Alami, Mohammad Taghi; Vousoughi, Farnaz Daneshvar: Wavelet-Entropy Data Pre-Processing Approach for ANN-Based Groundwater Level Modeling. Journal of Hydrology, 524:255–269, Mai 2015.

[Ri96] Richter, Brian D.; Baumgartner, Jeffrey V.; Powell, Jennifer; Braun, David P.: A Method for Assessing Hydrologic Alteration within Ecosystems. Conservation Biology, 10(4):1163–1174, August 1996.

[TS17] Tongal, Hakan; Sivakumar, Bellie: Cross-Entropy Clustering Framework for Catchment Classification. Journal of Hydrology, 552:433–446, September 2017.

[Ve00] Vesanto, Juha: , Neural Network Tool for Data Mining: SOM Toolbox, März 2000.

[Wa18] Walz, Michael A.; Befort, Daniel J.; Kirchner-Bossi, Nicolas Otto; Ulbrich, Uwe; Leckebusch, Gregor C.: Modelling Serial Clustering and Inter-Annual Variability of European Winter Windstorms Based on Large-Scale Drivers. International Journal of Climatology, 38(7):3044–3057, 2018.

[WLB20] Wunsch, Andreas; Liesch, Tanja; Broda, Stefan: Feature-Based Groundwater Hydrograph Clustering Using Unsupervised Self-Organizing-Map-Ensembles. Water Resources Management (submitted), 2020.

[Wu20] Wunsch, Andreas: , Groundwater-Dynamic-Clustering. GitHub repository, Zenodo, 2020.

R. Reussner, A. Koziolek, R. Heinrich (Hrsg.): INFORMATIK 2020,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2021 1043

Übertragung eines Vorgehensmodells zur KI-Integration von der Industrie auf Umweltinformationssysteme

Désirée Hilbring[1] Julius Pfrommer[2]

Abstract: Maschinelles Lernen stößt in der Umwelt-Domäne auf großes Interesse. Allerdings ist der Einsatz von entsprechenden Algorithmen in Umweltinformationssystemen (UIS) bisher nicht weit verbreitet. Eine offene Frage ist zum Beispiel wie von Umweltbehörden bestehende Informationssysteme mit seit Jahrzehnten aufgebauten und gepflegten Umweltdatenbanken in einem hierarchischen föderalem System weiterentwickelt und für den Einsatz von neuen Technologien fit gemacht werden können. Hierbei sind nicht nur technische Aspekte von Interesse, sondern vor allem auch die Entwicklung geeigneter Prozesse in der Zusammenarbeit verschiedener Behörden. Für die Industrie wurde mit dem ML4P Vorgehensmodell ein toolgestütztes Verfahren für die Einführung von maschinellem Lernen in der Produktion entwickelt. Dieser Artikel untersucht dieses Vorgehensmodell im Hinblick der Übertragbarkeit des Modells auf die Einführung von maschinellem Lernen in von Behörden betriebenen Umweltinformationssystemen.

Keywords: Umweltinformationssysteme; Maschinelles Lernen; Machine Learning for Production (ML4P); Vorgehensmodell

1 Einleitung

Die öffentlichen Verwaltungen haben die letzten Jahrzehnte genutzt, im Bereich des Umweltmonitorings für die Erfüllung ihrer täglichen Aufgaben Fachinformationssysteme und Umweltdatenbanken aufzubauen. Die Funktionen behördlicher Systeme orientieren sich dabei primär an der Erfüllung von gesetzlichen Pflichten. Ein typisches Beispiel ist die Umsetzung der Wasserrahmenrichtlinie (WRRL), welche eine integrierte Gewässerschutzpolitik in Europa über Staats- und Ländergrenzen hinweg zum Ziel hat [PE00]. Dafür sind komplexe fachspezifische Algorithmen, Visualisierungen und Plausibilisierungen von Fachdaten unumgänglich. Die zugehörigen Fachalgorithmen werden von Experten häufig in Gremien abgestimmt und in den Systemen der nachgeordneten Behörden durch Software-Entwickler umgesetzt.

KI-Anwendungen im Umweltbereich sind verglichen mit anderen Disziplinen seltener. In den letzten Jahren werden jedoch verstärkt Methoden im Bereich Machine Learning

[1] Fraunhofer IOSB, Abteilung Informationsmanagement und Leittechnik, Fraunhoferstr. 1, 76131 Karlsruhe, Deutschland, desiree.hilbring@iosb.fraunhofer.de
[2] Fraunhofer IOSB, Abteilung Informationsmanagement und Leittechnik, Fraunhoferstr. 1, 76131 Karlsruhe, Deutschland, julius.pfrommer@iosb.fraunhofer.de

doi:10.18420/inf2020_99

entwickelt. Eine Studie des Umweltbundesamtes „Künstliche Intelligenz im Umweltbereich“ untersucht die Forschungsaktivität aus dem KI-Bereich im Sinne der Nachhaltigkeit. Gefundene Themengebiete sind die Verbesserung des Recycling in der Abfallwirtschaft, der Einsatz von Verfahren der Mustererkennung bei der Erdbeobachtung und Naturkatastrophen, die Erhöhung von Energie- und Gebäudeeffizienz oder die Verbesserung von Produktionsverfahren zur Erhöhung der Nachhaltigkeit [TJ19].

Für die Integration von KI-Methoden in existierende produktive Umweltinformationssysteme ist neben der Entwicklung geeigneter Verfahren ein abgestimmtes Vorgehensmodell für die Anpassung bestehender Prozesse notwendig.
Die Anforderung existierende produktive Systeme mit neuen Methoden zu erweitern besteht auch in der Industrie. Deswegen wurde mit dem ML4P Vorgehensmodell ein toolgestütztes Verfahren für die Einführung von maschinellem Lernen in der Produktion entwickelt. Dieser Artikel analysiert die Prinzipien und Komponenten des Vorgehensmodells auf die Übertragbarkeit in Umweltinformationssysteme. Für die Diskussion der Abbildung werden beispielhaft Gewässerinformationssysteme herangezogen, welche in den Umweltverwaltungen der Bundesländer Baden-Württemberg, Bayern und Thüringen für die Umsetzung der Wasserrahmenrichtlinie eingesetzt werden [Us05].

Nachfolgend wir zunächst der als Beispiel herangezogene Anwendungsfall erläutert, ein Überblick über den Stand der Wissenschaft gegeben und anschließend das ML4P Vorgehensmodell vorgestellt, welches im Abschnitt „Abbildung des ML4P Vorgehensmodells auf Umweltinformationssysteme“ hinsichtlich der Aspekte Rollenkonzept, Phasenmodell, Maschine Learning Pipeline Diagramm, virtuelle Prozessakte und technischer Tools auf die Übertragbarkeit auf UIS untersucht wird. Der Artikel schließt mit Empfehlungen für nächste Schritte.

2 Anwendungsfall Umsetzung der WRRL in Gewässerinformationssystemen

Die Abbildung des Vorgehensmodells wird in diesem Paper beispielhaft an folgendem Anwendungsfall diskutiert. Ziel der WRRL ist eine einheitliche Gewässerschutzpolitik in Europa zu erreichen, um die Gewässerqualität zu verbessern. Hierfür werden Vorgaben auf EU und Staatsebene definiert, welche in den einzelnen Staaten umgesetzt werden. Das bedeutet, dass der Ist-Zustand durch geeignetes Umweltmonitoring der Gewässer zu erfassen ist und der Soll-Zustand mit geeigneten Maßnahmen erreicht werden soll.

Für die föderale Struktur in Deutschland bedeutet dies, dass Bund und Länder ihr Handeln aufeinander abstimmen müssen. Dafür existieren länderübergreifende Fachgremien, welche die Vorgaben für die Umsetzung der Richtlinie definieren. Die aus der WRRL abgeleitete Erfassung und Verwaltung von Gewässerqualitätsdaten ist die Aufgabe der Länder und wird auf unterschiedliche Arten umgesetzt, z.B. durch die Entwicklung von geeigneten Umweltinformationssystemen. In den Bundesländern Baden-Württemberg, Bayern und

Thüringen existiert hierfür eine Entwicklungskooperation für die Realisierung von Gewässerinformationssystemen. Die resultierenden Fachanwendungen FIS GeQua, LIMNO und FIS Gewässer werden auf Basis der gemeinsamen Produktlinie WaterFrame® des Fraunhofer IOSB entwickelt [JS14]. Die Erfahrungen mit Weiterentwicklungen im Rahmen dieser real bestehenden Entwicklungskooperation dienen als Basis für die Abbildungsdiskussion des ML4P-Vorgehensmodells in diesem Artikel.
Die von der WRRL geforderten Daten und Analyseergebnisse werden zum Abschluss des Monitoring- und Analyseprozesses von allen Ländern in das bundeseinheitliche Tool WasserBLIcK gemeldet.
Als Hypothese für die Einführung von ML-Methoden wird als potentielles Beispiel die Idee der Einführung von Nitratprognosen in diesen Monitoringprozess diskutiert.

3 Stand der Wissenschaft

Das Umweltbundesamt hat mit seiner Studie „Künstliche Intelligenz im Umweltbereich“ Zukunftsperspektiven für den Einsatz von KI in der Domäne Umwelt untersucht [TJ19]. Die Studie stellt fest, dass wirtschaftliche Akteure Haupttreiber der Entwicklung von KI-Anwendungen sind. Dargestellt wird das Microsoft Programm „AI for earth“, oder eine Crunchbase-Analyse von Start-Up Unternehmen, die in der Umwelt-Domäne tätig sind und KI-Methoden einsetzen. Außerdem werden im Bereich der wissenschaftlichen Forschung die Möglichkeiten von KI erforscht. Im Abschnitt 4.2 „KI und (IT-)Infrastruktur für KI“ wird auf Voraussetzungen wie die Verfügbarkeit von Daten, die Nutzung leistungsstarker Rechenkapazitäten und das Vorhandensein sicherer Speicherarchitekturen eingegangen. Abschnitt 4.4 „KI als Sicherheits- und Warnsystem für den Umweltschutz“ stellt heraus, dass intelligente Systeme komplexe Zusammenhänge deuten und Muster erkennen können und damit als flexible Warnsysteme für den Umweltschutz dienen können.
Zieht man diese Aspekte in Betracht, ist es sinnvoll bestehende Datenbanken und Infrastrukturen bisher konventionell entwickelter Umweltinformationssysteme als Basis für die Einführung von ML-Methoden zu nutzen. Eine dafür potentiell geeignete Vorgehensweise aus der industriellen Produktion wird in den nächsten Abschnitten diskutiert. Alternative Vorgehensweisen für den Einsatz KI in behördlichen UIS sind in Veröffentlichungen bisher nicht zu finden.
Für die schriftliche Spezifikation von verteilten Software-Systemen existieren allerdings schon lange Vorgehensweisen für die Beschreibung von Software-Architekturen. Das Reference Model für Open Distrubited Processing wurde in diversen Projekten (z.B. ORCHESTRA, SANY) für die Beschreibung von SOA-Architekturen angepasst, berücksichtigt aber nicht die speziellen Anforderungen der Integration von ML-Methoden [IS19] [Us07]. Unabhängig von der Vorgehensweise bei der Beschreibung gibt die Dissertation „Ein Rahmenwerk für die Architektur von Frühwarnsystemen“ einen Überblick über verschiedene Software-Architekturen von Umweltinformationssystemen [Mo17]. In der föderalen Struktur in Deutschland ist keine einheitliche technische Umsetzung von behördlichen UIS vorgeschrieben. Dies sollte ein Vorgehensmodell im Prozess berücksichtigen.

Existierende Vorgehensmodelle aus dem klassischen Data Mining, wie zum Beispiel CRISP-DM [WH00], sind anwendungsneutral und ohne den Hintergrund der industriellen Produktion entwickelt worden. Im Fraunhofer Leitprojekt *ML4P: Machine Learning for Production* wurde, gemeinschaftlich mit sechs beteiligten Fraunhofer Instituten, das toolgestützte *Fraunhofer Vorgehensmodell speziell für maschinelles Lernen in der Produktion* entwickelt. Das Ergebnis der Nutzung des Vorgehensmodells ist eine ML-basierte Anwendung im kontinuierlichen Betrieb in der Produktion. Das Vorgehensmodell kapselt Best Practices und ermöglicht die Skalierung auf große Teams durch entsprechende Planung, die Quantifizierung des Fortschritts und klare Schnittstellen zwischen Verantwortlichkeiten. Entsprechende Werkzeuge sorgen für die direkte Anwendbarkeit der Konzepte und eine hohe Geschwindigkeit in der Ausführung. In diesem Paper wird das Vorgehensmodell in seinen Grundzügen vorgestellt.

Eine zentrale Eigenschaft des ML4P Vorgehensmodell ist die Visualisierung des Umzusetzenden Prozesses in einem Maschine Learning Pipeline Diagramm, welches von Beginn an eine anschauliche zu verfeinernde und umzusetzende Kommunikationsbasis für alle Beteiligten bietet. Da dies durch die komplexe Verteilung von Zuständigkeiten in den Behörden ein wichtiger Aspekt bei der Einführung von ML-Methoden in Umweltinformationssystemen ist, wurde das ML4P Vorgehensmodell als Basis für die Untersuchung der Übertragbarkeit ausgewählt.

4 Beschreibung des ML4P Vorgehensmodells

ML4P ist ursprünglich für den Einsatz von maschinellem Lernen in der Produktionstechnik entwickelt worden, um über datengebtriebene Prozessmodelle Verbesserungen einzubringen zu können. In der folgenden Beschreibung wird der Bezug zur Produktion auch ersichtlich sein. In der Folge findet die Übertragung auf Umweltinformationssysteme statt.

Das Vorgehensmodell ist in sechs Phasen mit klaren definierten Ergebnissen gegliedert. Für jede Phase ist eine Reihe von Phasenergebnissen definiert. Diese münden in einen Meilenstein, an dem die Phasenergebnisse begutachtet werden und eine Planung der nächsten Schritte vorgenommen wird [JP20].

Das Vorgehensmodell befindet sich in einem Spannungsfeld zwischen agilem und linearem Vorgehen. Viele ML-Projekte starten mit einer großen Unsicherheit bezüglich der Datenlage und der Eignung verschiedener Lösungsansätze. Diese Unsicherheit begünstigt ein agiles Vorgehen, bei dem im Projektverlauf laufend nachjustiert wird. Auf der anderen Seite sind Änderungen und Experimente an realen Produktionsanlagen kostenintensiv und setzen eine gute Planung voraus. Um dieses Spannungsfeld aufzulösen werden agiles und lineares Vorgehen kombiniert

1. Agiles Vorgehen innerhalb der Phasen

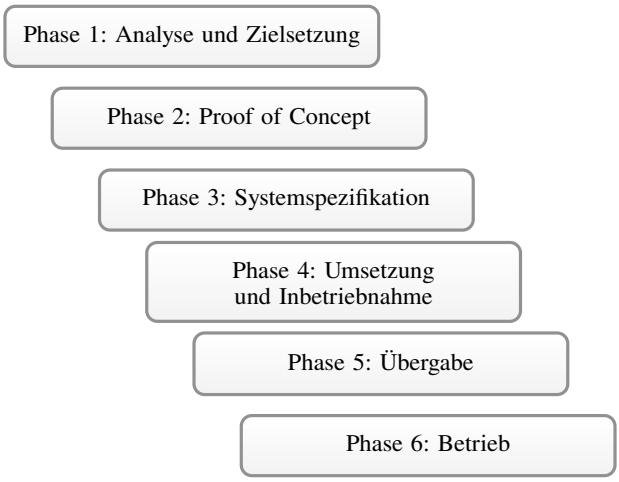


Abb. 1: Die sechs Phasen des ML4P Vorgehensmodells

2. Lineares Vorgehen über Phasen hinweg

Ein Zurückspringen auf vergangene Phasen sollte nur in Ausnahmefällen, wie etwa bei einer extern veranlassten Veränderung des Anwendungsbereichs und der Anwendungsziele, vorgenommen werden.

Das Ergebnis eines maschinellen Lernverfahrens hängt von der Verfügbarkeit qualitativ hochwertiger Datenbestände ab. In der industriellen Produktion ist eine nachträgliche Aufbereitung der Daten aber nur mit großem Aufwand machbar. Wenn zum Beispiel eine Zuordnung zwischen Datenströmen nicht möglich ist, kann eine nachträgliche Aufbereitung sogar unmöglich sein! Etwa wenn Messungen von Produktproben in externen Labors durchgeführt werden und deren Ergebnisse erst nach einigen Wochen zurück geführt werden. Nur bei einer guten Datenhaltung und mit Hilfe von geeigneten Datenmodellen ist bei zeitlich verzögerter Integration eine Zuordnung der Laboranalysen zu den Rahmenbedingungen des Probenahmezeitpunkt in der Produktion möglich. Ein entsprechendes Beispiel aus dem UIS Umfeld wäre die Zusammenführung der erfassten Umweltbedingungen zum Zeitpunkt der Probenahme mit dem zeitlich gestallelten Rückfluss der analysierten Labor-Messwerten aus verschiedenen Proben des Probenahme.

Im Vorgehensmodell wird die Datenverarbeitung ganzheitlich im Sinne einer durchgängigen Verarbeitungskette betrachtet. Dafür wird bei der Durchführung der Phasen des Vorgehensmodells ein sogenanntes Machine Learning Pipeline Diagramm aufgestellt. Beispiele finden sich in [JP20]. Ein solches ML Pipeline Diagramm ist zunächst ein visuelles Hilfsmittel zur Unterstützung menschlicher Kommunikation und der Dokumentation des gemeinsamen Prozessverständnisses. Daneben sind ML Pipeline Diagramme ein gutes Mittel zur Kommunikation mit dem Management.

In kleinen Projekten können die Schritte von der Datenerfassung bis zur Modellbildung noch von Hand ausgeführt werden. Für den automatisierten kontinuierlichen Betrieb und die Verarbeitung großer Datenmengen ist dies jedoch nicht mehr möglich. Für den kontinuierlichen und wiederholbaren Betrieb wird die ML Pipeline auf konkrete Technologien abgebildet und umgesetzt.

Die Phasen des Vorgehensmodells beziehen sich auf die (Weiter-) Entwicklungen und den Betrieb einer Machine Learning Pipeline:

In Phase 1 wird die Ist-Situation und die Ziel-Situation für die ML Pipeline aufgestellt. Diese Aktivität ist ein wichtiger Schritt um ein gemeinsames Verständnis zwischen ML-Experten und Prozessexperten sicher zu stellen. Die Darstellung erfolgt in einem Maschine Learning Pipeline Diagramm und ist dabei noch rein konzeptionell und unabhängig von konkreten Technologien. Hierbei ist es wichtig einerseits den aktuellen Ist-Zustand der in der Produktion genutzten Prozesse ohne ML-Methoden abzubilden und diesen von einem Entwurf des idealisierten Zielzustand zu unterscheiden.

In Phase 2 wird die ML Pipeline als Proof of Concept (PoC) verfeinert und validiert. Dafür findet eine partielle technologische Umsetzung statt. Diese ist aber in der Regel noch nicht für den dauerhaften Betrieb geeignet. Gängige Mittel zur Umsetzung des PoC umfassen die manuelle Ausleitung von Stichproben aus der Produktion und Datenabzüge von freistehenden (nicht vernetzten) Datenbanken der einzelnen Maschinen.

In Phase 3 wird die ML Pipeline in einer Systemspezifikation auf Technologien für den Dauerbetrieb abgebildet. Insbesondere Schnittstellen zwischen technischen Komponenten der ML Pipeline und Mensch-Maschine-Schnittstellen für die Interaktion mit dem Anlagenbediener sind im Detail zu betrachten. Die Systemspezifikation ermöglicht die Schätzung von Kosten einer Lösung für den dauerhaften Betrieb.

In Phase 4 wird die ML Pipeline als Bestandteil der Produktionsanlage für den kontinuierlichen Betrieb umgesetzt und in Betrieb genommen. Häufig wird durch Abhängigkeit zwischen Teillösungen eine gestaffelte Umsetzung gewählt. Zum Beispiel werden häufige strukturelle Anpassungen an Anlagen umgesetzt bevor die regelungstechnischen Anpassungen darauf aufgesetzt werden.

In Phase 5 wird die ML Pipeline an den Anlagenbetreiber übergeben. Dies ist ein sehr wichtiger Schritt um den Anlagenbetreiber zu befähigen das System langfristig zu betreiben. Dazu gehört auch die Schulung der Mitarbeiter in den entsprechenden Rollen.

Mit Beginn der Phase 6 ist die fertig implementierte ML Pipeline im Produktivsystem umgesetzt. Innerhalb der Phase 6 können Betrieb, sowie Wartung, Pflege und kontinuierliche Verbesserung der ML Pipeline durchgeführt werden. Dies ist notwendig insbesondere wenn der physische Produktionsprozess durch Verschleiß, Veränderung der Rahmenbedingungen oder strukturelle Umbauten im Laufe des Betriebs fortentwickelt wird.

Durchgängige Artefakte Neben den definierten Phasenergebnissen fußt das Vorgehensmodell auf zwei durchgängigen Artefakten — Dokumente und Datenstrukturen, die im Vorgehensmodell initial erstellt und in allen folgenden Phasen angewandt und fortgeführt werden. Sie repräsentieren den aktuellen Entwicklungsstand und sind zentral für das Wissensmanagement:

1. Machine Learning Pipeline Diagramm: Ein Machine Learning Pipeline Diagramm gibt eine Übersicht über die Datenerhebung und -verarbeitung, Modellbildung und Entscheidungsfindung in einem ML-unterstützten Produktionsprozess, siehe Abb. 3. Die verschiedene Phasen erfordern ein Pipeline Diagramm in unterschiedlicher Granularität.
2. Virtuelle Prozessakte: Die virtuelle Prozessakte ist eine Informationsquelle, welche für das Projektteam relevantes Wissen über die Anlage bereithält und während der gesamtem Projektzeit immer wieder adaptiert und aktualisiert wird.

Rollenmodell In den einzelnen Phasen werden unterschiedliche Kompetenzen und Disziplinen benötigt. In einem Rollenmodell sind die benötigten Disziplinen für die Phasen ausgeführt. Dabei kann eine Person auch mehrere Rollen ausfüllen. Das Fehlen einer Projektrolle im Team ist ein starker Hinweis für zu erwartende Probleme und Verzögerungen im weiteren Projektverlauf.

Technische Architektur und Werkzeugunterstützung Das Vorgehensmodell zielt auf eine planbare und reproduzierbare Durchführbarkeit von ML-Projekten in der industriellen Produktion ab. Durch die Entwicklung von Best-Practices und strukturierter Vorgänge bietet es sich an, Projekte anhand des Vorgehensmodells durch einheitliche technische Werkzeuge zu unterstützen. Dafür wird eine Architektur vorgestellt, in der die Werkzeuge zu einer ganzheitlichen Lösung im Sinne einer Machine Learning Pipeline integriert werden.

5 Abbildung des ML4P Vorgehensmodells auf Umweltinformationssystme

Dieser Abschnitt diskutiert eine mögliche Abbildung des ML4P Modells auf UIS im Hinblick auf essentielle Unterschiede. Das ML4P Vorgehensmodell beschreibt das Vorgehen für Produktionsanlagen. Umweltinformationssysteme hingegen bestehen nur zu einem kleinen Teil aus fachspezifischer Hardware (z.B. Umweltsensoren). Ihre zentralen Komponenten bestehen aus Software, deren verschiedene Funktionen für die Erfüllung von gesetzlichen Aufgaben dienen. Die Systeme sind an vielen Stellen mit Schnittstellen zu Softwaresystemen verschiedener Hersteller z.B. für Datenlieferungen oder spezielle Auswertungen ausgestattet.

5.1 Abbildung von Rollenkonzepten

Wichtig für einen funktionierenden Einführungsprozess ist die Definition wichtiger Projektrollen und die Zuordnung von entsprechenden Mitarbeitern. Genauso wie in der industriellen Produktion erfordert die Einführung von ML-Methoden in Umweltinformationssystemen die Zusammenarbeit verschiedener fachlicher Disziplinen. Die entsprechenden Mitarbeiter, arbeiten jedoch nicht zwangsweise in der gleichen Behörde. Eine Kommunikation über Behörden und Landesgrenzen hinweg ist notwendig. Dieser Abschnitt analysiert, inwieweit das ML4P Rollenkonzept auf existierende Strukturen abgebildet werden kann:

- Der Projektsponsor ist für die Finanzierung des Projektes zuständig. Im Umweltbereich sind dies häufig die Führungskräfte verschiedener Referate in den Landesbehörden.
- Prozessexperten sind die Fachexperten der Umweltinformationssysteme. Die Systeme sind komplex und realsisieren eine Vielzahl von verschiedenen Monitoring und Auswertemöglichkeiten, um gesetztliche Vorgaben aus verschiedenen Fachbereichen erfüllen zu können. Der Prozessexperte ist also üblicherweise keine einzelne Person, sondern eine Gruppe von Angestellten der übergeordneten Landesbehörden, die Anforderungen aus gesetzlichen Grundlagen und anderen Anforderungen für die Umsetzung im UIS definiert.
- Die Prozessbediener nutzen die Funktionen des UIS im täglichen Betrieb, um die gesetzliche Vorgaben zu erfüllen oder Anfragen aus der Bevölkerung zu beantworten. Prozessbediener gibt es sowohl in übergeordneten als auch in nachgeordneten Landesbehörden. Die Personalressourcen in Behörden sind knapp, deswegen haben Prozessexperten üblicherweise eine Doppelrolle als Prozessbediener.
- Der Automatisierungsingenieur des ML4P Modells kann in nächster Näherung mit den System-Administratoren der Prozessexperten verglichen werden. Diese Personen bilden innerhalb der Gruppe der Prozessexperten eine Entscheidungsebene für den Betrieb der UIS und können die Konfiguration des UIS ggf. in Zusammenarbeit mit den Software-Entwicklern entsprechend anpassen.
- Im ML4P Vorgehensmodell nicht vorgesehen ist die Rolle des Softwareentwicklers. Diese Personen sind meist nicht in den Landesbehörden, sondern in Software-Firmen angestellt und für die Wartung und Weiterentwicklung der UIS zuständig. Je nach Aufgabe stehen sie in direktem Austausch mit den System-Administratoren der Prozessexperten und den Prozessexperten.
- Bei der Rolle des Verantwortlichen für IT-Sicherheit gibt es keine speziellen Unterschiede im Hinblick auf die Übertragung des Konzeptes auf UIS.
- Die ML-Experten sind eine neue Gruppe von Personen, die für die Integration von ML-Methoden in bestehende UIS in die eingespielten Rollen der Entwicklung von Umweltinformationssystemen integriert werden müssen. Notwendig für eine

erfolgreiche Integration ist eine Abstimmung mit Prozessexperten und Software-Entwicklern.

5.2 Abbildung des Phasenmodells

Für ein geeignetes Vorgehen bei der Einführung von ML-Methoden wurden im ML4P Vorgehensmodell sechs zentrale Phasen definiert, welche im Folgenden auf ihre Übertragbarkeit auf Umweltinformationssysteme untersucht werden.

Das in Phase 1 *Zieldefinition und Lösungsansatz* zu erstellende Konzept inklusive Beschreibung der vorhandenen Ist-Situation und einer gewünschten Soll-Situation ist für die Weiterentwicklung von Umweltinformationssystemen essentiell und kann übertragen werden. Es soll hierbei ein gemeinsames Verständnis von den mit dem UIS vertrauten fachlichen Prozessexperten, den Softwareentwicklern der betroffenen UIS Komponenten und den in dieser Phase das System kennenlernenden ML-Experten entstehen. Eine Beschreibung der Zieldefinition, eine initiale Versionen der Maschine Learning Pipeline Diagramm und der virtuellen Prozessakte bilden das Ergebnis dieser Phase.

Ausgehend von diesem konzeptionellen Verständnis sollen im ML4P Modell in einem agilen Prozess geeignete Lösungsansätze gefunden werden, wobei die zu testenden Lösungsansätze entwickelt, mit Testdaten erprobt und verfeinert werden, bis ein geeignetes Verfahren gefunden wird. Bei der konventionellen Weiterentwicklung von UIS werden bestimmte fachliche Lösungsansätze häufig durch übergeordnete Gremien festgelegt und für die Umsetzung in die Systeme vorgegeben. Für die im Anwendungsfall als Beispiel herangezogenen Gewässerinformationsssysteme ist dies z.B. die Bund/Länder Arbeitsgemeinschaft Wasser - LAWA [Gr]. Ziel ist die einheitliche Umsetzung von EU- oder bundesweit vorgegeben Monitoring oder Analyseaufgaben: z.B. die der Wasserrahmenrichtlinie. Die eigentliche Umsetzung wird im föderalen deutschen System an die nachgeordneten Behörden der Länder delegiert.
Bei der bisherigen Umsetzung der Vorgaben in den verschiedenen in den Ländern eingesetzten Gewässerinformationssystemen bestehen Freiheitsgrade, welche von Prozessexperten der Behörde, die das UIS betreibt, und Softwarentwicklern in einem iterativen und agilen Entwicklungsprozess mit von Prozessexperten definierten Testdaten über die Erzeugung von Piloten und Previews ausgestaltet werden können, bis ein geeignetes Umsetzungsergebnis erzielt wird. Bei der Einführung von ML-Methoden, welche unabhängig von anderen Systemen nur für ein bestimmtes Landes-UIS integriert werden sollen, kann das aus Entwicklungssicht bestehende Vorgehen als Phase 2 *Proof of Concept* des ML4P Modells problemlos übertragen werden.
Für die erfolgreiche länderübergreifende Einführung von ML-Methoden muss die bisherige länderübergreifende fachliche Abstimmung von vorgegeben Lösungen angepasst werden, da ein iterativer Prozess zur Ermittlung eines geeigneten KI-Algorithmus notwendig ist. Hierfür werden Tests mit spezifischen UIS, wie z.B. einem Gewässerinformationssystem (FIS GeQua, LIMNO oder FIS Gewässer aus der Entwicklungskooperation) notwendig.

Daraus folgt, dass einerseits erfahrende Prozessexperten aus verschiedenen UIS bereits in Phase 1 *Zieldefinition und Lösungsansatz* als auch in die Phase 2 *Proof Of Concept* in die übergeordnete fachliche Arbeit der Gremien eingebunden werden sollten. Andererseits ist in diesen Phasen auch bereits die Beteiligung der ML-Experten notwendig. Einige spezifische UIS sollten als *Proof Of Concept* Kandidaten ausgewählt und in dieser Phase getestet werden. Die Besetzung der übergeordneten fachlichen Gremien muss also überprüft und erweitert werden. Das Ergebnis könnte eine ML Pipeline sein, die für die Umsetzung in den Behörden aller Länder in den späteren Phasen entsprechend verfeinert und angepasst wird.

Eine korrekte und konkrete Spezifikation von Schnittstellen ist für konventionelle Umweltinformationssysteme ebenfalls essentiell. Einerseits müssen die Schnittstellen zu Softwaretools verschiedener Hersteller für Softwarentwickler entsprechend dokumentiert werden, andererseits wird für die Bedienung neuer Funktionalitäten eine Anwenderdokumentation für Prozessbediener erstellt. Da die Umsetzung der Vorgaben in konkrete UIS zu den Aufgaben der Länder gehört, unterscheiden sich die länderspezifischen UIS und somit auch die Spezifikationen stark. Bei der Umsetzung des ML4P Vorgehensmodells schlagen wir vor, in dieser Phase die Entwicklung der Systemspezifikation an die zuständigen Prozessexperten und Softwareentwickler der Entwicklungskooperationen mit zugehörigen Behörden zu übergeben. Damit entsteht abhängig von der technischen Ausgestaltung eines UIS eine angepasste konkrete ML Pipeline pro technischer Umsetzung eines UIS. Für die Abbildung der UIS-Anforderungen an die jeweilige, zumeist serviceorientierte UIS-Infrastruktur einer Behörde oder eines Behördennetzwerks kann auf dedizierte Methoden wie z.B. SERVUS (Design Methodology for Information Systems based upon Service-oriented Architectures and the Modelling of Use Cases and Capabilities as Resources) zurückgegriffen werden [UB18]. Damit ist die Phase 3 *Systemspezifikation* übertragbar [JS14] [DHvdS19].

Ausgehend von der Systemspezifikation aus Phase 3 und den in Phase 2 entwickelten technischen Vorarbeiten kann in Phase 4 *Umsetzung und Inbetriebnahme* die eigentliche Implementierungsarbeit durch die Softwarenentwickeler und ML-Experten pro UIS stattfinden. Die dabei entstehenden neuen Auslieferungsversionen der Software werden von den fachlichen Prozessexperten der einzelnen Länder getestet und auftauchende Probleme an die Softwareentwickler und ML-Experten zurückgemeldet. In dieser Phase wird die neue Version in eine aus der Produktion gespiegelten Umgebung installiert.

Nach erfolgreichem Test erfolgt die Abnahme der Software durch die Prozessentwickler: Phase 5 *Übergabe*. Die Übernahme der Installation in den Produktionsbetrieb erfolgt dann üblicherweise durch die zuständigen IT-Referate oder Rechenzentren der Behörden der Prozessentwickler.

Damit ist die Übergabe vom Softwareentwickler und den ML-Experten an die Behörden der Prozesssentwickler abgeschlossen und die Wartungsphase Phase 6 *Betrieb* beginnt. In dieser Phase können durch entsprechende Nachlieferungen Anpassungen und Fehlerbehebungen an einzelnen Softwarekomponenten umgesetzt werden.

Phasen	Hauptbeteiligte der Phase	Beteiligte Verwaltungseinheit der Phase
Phase 1: Analyse und Zielsetzung	Projektsponsor / Prozess-Experten (PE) / ML Experten (MLE)	Behörde des Projektsponsors
Phase 2: Proof of Concept	PE / MLE	Fach-Gremium / Test-Entwicklungs-Behörde(n) des UIS
Phase 3: Systemspezifikation	PE / MLE / Software-Entwickler	Alle Entwicklungs-Behörden der UIS
Phase 4: Umsetzung und Inbetriebnahme	PE / MLE / Software-Entwickler	Alle Entwicklungs-Behörden der UIS
Phase 5: Übergabe	PE, IT-Sicherheit	Rechenzentren der Behörden der UIS
Phase 6: Betrieb	Prozessbediener / PE / MLE / Software-Entwickler / IT-Sicherheit	Einsatz-, Entwicklungsbehörden und Rechenzentren der UIS

Abb. 2: Abbildung des Phasenmodells auf UIS

Abbildung 2 zeigt die Übertragung der Phasen auf UIS visuell. Die Hauptverwantwortlichen einer Phase und die zuständigen Verwaltungseinheiten (z.B. die LAWA als Fachgremium mit Prozess-Entwicklern und ML Experten) sind den einzelnen Phasen zugeordnet.

5.3 Maschine Learning Pipeline Diagramm

Im folgenden wird beispielhaft ein Maschine Learning Pipeline Diagramm für ein Umweltinformationssystem definiert. Ein Ziel für die Nutzung eines ML-Algorithmus in einem UIS könnte die Vorhersage über die Entwicklung der Nitratbelastung an Messstellen sein. Das Ziel dieses Artikels ist die Diskussion einer möglichen Übertragbarkeit des ML4P Modells auf UIS, nicht die Entwicklung eines Vorhersageverfahrens, deswegen ist ein möglicher Algorithmus nur grob angerissen.

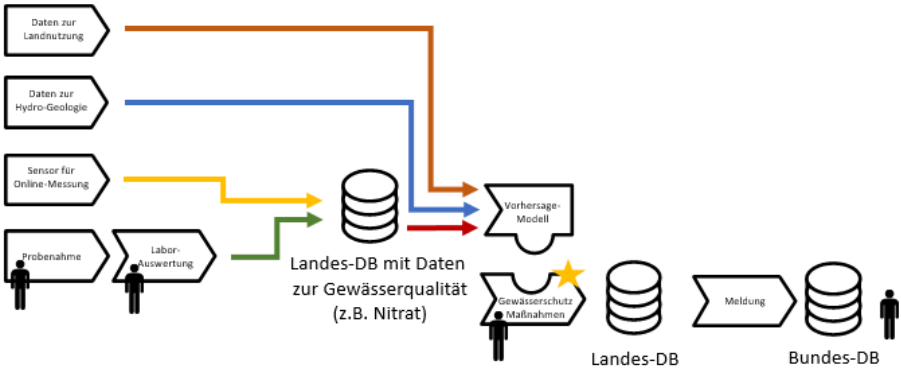


Abb. 3: Pipeline Diagramm

Abbildung 3 zeigt mögliche Eingabedaten:

- Geoinformationen aus unterschiedlichen Quellen, z.B. geologische Karten und Informationen über die Landnutzung (Ackerland, Grünland, etc.)
- Direkt gemessene Nitratwerte aus einzeln verfügbaren Sensoren.
- Produktiv gemessene Nitratwerte aus der Datenbank des Gewässerinformationssystems der letzten 20 Jahre, analysiert im Labor.

Die Eingabedaten speisen das Vorhersagemodell. Das Ergebnis des Algorithmus auf der rechten Seite der Abbildung sind aus dem Ergebnis der Vorhersage abgeleitete Gewässerschutzmaßnahmen. Die Prognoseergebnisse und Maßnahmen werden in die landeseigene Datenbank zurück gespielt und abschließend an die Datenbank der Bundesbehörde weitergeleitet.

Es ist in Planung das Vorgehensmodell an einem praktischen Projektbeispiel für die Erweiterung eines UIS mit Nitratprognosen zu testen. Im Rahmen dieses Projektbeispiels wird das Pipeline Diagramm während des Durchlaufs der Phasen des Vorgehensmodells kontinuierlich verfeinert und die für das Vorhersagemodell verwendeten konkreten KI-Methoden integriert.

5.4 Virtuelle Prozessakte

Die virtuelle Prozessakte des ML4P Vorgehensmodell beschreibt den Produktionskontext einer industriellen Anlage. Übertragen auf Umweltinformationssysteme bedeutet dies, dass in der virtuellen Prozessakte eines UIS alle relevanten Informationen bezüglich des ML-Algorithmus in der Prozessakte zusammengestellt werden. Dies sind insbesondere die Folgenden:

- Die grundlegende Bestandsdaten umfassen die Softwareversionen und Hersteller aller Komponenten des UIS und welche Daten und Funktionen die einzelnen Komponenten zur Verfügung stellen.
- Die erwartbaren Anlagendaten entsprechen dem zugrundeliegenden Datenmodell des UIS, sowie dessen Dokumentation.
- Weiterhin müssen Schnittstellen zwischen bestehenden Komponenten verschiedener Hersteller beschrieben werden. Die Schnittstellen sind auf viele unterschiedliche Arten realisiert und müssen entsprechend dokumentiert werden. Die Daten und zugehörigen Schnittstellen für den Zugriff auf die Daten können im Vorgehensmodell auf die Zusatzdaten abgebildet werden.

- Die zusätzlich erforderlichen Daten sind die, welche bisher nicht im UIS verfügbar sind, aber für die Lösung des zu implementierenden ML-Algorithmus erforderlich sind.

5.5 Eignung vorhandener ML4P Tools

Das ML4P Vorgehensmodell wird im Rahmen von Best-Practices durch die Entwicklung von technischen Komponenten unterstützt. Diese Tools können für die Realisierung einer ML Pipeline verwendet werden. Die bisher entwickelten ML4P-Komponenten basieren zumeist auf offener und frei verfügbarer Software. Diese Komponenten konzentrieren sich auf die Einbindung von Maschinen und Anlagen aus der Domäne der industriellen Produktion. Typische Standardschnittstellen für UIS unterscheiden sich allerdings zumeist von den in der Industrie eingesetzten Schnittstellen. Übliche UIS-Schnittstellen sind zum Beispiel:

- Direkter Austausch von Daten zwischen verschiedenen UIS Datenbanken einer Behörde via Datenbankaustauschtabellen.
- Im- und Export von Daten via Excel oder CSV-Dateien.
- Bedienung von proprietären Datenaustauschformaten (teilweise via XML-Schemata) von speziell für die Domäne Umwelt entwickelter Fach-Software. Typisches Beispiel ist der WasserBLIcK für die Berichterstattung gemäß der WRRL, der von den Gewässerinformationsssystemen der Länder mit Daten beliefert wird. Weiterhin bestehen Schnittstellen zu allgemeiner proprietärer Software wie Laborinformationssystemen für die Integration von Analysedaten aus erhobenen Proben.
- Datenaustausch über offene Standard-Schnittstellen des Open Geospatial Consortiums (OGC). Ein Beispiel ist die Nutzung der OGC Sensor Things API. Dieser Standard ist sowohl für die Anbindung von Messdaten aus Umweltsensoren als auch als Schnittstelle für die Integration von zeitreihenbasieren Messdaten aus Umweltdatenbanken geeignet.

Für eine erfolgreiche technische Abbildung des ML4P Vorgehensmodells sollten für verbreitete Open Source Implementierungen, welche in den UIS eingesetzt werden, ML4P-Konnektoren entwickelt werden.

6 Ausblick

Da der Einsatz von Maschine Learning Methoden im industriellen Bereich weiter fortgeschritten ist als in der Domäne des Umweltschutzes wurde in diesem Artikel analysiert, inwieweit eine Methode aus der Industrie, nämlich das ML4P Vorgehensmodell für die

Einführung von ML in der Industrie, auf die Integration von KI-Methoden in bestehende behördliche Umweltinformationssysteme übertragbar ist.
Grundsätzlich ist eine an das ML4P angelehnte Vorgehensweise sehr sinnvoll, diese muss allerdings in bestimmten Bereichen für den Einsatz von behördlichen UIS angepasst oder erweitert werden. Dabei muss vor allem in der Phase 2 *Proof Of Concept* berücksichtigt werden, dass fachliche Lösungsansätze, die gesetzlichen Vorgaben entsprechen sollen, bisher häufig in behördenübergreifenden Gremien mit Fachexperten vor der Umsetzung in den im Einsatz empfindlichen UIS festgelegt werden. Für die erfolgreiche Entwicklung von ML-Algorithmen müssen bereits in dieser Phase ML-Experten herangezogen und Lösungsstrategien in geeigneten Test-UIS erarbeitet werden. Um eine angepasste Strategie für behördliche UIS zu entwickeln sind dabei die verschiedenen Phasen des industriellen Vorgehensmodells und die definierten Rollenkonzepte sehr hilfreich.
Als nächster Schritt ist der Test des Vorgehensmodells an einem praktischen Projektbeispiel in einem UIS sinnvoll. Im Rahmen eines Projektes können aus technischer Sicht bestehende ML4P-Konnektoren für den Einsatz in UIS gestestet werden oder für UIS geeignete ML4P-Konnektoren entwickelt werden. Mit erfolgreicher Umsetzung kann so auf dem Machine Learning 4 Production Vorgehensmodell ein Machine Leanring 4 Environment Modell (ML4E) aufsetzen.

Literaturverzeichnis

[DHvdS19] Desiree Hilbring, Robert Saenger, Jörg Stumpp; van der Schaaf, Hylke: Gewässerinformationssysteme auf Basis von XCNF und die Nutzung der Sensor Things API zur Veröffentlichung von Daten. AK-UIS, 2019.

[Gr] Grambow, LAWA-Vorsitzender Herr Ministerialdirigent Prof. Dr. Martin: , Bund/Länder-Arbeitsgemeinschaft Wasser (LAWA). https://www.lawa.de/.

[IS19] ISO: , ISO/IEC 10746-3:2009 Information technology — Open distributed processing — Reference model: Architecture — Part 3. https://www.iso.org/standard/55724.html, 2019.

[JP20] Julius Pfrommer, Lars Wessels, Christian Frey Jürgen Beyerer: ML4P: Ein Standard-Vorgehensmodell für die Anwendung Maschinellen Lernens in der industriellen Produktion. In: Automatisierungskongress. 2020.

[JS14] Jörg Stumpp, Desiree Hilbring, Thomas Gülden Anette Maetze: WaterFrame® - Neue Entwicklungen in den Gewässerinformationssystemen in Baden-Württemberg, Thüringen und Bayern. Fachdokumente Baden-Württemberg, 2014.

[Mo17] Moßgraber, Jürgen: Ein Rahmenwerk für die Architektur von Frühwarnsystemen, Jgg. 29. KIT Scientific Publishing, 2017.

[PE00] PARLAMENT, EUROPÄISCHES; EUROPÄISCHER, RAT: Richtlinie 2000/60/EG vom 23. Oktober 2000 zur Schaffung eines Ordnungsrahmens für Maßnahmen der Gemeinschaft im Bereich der Wasserpolitik–Wasserrahmenrichtlinie (WRRL). Amtsblatt der Europäischen Gemeinschaften, L, 327(1):22–12, 2000.

[TJ19] Tobias Jetzke, Stephan Richter, Jan-Peter Ferdinand und Samer Schaat: , Künstliche Intelligenz im Umweltbereich. https://www.umweltbundesamt.de/sites/default/files/medien/1410/publikationen/2019-06-04_texte_56-2019_uba_ki_fin.pdf, 2019.

[UB18] Usländer, Thomas; Batz, Thomas: Agile Service Engineering in the Industrial Internet of Things. MDPI Future Internet, 10(100):doi:10.3390/fi10100100, 2018.

[Us05] Usländer, Thomas: Trends of environmental information systems in the context of the European Water Framework Directive. Environmental Modelling & Software, 20(12):1532–1542, 2005.

[Us07] Uslander, Thomas: Reference Model for the ORCHESTRA Architecture (RM-OA). Version 2 (Rev 2.1). 2007.

[WH00] Wirth, Rüdiger; Hipp, Jochen: CRISP-DM: Towards a standard process model for data mining. In: Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining. S. 29–39, 2000.

R. Reussner, A. Koziolek, R. Heinrich (Hrsg.): INFORMATIK 2020,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2021 1059

Enabling decentralized demand side management in industrial energy supply systems: A modular framework to implement control add-ons and external interfaces

Daniel Bull[1], Adrian Bürger[1], Markus Bohlayer[1], Markus Fleschutz[1], Marco Braun[1]

Abstract: Due to the increasing share of fluctuating renewable energy resources in the energy supply, the supply-demand balance needs to be increasingly supported by prosumers, who are able to adapt their energy demand and production depending on the current supply. Since small and medium-sized companies are expected to yield the potential for providing a significant share of the required flexibility, we propose an approach that enables an efficient development, testing and implementation of advanced control strategies and further data applications in decentralized energy supply systems of medium-sized companies to support the integration of such technologies and the increase of prosumer-side flexibility. The approach is based on an adaptable control framework, which is at first applied to a physical simulation model of the industrial energy system to test and train new control strategies and can afterwards be moved to the actual energy supply system of the plant.

Keywords: Adaptable Control Framework; Decentral Energy Supply; Energy System Modeling; Optimization; Load Forecasts; External Data Sources

1 Introduction

Due to the increasing share of fluctuating renewable energy resources in the energy supply, the balancing of supply and demand is getting more challenging [EC16]. As electricity is hard to store on a large scale, the supply-demand balance needs to be increasingly supported by prosumers, who are able to adapt their energy demand and local energy production depending on the current supply.

Recent studies indicate, that the prosumer-side in small and medium-sized companies could provide a significant share of this flexibility [Gi14]. The size of industrial energy system aggregates and the mostly centrally controlled energy devices enable a more profitable prosumer-side flexibility and an easier adaption compared to e.g., household energy systems. However, practical and easy to implement solutions to control a decentralized energy supply system based on external factors, like the energy spot price or control signals of the network operator, are still rare [PPB16]. The individual industrial energy system design and its technical restrictions make it difficult to develop a single infrastructure with various interfaces to external services and markets, energy supply units and additional Internet of

[1] Institute of Refrigeration, Air-Conditioning, and Environmental Engineering, Karlsruhe University of Applied Sciences, Moltkestr. 30, 76133 Karlsruhe, Germany. Email corresponding author: daniel.bull@hs-karlsruhe.de

doi:10.18420/inf2020_100

Things (IoT) devices [SKH15]. Because of that, most of the industrial energy systems are still operated by conventional controllers, which rely on current local signals and fixed set points.

Therefore, we propose an approach that enables energy practitioners and scientists to efficiently develop, test and implement new control strategies in decentralized energy supply systems, by introducing a modular and adaptable control framework. The framework is based on a central messaging system as presented by [Mü17], that enables the communication to various devices and controllers as well as the integration of additional external and internal data sources. The control framework is lightweight and can be used without excessively interfering with the existing energy system controllers of the companies. Since all data can be stored in a database, all occurring data in the framework is permanently recorded and can afterwards be used to validate control decisions or to train machine learning algorithms for e.g., thermal load or energy demand predictions of the company. Furthermore, with the introduced approach, control strategies can be tested on physical simulation models, before they are deployed in the actual energy supply system of the plant. With the simulated steady conditions, different types of advanced controls and learning algorithms can be compared and tested.

In the following, we focus on the control of the supply units of the decentralized energy system rather than on the demand side activation of actual industrial production processes, as the utilization of thermal storage capacities and the switch inbetween supply units (e.g., heat pumps and CHPs) can facilitate an improved energy consumption and production behavior. This can be done without interfering with production processes and working schedules, which might simplify the introduction of such concepts and increase their acceptance.

In Section 2, an overview of the framework and its major elements is given. A real-life application of the framework is presented in Section 3. Further possibilities for implementations of advanced optimization-based control strategies and machine learning algorithms for energy demand and availability predictions are discussed in Section 4. A conclusion is given in Section 5.

2 Modular adaptable control framework

The aim of the framework is to enable a fast implementation of new advanced control strategies in decentralized energy supply systems, which can be tested and recorded on physical simulation models before the control is applied in a real-life system. In order to provide the therefore needed flexibility and storage, the introduced modular adaptable control framework is based on separate modular elements, which include, amongst other things, a messaging system, a datastorage and multiple interfaces.

2.1 Layout and elements of the framework

The framework consists of four parts: The *Control Application*, the *Data Platform*, the *System Environment* and the *Model Environment* (see Fig. 1). While the *Data Platform* is universally usable, the *Control Application*, the *System Environment* and the *Model Environment* are tailored to meet the respective demand side flexibility target, the current company structure and the external signals which are taken into account.

The *Data Platform* is the central part of the modular adaptable framework. It contains the transmission, storage, visualization and processing of the data which appear in the framework. The data is sent over a central Messaging System, which is based on a publish subscribe pattern. All interfaces must therefore publish their data under an assigned topic name, which will then be automatically published to all subscribers of the topic. If an interface needs to obtain data, e.g., the System Interface, it can subscribe the required 'control value' topics. Every upcoming topic of the Messaging System is also subscribed by the Database Interface, which writes all published data into the Time Series Database. To allocate or create a new table, the topic name of the message is used. The message consists of a timestamp key followed by key/ value pairs, set up in the JavaScript Object Notation (JSON). The Time Series Database is utilized as a system data cache for the advanced control as well as a data source and storage for Machine Learning Algorithms, Forecast Values, the Visualization and the Event Processing. It can be accessed from outside the *Data Platform* via the the Control Interface. The Visualization obtains its data directly from the Database. It can be used to monitor data in real time, e.g. during regular operation of the industrial energy system, or to view recorded data for defined time periods, e.g., after a test run in the case of an accelerated simulation in combination with the *Model Environment*. Due to that, test runs of new advanced control strategies can be reviewed directly using the Visualization without the need of any further tools. The additional Basic Event Processing is established to receive automated feedback if the energy system is running into a prohibited state, like e.g., an overheated storage, an electric peak load or in case of a defective device.

The *Control Application* contains the to be tested and implemented advanced control strategy and its interface. The advanced control strategy can be based on conventional controllers like a Proportional-Integral-Derivative (PID) controller and set point based controllers, or controllers based on heuristic or mathematical optimization, possibly taking external control signals into account as e.g., presented in [Bo20]. Machine learning algorithms for reliable predictions of e.g., thermal load profiles as well as further applications can be implemented here. The controls and further applications are connected to the interface, which consists of a database reader and a messaging client. The database reader requests as needed the latest energy system measurements, e.g., temperature measurements or time dependent values such as hourly energy spot-prices. The messaging client publishes the control signals (e.g., On/Off signals) or other generated data (e.g. predictions of the averaged 15 min. electricity demand) to the central messaging broker. If new external or system signals need to be established (e.g., energy spot prices), the interface can be extended by adding a new

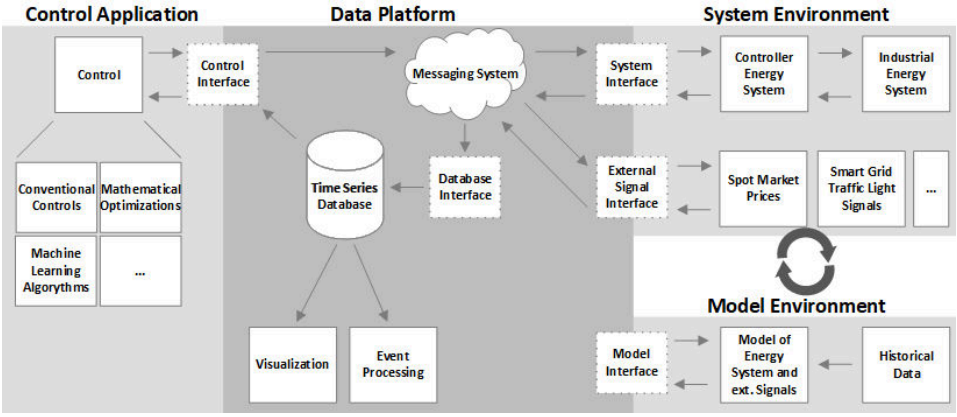


Fig. 1: Schematic description of the framework and data exchange.

database call. New control signals or other data can be sent by adding a new messaging client with a new topic.

The *System Environment* consists of two parts: The controller of the energy system together with its system interface and the external signal interface. The interfaces are used to exchange data between the *Data Platform* and the controllers of the Industrial Energy System as well as the External Signals Sources. While the interface between the Control Application and the Data Platform can be implemented as desired, the data exchange of the controller interface and the external signal interface are usually inherent by the existing controller systems and data providers. Therefore, the connection to the system controllers or external data source needs to be established by building a unique controller-messaging client interface, or a datasource-messaging client interface, respectively. Since most industrial energy systems are centrally controlled by a Programmable Logic Controller (PLC), usually a Transmission Control Protocol (TCP) based communication like Profinet or ADS can be used. To publish the current system data to the Messaging System, a constant read out cycle time is established.

The *Model Environment* contains a physical model of the industrial energy system. The aim of the model is to safely test and improve the newly developed advanced control strategies before implementing them in the real energy system of the plant as well as to train and verify e.g., machine learning algorithms. To get an adequate replica, all energy devices should be physically described by suitable dynamic system models as e.g., in [Bü20]. Therefore, detailed knowledge of the functionality of the devices as well as performance data are needed. Furthermore, historical data such as thermal and electric load profiles of the company and historical external signals like spot market prices are required to simulate the system behavior and the external signals influencing it during a week, a month or a year. Test runs can therefore be conducted within a fraction of the real cycle time, which helps to monitor long term developments and investigate rare but critical system states. The simulation

model also enables a fast generation of data sets of system data (e.g., temperatures and operation levels of the devices) and therefore builds the basis for further training data. To be able to change from the physical model to the real energy system without changing the *Control Application* or the *Data Platform*, the interface of the *Model Environment* needs to be identical to the interface of the *System Environment*. Hence, a detailed analysis of the existing system should be conducted before creating the model.

2.2 Procedure for testing new controls

Once all parts are set up, a test run of the advanced control strategy can be started. For this, a computer off site can be used, to run the architecture consisting of the *Control Application*, the *Data Platform* and the *Model Environment*. The impact of the advanced control on the modeled energy system as well as other applications like machine learning algorithms for load predictions can thereby be tested and/or trained. To test new demand side functionalities, external signals like a spot market price or a traffic light signal from the grid, which indicates the current status of the electric network and therefore the current rules of interaction for the prosumer (see also [BD15]), can be implemented gradually. The occurring changes in the control can then be evaluated by the Visualization, or by using the stored system data and further data analysis tools.

If all system parts are tested and rated stable, the *Control Application* will be moved to a computer on site. Preferably, this device is connected to the existing system controller via e.g., a stable Ethernet connection, to have as less change in the existing industrial energy system as possible. The *Data Platform* and the adapted *System Environment* should already be implemented in advance before implementing the *Control Application*, to assure a stable operation of the parts. Furthermore, data of the energy system and the external signals can already be collected for long-term processings. Once all parts are tested and working successfully, the advanced demand side control can be started as an add-on to the existing controller.

3 Real-life implementation

The introduced control framework was used to implement an energy-price and electricity demand driven control at a medium-sized electroplating company in Southern Germany. Due to the high energy and heat consuming processes, three gas driven Combined Heat and Power plants (CHP) with 14.5 kW electric power and 30.8 kW thermal power each (see Fig. 2, left) and another CHP with 50 kW electric and 79 kW thermal power (see Fig. 2, right) are installed. As thermal storage, a 3 m³ heat storage for the smaller CHPs and a 10 m³ heat storage for the bigger CHP are available. On the roof of the company building, several photovoltaic fields with 147 kW peak power are installed, which produce a large share of the required energy during noon. Moreover, two conventional gas-fueled boilers,

Fig. 2: CHPs of the electroplating company.

which were used to heat the electroplating shops, are now used as peak-load resource. The company uses the hourly day-ahead electricity price, which is determined the day before the provision at noon.

3.1 Set-up of framework components

As a first step, a conventional control with set points and PID based controls has been installed. This control strategy is presented in this section, together with the implementation of the framework. In Section 4, an overview of the planned model based control and the demand forecasts required for this task is given.

For a first insight on the industrial energy system, a model was built in the object oriented modeling language Modelica. A graphical representation of the model is shown in Fig. 3. The model is part of the *Model Environment*, which is used to test new control strategies and to gain datasets of the system behavior. The model contains the physical processes of the CHPs, boilers and the heat storages, which are described by ordinary differential equations (see also [Mo20] and [Fr15] for more information). Due to that, an adequate start-up behavior (e.g., a warm-up phase of the CHPs) and the coherences between the devices can be mapped. The energy device models replicate the control strategy of the real energy system and are partly controlled by underlying controls like e.g., circulating pumps which keep a constant outlet temperature of the CHPs. The model interface contains the same signals as the interface of the real energy system, which are the control signal of the CHPs, the temperature values of the storages and the consumer lines, the current electricity production and demand, the current gas demand and the hourly electricity day-ahead price. To get an adequate replica of a typical working day, load changes which were calculated from historic measured gas demands are implemented as thermal load in the model (see Thermal Demand in Fig. 3, which is split in two parts and then connected to the two circuits

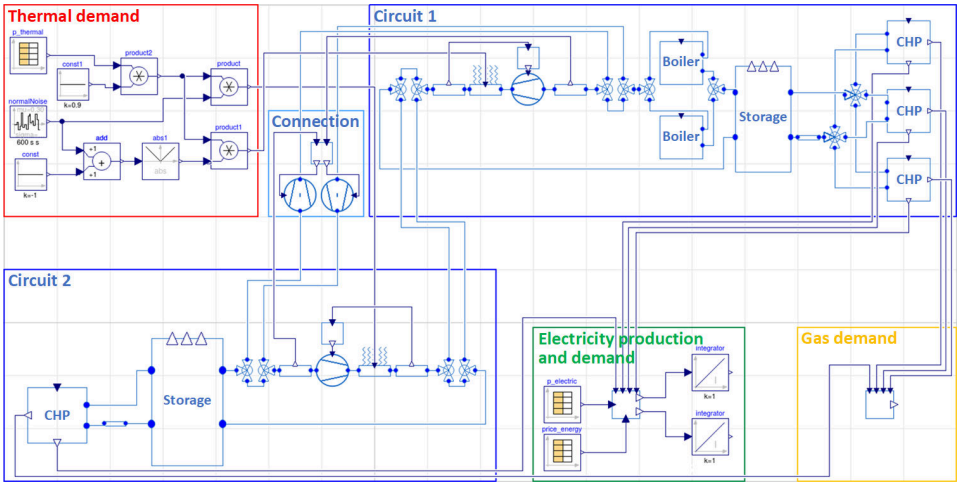


Fig. 3: Graphical representation of the physical energy system model in Modelica.

of the plant). Furthermore, the hourly electricity demand and the hourly resolved electricity day-ahead price are included as historical data input.

Using the knowledge on the behavior of the system, a first control concept was designed and implemented in Python. The control is based on the current electricity demand and the current electricity price, which is regarded as an indicator of the current supply-demand balance in the electric grid. If the energy price is high, the CHPs are working at full power, to minimize the consumed electricity drawn from the grid. If the energy price is low, e.g., on a windy and sunny weekend day where low demand and high energy supply occur, the CHPs are turned off, to utilize as much energy from the grid as possible and therefore using a high share of renewable energy. During this time, the required hot water for the electroplating shops are either provided by the thermal storages, or by the boiler. Furthermore, if in any situation a peak load of the electricity demand is detected, all CHPs are automatically started to avoid high demand fluctuation. For the same reason, in a regular operation state where the energy price is within specified boundaries, a too high feed in into the electric grid is also avoided.

The Messaging System of the *Data Platform* was set up using the lightweight and widely used MQTT-protocol [Na17]. A Database Interface is implemented in Python, which subscribes all topics of the MQTT-broker and stores all occurring data depending on their topic name into the database. Incoming measurements like temperatures, electricity demands which are published from the System Interface are stored in an energy system table, while control signals which are sent from the Control Application to the CHPs are stored separately in seperate tables. Furthermore, a table for the additional photovoltaik production and a table for the (future) hourly electricity price is established.

While in a simulated run the industrial energy system data and the hourly electricity prices are provided by the Historical Data of the *Model Environment*, in live operation they are provided by the System Interface and the External Signal Interface.

To monitor the newly implemented control strategy, a dash board was set up to visualize the current and past system operation data. The dashboard displays the key measurements of the industrial energy system, like e.g., storage temperatures, pipe temperatures, operating modes of the CHPs and the current electricity demand, which are required to ensure a safe operation. A simple Event Processing is already installed in the dashboard environment, that sends automated messages if critical temperatures, e.g., low tank temperatures, overheating CHPs or error messages, appear.

3.2 Initial operation of the control

In a first step, the new advanced control was tested together with the *Model Environment*. A regular workplace computer was used, to simultaneously run the energy model, publish the current system data to the MQTT-broker, store them in the database, run the advanced control and push the control signals back to the energy model, while being stored in the database. The time steps of the model were set to 60 seconds while the process of one run took about 0.7 seconds. A test run of one week could therefore be simulated in around 2 hours. During the tests and afterwards, all system data and control decisions were monitored in the dashboard.

After testing and determining the control values, the architecture was moved to a computer which is located at the electroplating company. It is connected with the company network and can therefore interchange with the existing energy system controller which runs on a PLC. To use its measured system data, the system interface was implemented, which reads all currently measured system data and publishes them every 10 seconds to the MQTT-broker. After a few weeks of recording the energy system data and the spot-market prices, the earlier tested advanced control was implemented on the computer on site. Together with a control signal writer which sends the current control signals, the advanced control was set into operation.

At the time of writing, the collection of operation data for the advanced controller was still ongoing. Therefore, the evaluations of energy savings and carbon dioxide reductions were still pending.

4 Predictive control strategies and machine learning applications

For future work, it is planned to implement a Model Predictive Control (MPC) strategy in the Control Application. Such strategies utilize models of a system for optimized, predictive

control decisions with regard to a defined control objective, while system states, constraints and future operation conditions of a system can be taken into account explicitly cf. [CB07].

For the presented application, an MPC strategy could utilize the thermal storages of the electroplating company in consideration of future operation conditions, such as low or high energy prices, which makes their control more efficient than using a conventional control strategy. Since these kind of methods require information about future electricity and heat demands as well as weather forecasts to gain information about the expected photovoltaic electricity production, machine learning algorithms could be employed to obtain according predictions. These could be trained using the data recorded during the real-life operation of the system or simulated data obtained via the *Model Environment* from historical gas and electricity demands.

5 Conclusion

With the introduced modular adaptive framework, new control strategies of demand side driven decentralized energy supplies can be developed, tested and implemented in a protected environment. The physical model of the energy system is highly useful for performing long term test runs of new advanced control algorithms as well as basis for training and testing machine learning algorithms. Due to the steady Data Platform, no adjustments between the real energy system and the model need to be made on the developed control strategy which enables fast prototyping. Nevertheless, the modeling of the energy system and the initial implementation of the Data Platform can be an elaborate task.

Acknowledgment

This research was supported by the German Federal Ministry for the Environment, Nature Conservation and Nuclear Safety (BMU) via WIN4climate (03KF0094).

Bibliography

[BD15] Diskussionspapier - Smart Grids Ampelkonzept . Ausgestaltung der gelben Phase. BDEW Bundesverband der Energie- und Wasserwirtschaft e.V.. Berlin, 10.03.2015.

[Bo20] Bohlayer, Markus; Fleschutz, Markus; Braun, Marco; Zöttl, Gregor: Energy-intense production-inventory planning with participation in sequential energy markets. Applied Energy, 258:113954, 2020.

[Bü20] Bürger, Adrian; Bohlayer, Markus; Hoffmann, Sarah; Altmann-Dieses, Angelika; Braun, Marco; Diehl, Moritz: A whole-year simulation study on nonlinear mixed-integer model predictive control for a thermal energy supply system with multi-use components. Applied Energy, 258:114064, 2020.

[CB07] Camacho, Eduardo F.; Bordons, Carlos: Model predictive control. Advanced textbooks in control and signal processing. Springer London Limited, London, 2nd edition, 2007.

[EC16] Impact assessment study on downstream flexibility, price flexibility, demand response & smart metering. REQUEST NUMBER: ENER/B3/2015-641. EUROPEAN COMMISSION DG ENERGY, 2016.

[Fr15] Fritzson, Peter: Principles of object-oriented modeling and simulation with Modelica 3.3 : a cyber-physical approach. IEEE Press, Piscataway, NJ, 2. ed. edition, 2015.

[Gi14] Gils, Hans Christian: Assessment of the theoretical demand response potential in Europe. Energy, 67:1 – 18, 2014.

[Mo20] Modelica Association: Modelica Language. URL: `https://www.modelica.org/modelicalanguage`, 25.06.2020.

[Mü17] Müller, Stephan; Wiener, Patrick; Bürger, Adrian; Nimis, Jens: IoT for All: Architectural Design of an Extensible and Lightweight IoT Analytics Platform. 10 2017.

[Na17] Naik, N.: Choice of effective messaging protocols for IoT systems: MQTT, CoAP, AMQP and HTTP. In: 2017 IEEE International Systems Engineering Symposium (ISSE). pp. 1–7, 2017.

[PPB16] Paolo, Zancanella; Paolo, Bertoldi; Benigna, Kiss: Demand response status in EU member states. Publications Office of the European Union. 2016.

[SKH15] Steurer, Martin; Klempp, Nikolai; Hufendiek, Kai: Identifikation und Realisierung wirtschaftlicher Potenziale für Demand Side Integration in der Industrie in Deutschland: Management Summary. 2015.

R. Reussner, A. Koziolek, R. Heinrich (Hrsg.): INFORMATIK 2020,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2021 1069

Nitrat-Monitoring 4.0 – Intelligente Systeme zur nachhaltigen Reduzierung von Nitrat im Grundwasser

Tanja Liesch[1] Julian Bruns[2] Andreas Abecker[3] Désirée Hilbring[4] Divas Karimanzira[5] Tobias Martin[6] Martin Wagner[7] Andreas Wunsch[8] Thilo Fischer[9]

Abstract: Nitrat im Grundwasser stellt weltweit unter anderem für die Trinkwasserversorgung ein großes Problem dar. Die Verteilung von Nitrat im Grundwasser ist dabei das Ergebnis eines komplexen Zusammenspiels vieler Einflussfaktoren, welches sich mit herkömmlichen Modellen für große Gebiete aufgrund der hohen Komplexität der Domäne nur schwer modellieren lässt. KI-Anwendungen, insbesondere Neuronale Netze bzw. Deep Learning Verfahren, lassen als datenbasierte Modelle, die komplexe Zusammenhänge aus einer großen Datenmenge extrahieren und übertragen können, hier einen deutlichen Mehrwert bei der zeitlich-räumlichen Vorhersage von Nitratwerten erwarten. Im vorliegenden Projekt soll daher ein übergreifendes System entwickelt werden, welches KI Verfahren mit Methoden der Umweltinformatik und speziell der Wasserdomäne kombiniert. Hierzu kommen State-of-the-Art Machine Learning Methoden wie Convolutional Neural Networks und Long short-term Memory Netzwerke zum Einsatz, um so eine verbesserte räumliche und zeitliche Vorhersage von Nitrat im Grundwasser zu erzielen und damit zur effizienten und nachhaltigen Nitrat-Reduzierung beizutragen. Diese werden mit Methoden des Operation Research und der semantischen Datenintegration erweitert, um damit einen Endnutzer bei der Entscheidungsfindung intelligent zu unterstützen.

Keywords: Convolutional Neural Networks; Long-Short Term Memory Networks; Grundwasser; Nitrat

1 Einleitung

Grundwasser bildet weltweit die größten Süßwasservorkommen und ist in vielen Regionen die Grundlage der Trinkwasserversorgung. So wird in Deutschland etwa 70% des Trinkwasserbedarfs über Grundwasser gedeckt [Bu16]. Darüber hinaus versorgen oberflächennahe

[1] Karlsruher Institut für Technologie (KIT), Institut für Angewandte Geowissenschaften, Abteilung Hydrogeologie, Kaiserstr. 12, 76131 Karlsruhe, tanja.liesch@kit.edu
[2] Disy Informationssysteme GmbH, Ludwig-Erhard-Allee 6, 76131 Karlsruhe, julian.bruns@disy.net
[3] Disy Informationssysteme GmbH, Ludwig-Erhard-Allee 6, 76131 Karlsruhe, andreas.abecker@disy.net
[4] Fraunhofer IOSB, Fraunhoferstr.1, 76131 Karlsruhe, desiree.hilbring@iosb.fraunhofer.de
[5] Fraunhofer IOSB-AST, Am Vogelherd 90, 98693 Ilmenau, divas.karimanzira@iosb-ast.fraunhofer.de
[6] DVGW-Technologiezentrum Wasser (TZW), Außenstelle Dresden, Wasserwerkstraße 2, tobias.martin@tzw.de
[7] DVGW-Technologiezentrum Wasser (TZW), Außenstelle Dresden, Wasserwerkstraße 2, martin.wagner@tzw.de
[8] Karlsruher Institut für Technologie (KIT), Institut für Angewandte Geowissenschaften, Abteilung Hydrogeologie, Kaiserstr. 12, 76131 Karlsruhe, andreas.wunsch@kit.edu
[9] DVGW-Technologiezentrum Wasser (TZW), Karlsruher Straße 84, 76139 Karlsruhe, Thilo.Fischer@tzw.de

doi:10.18420/inf2020_101

Grundwasservorkommen Pflanzen mit Wasser, bilden wertvolle Feuchtbiotope und speisen über Quellen sowie Basisabfluss Bäche und Flüsse. Die Überwachung der Grundwasserbeschaffenheit im Rahmen der europäischen Wasserrahmenrichtlinie (WRRL) hat gezeigt, dass der gute Zustand des Grundwassers vielerorts gefährdet ist. Vor allem die diffusen Einträge von Stickstoff aus der Landwirtschaft stellen hierbei ein großes Problem dar. In dieser Arbeit soll eine übergreifende Vision zur Lösung der Problematik vorgestellt werden und wie dabei KI innerhalb eines Gesamtsystems einen Beitrag für diese Fragestellung liefern kann.

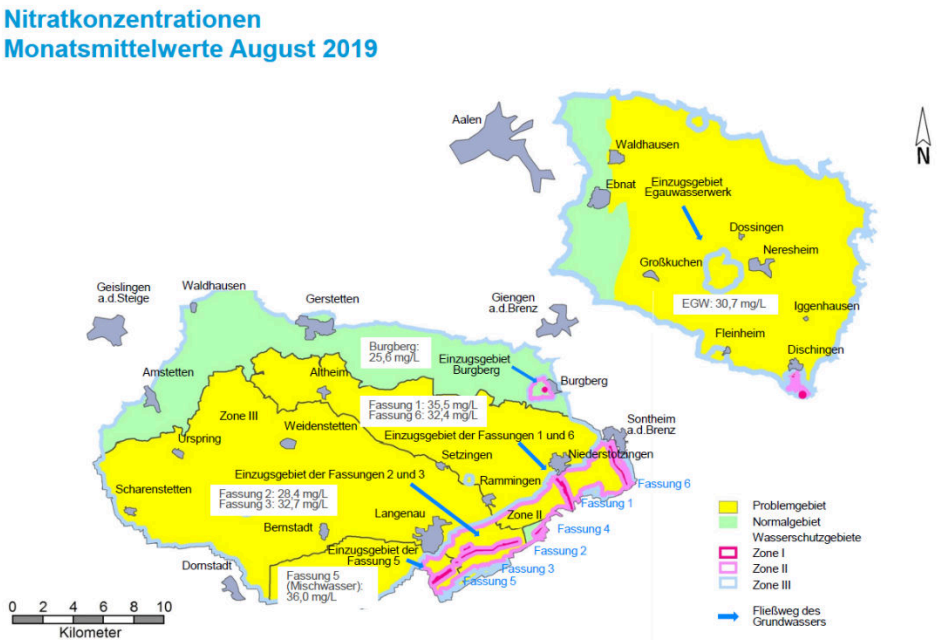


Abb. 1: Beispiel für aktuelles Nitratmonitoring in einer der Pilotregionen. Quelle: Zweckverband Landeswasserversorgung Baden-Württemberg

Die Verteilung von Nitrat im Grundwasser ist das Ergebnis eines komplexen Zusammenspiels vieler Einflussfaktoren, darunter neben dem Eintrag, der zum Großteil von der Landnutzung bestimmt wird, meteorologische Faktoren (Niederschlag, Verdunstung), chemisch-physikalische Eigenschaften der grundwasserüberdeckenden Schichten sowie Transport- und Reaktionsprozesse im Grundwasser selbst. Die Nitrat-Verteilung im Grundwasser stellt daher ein hochkomplexes, räumlich und zeitlich stark variables Muster dar, das regional und insbesondere vertikal eine ausgeprägte hydro-geochemische Differenzierung aufweist. Abb.1 zeigt ein Beispiel für die Situation und Ergebnisvisualisierung. Obwohl Eintrag, Transport und Ausbreitung von Nitrat im Grundwasser weitgehend bekannten chemisch-physikalischen Prozessen folgen, so ist eine Modellierung mit analytischen oder numerischen Modellen im gegebenen Maßstab auf Landes- oder Bundesebene in einer sinn-

vollen räumlichen Auflösung bisher nicht möglich. Räumliche und zeitliche Beurteilungen und damit verbundene Handlungsempfehlungen oder Maßnahmen, wie beispielsweise die novellierte Düngemittelverordnung, basieren daher aktuell weitgehend auf der Regionalisierung von punktuellen Messdaten bzw. der Auswertung von zeitlichen Trends. Gerade die Regionalisierung mit herkömmlichen Interpolationsverfahren bzw. die Beurteilung des Zustandes auf Basis der Grundwasserkörper ist auf Grund der hohen räumlichen Variabilität nur bedingt geeignet. Gleichzeitig ist hier die bisherige Datengrundlage und -integration außerhalb spezieller Pilotregionen häufig dünn. KI-Anwendungen, insbesondere Künstliche Neuronale Netze bzw. Deep Learning Verfahren, wie sie in anderen Disziplinen häufig zur Mustererkennung eingesetzt werden, bieten hier einen deutlichen Mehrwert gegenüber den etablierten Verfahren. Als datenbasiertes Modell sind sie in der Lage, komplexe Zusammenhänge aus einer großen Datenmenge zu extrahieren und zu übertragen. Gleichzeitig können KI-Anwendungen auch so trainiert werden, dass sich dadurch Generalisierungen für weitere Regionen ableiten lassen, was mit bisherigen Modellen nicht möglich ist.

2 Vision des Gesamtsystems

Unsere Vision ist es, ein übergreifendes, räumliches Entscheidungsunterstützungssystem (SDSS) für Nitrat-Monitoring zu entwickeln. Dieses System soll dabei innovative KI-Methoden für die Datenanalyse, die Datenintegration, die Visualisierung und die übergreifende Entscheidungsunterstützung, welche über konkrete Instanziierungen umgesetzt werden, sowie begleitende Verfahren und Prozessschritte beinhalten. Dies verfolgt das übergeordnete Ziel einer verbesserten räumlichen und zeitlichen Vorhersage von Nitrat im Grundwasser und darauf aufbauende intelligente Entscheidungsunterstützungssysteme, welche zum Beispiel durch Szenarien-Rechnungen zur Optimierung von Grundwasserschutzprogrammen einen Beitrag zur effizienten und nachhaltigen Nitrat-Reduzierung leisten können. Dabei werden jedoch nicht nur die Aspekte der KI oder reinen Umweltinformatik betrachtet sondern die Definition eines SDSS aus [JFP14] zu Grunde gelegt, wonach sich dieses aus der Kombination zweier Disziplinen zusammensetzt. Dies ist zum einen das Operation Research, gekennzeichnet nach dem bekannten drei Phasen Modell nach [Si60], und zum anderen die Geoinformatik mit Schwerpunkten in der multikriteriellen Evaluierung räumlich expliziter Modelle und der räumlichen Optimierung. Wir erweitern dies in dem geplanten System um die Komponenten der intelligenten Datenintegration unterschiedlichster Formate sowie die KI-basierte Analyse innerhalb des Systems.

Im Gegensatz zu [Jo17] verfolgt unser Ansatz die Vision, dass nur das Zusammenspiel der verschiedenen Akteure und Domänen einen Mehrwert bringt. Eine alleinige Anwendung von KI ohne Domänenwissen und Diskussion mit dem Endanwender bleibt im besten Fall ungenutzt und führt im schlimmsten Fall zu falschen Handlungsempfehlungen. Nur durch die Kooperation und Interaktion von Akteuren in einem interdisziplinärem Umfeld lassen sich tragfähige Lösungen für die Praxis entwickeln.

3 Übersicht relevanter Entwicklungen für den Einsatz von KI in der Umweltinformatik

Für die Einordnung der geplanten Arbeiten und Innovationen ist ein Blick jenseits des Themenfeldes Grundwassers hin zur gesamten Umweltinformatik hilfreich. Der KI-Einsatz in der Umweltinformatik in den letzten Jahren auf mehrere parallele Entwicklungen zurückführen, für welche im Folgenden ein kurzer, stichprobenartiger Überblick gegeben werden soll. Als entscheidender Einflussfaktor lässt sich dabei der zunehmende Reifegrad der KI-Forschung betrachten, der diese stärker in den gesellschaftlichen Fokus rückt. Im letzten Jahrzehnt hat sich insbesondere im Bereich der visuellen und sprachlichen Analyse eine starke Verbesserung der KI-Anwendungen ergeben, welche innovative Algorithmen und Ansätze hervorgebracht haben. Dies hat zu einem starken Interesse in Forschung, Wirtschaft und Politik geführt, diese Fortschritte in weitere Domänen und Anwendungsfälle zu überführen, welches im Bereich der Umweltwissenschaft und Umweltinformatik in Deutschland beispielsweise vom UBA über deren Forschungsstrategiebericht [Je19] oder international über Microsoft und deren Chief Environment Scientist Lucas Joppa in seinem NATURE Beitrag [Jo17] vorangetrieben wird. In [Jo17] wird das Anwendungsfeld der Umweltprobleme für die Informatik auf zwei Kernfragen fokussiert: „Wie kann KI bei der Lösung helfen?“ und „Wie kann die Anwendung von KI eingebracht werden?“.

Neben den Umweltwissenschaften und der Umweltinformatik lassen sich zusätzlich auch aktuelle Entwicklung in der angrenzenden Disziplin der Geoinformatik betrachten. Durch den gleichartigen Schwerpunkt auf Phänomenen mit einem starken geographischen Schwerpunkt lassen sich oft Entwicklungen aus dieser Disziplin auf die Umweltinformatik direkter anwenden als aus der klassischen Informatik. [Vo18] und [Br20] haben die Entwicklungen der KI-Forschung aus dieser Disziplin genauer betrachtet. [Vo18] beschreiben dabei den Einsatz von GeoAI für Umweltepidemiologie, bei der vor allem das Ziel verfolgt wird, eine Verbesserung der Modellierung und Beurteilung von Umweltprozessen durch den Einsatz von KI-Methoden zu erhalten. Dies kann z.B. die Exposition von Umweltbelastungen oder anderer Faktoren sein. Dadurch kann auch das Verständnis über die Zusammenhänge verbessert werden. Ein von der Informatik getriebenes Beispiel hierfür ist in [Bu17] zu finden, wo die Luftqualität feingranular erfasst und analysiert werden soll, um damit den lokalen Behörden einen besseren Überblick zu liefern. In [Br18] wurden KI-Methoden wie genetische Algorithmen für die Verbesserung der Datenqualität eingesetzt. In [Br20] wird vor allem auf die Bedürfnisse und Möglichkeiten, die sich durch und für Geodaten durch die neuesten Entwicklungen ergeben, wie das Internet der Dinge, linked-Data wie OpenStreetMap und neue, geo-temporale Datenmanagement Techniken, eingegangen. Daneben bietet der Workshop GeoAI der ACM SIG Spatial, z.B. [Ga20] eine weitere Übersicht über die aktuellen Entwicklungen der Geoinformatik.

Zuletzt haben sich weitere grundlegende Möglichkeiten in dem Betrieb von Komponenten und deren Einsatz in einem Gesamtsystem ergeben. Das bekannteste Stichwort hierbei ist Edge Computing, wodurch komplexe Berechnungen direkt an der Datenquelle durchgeführt

werden und dadurch dem Übertragungsnetz zugeführt werden können. Beispielhaft hierfür ist TinyML [PW19], welches die Ausführung komplexer KI-Modelle auf Arduinio-basierter Hardware zulässt und dann statt der Rohdaten nur die Ergebnisse weiterleitet. Gleichzeitig erlauben Ansätze aus dem Stream Processing und dortige Architekturen wie z.B. beschrieben in [Wi16] und [Wi20], dies auch in komplexen System umzusetzen.

4 KI-Anwendungen im Bereich Grundwasser

KI-Anwendungen im Bereich Grundwasser sind bisher verglichen mit anderen Disziplinen eher selten und beschränken sich weitgehend auf reine Forschungsergebnis-se. In den letzten Jahren werden jedoch auch hier verstärkt Methoden des Machine Learning erfolgreich eingesetzt. Der Ansatz einer räumlichen Nitratvorhersage mit Klassifikatoren aus dem Bereich des maschinellen Lernens wurde bereits mehrfach erfolgreich von verschiedenen Autoren an Testgebieten weltweit und auch in Deutschland ([Li06]; [KBB19], u.a.) angewendet. Klassifikatoren, die hierbei zum Einsatz kommen sind typischerweise Random Forests (RF) ([An12]; [KBB19]; [NFL15]; [Ra19]; [Ro14]; [Ro18]), Künstliche Neuronale Netze (i.d.R. einfache Feed Forward Netze) ([AK05]; [Ge09]; [HM14]; [Li06]; [NFL15]) und Boosted Regression Trees (BRT) ([KBB19]; [NFL15]; [Ra17]; [Sa18]), aber auch Support Vector Machines (SVM) ([Ra19]; [Sa18]) und Bayesische Netzwerke (BN) ([NFL15]). Auch in der zeitlichen Vorhersage von Grundwasserständen zeigen KI-Methoden gute Ergebnisse (z.B. [WLB18]). Einen guten Überblick von KI-Anwendungen im Grundwasserbereich geben die beiden Reviews von [MD00] und [REN19].

Besonders vielversprechend für den Ansatz einer räumlichen und zeitlichen Modellierung von Schadstoffen im Grundwasser erscheinen Convolutional Neural Networks (CNN). Sie stellen eine Sonderform von Multi-Layer-Perceptrons dar und gelten als State-of-the-Art-Methode für zahlreiche Anwendungen im Bereich der Klassifizierung. Diese Deep Learning-basierten Netze eignen sich vor allem zur Verarbeitung von Daten mit Raumbezug und werden daher äußerst erfolgreich zur Bilderkennung und -verarbeitung angewandt. CNN wurden im Umweltbereich bereits erfolgreich z.B. zur Klassifizierung von Hyperspektral-Daten in der Fernerkundung ([Ma15]; [YJX17]), zur Vorhersage von Niederschlag ([Xi15]), zur Identifikation von Quellen der Grundwasserbelastung ([Mo19]) und zur Vorhersage von Grundwasserständen ([As17]) eingesetzt.

5 Geplante KI-Methodik

Die räumliche Vorhersage einer Nitratbelastung mit Hilfe von Methoden des Maschinellen Lernens ist typischerweise über den Aufbau eines Klassifikators oder einer Regression möglich. Hierfür werden räumlich flächendeckende historische und aktuelle Datensätze herangezogen (z.B. CORINE-Landnutzungsdaten, Niederschlagsrasterdaten, Bodenart, Grundwasserneubildung, Flurabstand etc.), die Rückschlüsse auf die Nitratexposition und

Vulnerabilität des Untersuchungsgebietes zulassen. Anhand von punktuell verfügbaren Messwerten von Nitrat wird der Klassifikator trainiert und ist anschließend in der Lage, auch über Gebiete, für die keine punktuellen Schadstoffinformationen zur Verfügung stehen, Aussagen zu treffen. Da es sich bei Nitrat um einen äußerst redox-sensitiven Parameter handelt, welcher abhängig von den hydrochemischen Bedingungen in unterschiedlichen Spezies vorliegt (De-/Nitrifikation), werden in die Analyse entsprechende zusätzlich verfügbare Parameter wie Redox-Potential, elektrische Leitfähigkeit und pH-Wert mit einbezogen.

CNN sind State-of-the-Art Deep Learning (DL) basierte Netze zur Verarbeitung von Daten mit Raumbezug. Im Projekt werden daher 3D-CNNs für die räumliche (2D) und zeitliche (dritte Dimension) Vorhersage des Nitratgehaltes in Grundwasser eingesetzt. Üblicherweise sind DL-Ansätze besser für sehr große und möglichst vollständige Datensätze geeignet, Nitratmessung liegen jedoch vor allem im Vergleich zu Punkten ohne Nitratmesswerte verhältnismäßig spärlich vor. Hierfür gibt es neuere Ansätze wie Sparse 3D CNNs, deren Einsatz geprüft wird. Zur verbesserten Vorhersage der zeitlichen Nitratbelastung soll zudem eine Nachschaltung eines LSTM-Netzwerkes (Long-Short Term Memory) an das CNN sowie der Einsatz von Temporal Convolutional Netzwerken (TCN) getestet werden. Neben dem neuartigen CNN-basierten Ansatz wird auch der Aufbau und die Anwendung des zuvor beschriebenen, klassischerweise gewählten Ansatzes eines KI-Klassifikators umgesetzt und die Leistungsfähigkeit beider Ansätze verglichen und evaluiert, um optimale Vorhersageergebnisse zu erzielen.

6 Nutzung offener Standards für die Integration von Sensor- oder Messdaten

Für die Integration von KI-Methoden in bestehende Systeme ist die Anbindung verschiedener heterogener Datenquellen essentiell. Es wird ein intelligentes Datenmanagement benötigt, welches existierende und neue Datenquellen harmonisiert und in ein übergreifendes System zur Auswertung von Daten integrieren kann. Dabei sollen offene neue Standards eingesetzt werden. Die dadurch entstehende Kombination des harmonisierten Datenzugriffs über offene Schnittstellen und die Nutzung dieser Methoden für die KI-Integration eröffnet bisher nicht realisierte Möglichkeiten der systemübergreifenden Datenauswertung. Im Bereich des Nitrat-Monitoring erfolgt eine länderübergreifende Detail-Auswertung oder Einbeziehung aller Daten z.B. von Wasserversorgern und auf Basis von offenen Standards bisher nicht.

Für die Beschreibung von Sensoren mit Raumbezug spielt das OGC als eines der wichtigsten internationalen Standardisierungsgremien eine wesentliche Rolle, deswegen werden im Projekt die offenen Standards OGC analysiert und für das intelligente Datenmanagement getestet.

Mit dem Internet of Things haben Geodienste für die Modellierung von Sensordaten an Bedeutung gewonnen. Deswegen wurde von der OGC im Gegensatz zu bisherigen Sensordaten-Standards mit der SensorThings API ein leichtgewichtiger neuer Ansatz für

die Anbindung und Datenhaltung entwickelt, der auf langjährig bewährte Konzepte mit zeit- und georeferenzierten Sensordaten zurückgreift [LHK15]. Die SensorThings API ist grundsätzlich sowohl für den direkten Zugriff auf Sensordaten als auch für die Anbindung von bestehenden Datenbanken geeignet. Die SensorThings API steht nun zusammen mit dem Feature Service (OGC API Features) im Fokus des JRC für die Eignung der Veröffentlichung von räumlichen Daten. Im Projekt wird die Einbindung von Schadstoffdaten langjährig bestehender Quellsysteme über die Open Source Implementierung FROST ®getestet werden [IO].

Ein weiterer für das Projekt wichtiger Standard des OGC ist GroundWaterML, dessen Grundwassermodell für die Zusammenführung von Grundwasserdaten verschiedener Systeme Potential hat, bisher aber in Deutschland nicht verwendet wird. Im Projekt soll untersucht werden, inwieweit eine Harmonisierung des Datenmodells von Grundwasserdaten über diesen Standard möglich und sinnvoll ist.

Essentiell für das intelligente Datenmanagement ist außerdem eine geeignete Integration von KI-Modulen in das System. Hierfür sollen verschiedene Möglichkeiten analysiert und getestet werden. Beispiele hierfür sind die Erweiterung Task Core der Sensor Things API [LK17] oder der Einsatz von ML4P-Konnektoren, die im industriellen Bereich für die Einführung von KI-Methoden verwendet werden können.

7 Entscheidungsunterstützung

Die Optimierung von Grundwasserschutzprogrammen und der dazugehörigen Messkampagnen basiert derzeit auf geeigneten konzeptionellen und zum Teil numerischen Modellen, welche in der Lage sind, Ursache-Wirkungs-Zusammenhänge zwischen der Umsetzung konkreter Maßnahmen wie der Verringerung der Nitrat-Einträge und der Verminderung der gemessenen Nitrat-Konzentrationen im Grundwasser zu quantifizieren. Solche Modelle müssen die zu Grunde liegenden hydrogeologischen Prozesse (u.a. Stickstoffumsatz in der Bodenzone, Nitratauswaschung mit dem Sickerwasser, Stofftransport und Stoffumsatz im Grundwasserleiter) in Abhängigkeit der Komplexität der naturräumlichen Gegebenheiten abbilden [We15]. KI-basierte, datengetriebene Ansätze erscheinen hier weitaus besser geeignet, werden jedoch bisher nicht eingesetzt. Diese bieten jedoch Vorteile für die Entscheidungsunterstützung, was für die finale Nutzung und Verwertung essentiell ist.

Ergebnisse der im Projekt erarbeiteten KI-Methoden müssen von Entscheidern analysiert werden können. Aus den Ergebnissen sollen darüber hinaus geeignete Maßnahmen abgeleitet werden können. Beispiele sind Vorgaben oder Empfehlungen hinsichtlich dynamischen Düngemanagements, Schutzzonenausweisung oder Wasseraufbereitung. Konkret können Maßnahmengebiete (z.B. besonders gefährdete Gebiete mit geringem Nitratabbauvermögen im Grundwasserleiter) ausgewiesen und priorisiert werden

Für die Erreichung dieser Ziele benötigen die Entscheider eine geeignete Entscheidungsunterstützung. Hier ist zwischen der reinen Visualisierung und der Methodik der Optimierung

zu unterscheiden. Die Visualisierung wird dabei meist über Webplattformen und Dashboards gelöst. [Su13] hat den Einsatz von Dashboards im Rahmen einer Smart City Anwendung für die Integration verschiedener Sensordaten untersucht. In der Optimierung stehen hierbei weitere Verfahren zur Verfügung, wie z.B. die Standortplanung der Messstellen oder die Wissensfusion, die aus den erzielten Erkenntnissen Handlungsempfehlungen generieren. Für die Ableitung von geeigneten Handlungsanweisungen ist sowohl die Modellierung der Situation über Ontologien hilfreich, als auch die Schaffung von Simulationsmöglichkeiten für die Analyse verschiedener Szenarien. Ein Beispiel für die Aufgabe, verschiedene Sensordatenquellen zu fusionieren und für die Entscheidungsunterstützung entsprechend zu kombinieren, wird in „The sensor decision chain in crisis management“ von Moßgraber beschrieben [Mo18]

8 Ausblick

Im Rahmen des Projekts Nitrat-Monitoring 4.0 – Intelligente Systeme zur nachhaltigen Reduzierung von Nitrat im Grundwasser (NiMo 4.0) wird der Transfer von innovativen Lösungsansätzen von KI-Anwendungen im Grundwasser-Bereich von der universitären und institutionellen Forschung in die industrielle, anwendungsnahe Forschung und Praxis vorangetrieben. Übergeordnetes Ziel ist eine verbesserte räumliche und zeitliche Vorhersage von Nitrat im Grundwasser und darauf aufbauende intelligente Entscheidungsunterstützungssysteme, welche z.B. durch Szenarienrechnungen zur Optimierung von Grundwasserschutzprogrammen und damit zur effizienten und nachhaltigen Nitrat-Reduzierung beitragen. Die betrachteten Lösungsansätze und Methoden werden anhand realer Daten aus zwei wasserwirtschaftlich bedeutenden Pilotregionen entwickelt, demonstriert und validiert. Diese weisen hinlänglich große hydrogeologische Variabilität auf, um hieraus Aussagen zur Allgemeingültigkeit und Übertragbarkeit der entwickelten Lösungen treffen zu können. Um die resultierenden, wachsenden Echtzeit-Datenströme effizient überwachen zu können, werden intelligente Monitoring-Algorithmen entwickelt, die beispielsweise automatisierte Datenplausibilisierungen, Auffälligkeitserkennung oder Frühwarnmechanismen realisieren. Schließlich ermöglicht die räumliche Vorhersage in Verbindung mit modernen Methoden der Geostatistik und des Operations-Research auch Empfehlungen zur Messnetzoptimierung. Final werden dabei diese Informationen auch an den Endanwender so vermittelt, dass dieser damit agieren kann und so ein Beitrag für die Praxis geschaffen wird.

Acknowledgement

Die Vision, die in diesem Paper entwickelt wurden, sollen im Projekt Nitrat-Monitoring 4.0 - Intelligente Systeme zur nachhaltigen Reduzierung von Nitrat im Grundwasser (NiMo 4.0), gefördert vom BMU unter Förderkennzeichen FKZ 67KI2048, umgesetzt werden. Informationen zum Projekt werden sich unter https://nimo-projekt.de/ finden.

Literaturverzeichnis

[AK05] Almasri, Mohammad N; Kaluarachchi, Jagath J: Modular neural networks to predict the nitrate distribution in ground water using the on-ground nitrogen loading and recharge data. Environmental Modelling & Software, 20(7):851–871, 2005.

[An12] Anning, David W; Paul, Angela P; McKinney, Tim S; Huntington, Jena M; Bexfield, Laura M; Thiros, Susan A: Predicted nitrate and arsenic concentrations in basin-fill aquifers of the southwestern United States. US Department of the Interior, US Geological Survey, 2012.

[As17] Assem, Haytham; Ghariba, Salem; Makrai, Gabor; Johnston, Paul; Gill, Laurence; Pilla, Francesco: Urban water flow and water level prediction based on deep learning. In: Joint European Conference on Machine Learning and Knowledge Discovery in Databases. Springer, S. 317–329, 2017.

[Br18] Bruns, Julian; Riesterer, Johannes; Wang, Bowen; Riedel, Till; Beigl, Michael: Automated Quality Assessment of (Citizen) Weather Stations. GI_Forum 2018,, 6:65–81, 2018.

[Br20] Breunig, Martin; Bradley, Patrick Erik; Jahn, Markus; Kuper, Paul; Mazroob, Nima; Rösch, Norbert; Al-Doori, Mulhim; Stefanakis, Emmanuel; Jadidi, Mojgan: Geospatial Data Management Research: Progress and Future Directions. ISPRS International Journal of Geo-Information, 9(2):95, 2020.

[Bu16] Bundesamt, Statistisches: , Tabellen zu Wassergewinnung: Bundesländer, Jahre, Wasserarten. https://www-genesis.destatis.de/genesis/online/link/tabelleErgebnis/32211-0002, 2016. Accessed: 2020-07-13.

[Bu17] Budde, Matthias; Riedel, Till; Beigl, Michael; Schäfer, Klaus; Emeis, Stefan; Cyrys, Josef; Schnelle-Kreis, Jürgen; Philipp, Andreas; Ziegler, Volker; Grimm, Hans et al.: SmartAQnet: remote and in-situ sensing of urban air quality. In: Remote Sensing of Clouds and the Atmosphere XXII. Jgg. 10424. International Society for Optics and Photonics, S. 104240C, 2017.

[Ga20] Gao, Song; Newsam, Shawn; Zhao, Liang; Lunga, Dalton; Hu, Yingjie; Martins, Bruno; Zhou, Xun; Chen, Feng: GeoAI 2019 workshop report: The 3nd ACM SIGSPATIAL International Workshop on GeoAI: AI for Geographic Knowledge Discovery: Seattle, WA, USA-November 5, 2019. SIGSPATIAL Special, 11(3):23–24, 2020.

[Ge09] Gemitzi, A; Petalas, C; Pisinaras, V; Tsihrintzis, VA: Spatial prediction of nitrate pollution in groundwaters using neural networks and GIS: An application to South Rhodope aquifer (Thrace, Greece). Hydrological Processes: An International Journal, 23(3):372–383, 2009.

[HM14] Hosseini, Seiyed Mossa; Mahjouri, Najmeh: Developing a fuzzy neural network-based support vector regression (FNN-SVR) for regionalizing nitrate concentration in groundwater. Environmental monitoring and assessment, 186(6):3685–3699, 2014.

[IO] IOSB, Fraunhofer: , FROST-Server: Der: Fraunhofer Open Source SensorThingsAPI Server:. https://www.iosb.fraunhofer.de/servlet/is/80113/. Accessed: 2020-07-13.

[Je19] Jetzke, Tobias; Richter, Stephan; Ferdinand, Jan-Peter; Schaat, Samer: Künstliche Intelligenz im Umweltbereich. Umweltbundesamt, 2019.

[JFP14] Jankowski, Piotr; Fraley, Grant; Pebesma, Edzer: An exploratory approach to spatial decision support. Computers, Environment and Urban Systems, 45:101–113, 2014.

[Jo17] Joppa, Lucas N: , The case for technology investments in the environment, 2017.

[KBB19] Knoll, Lukas; Breuer, Lutz; Bach, Martin: Large scale prediction of groundwater nitrate concentrations from spatial data using machine learning. Science of the total environment, 668:1317–1327, 2019.

[LHK15] Liang, S; Huang, CY; Khalafbeigi, T: , SensorThings API Part 1: Sensing, OGC Doc. No. 15-078r6, 2015.

[Li06] Liesch, Tanja: Ermittlung der Grundwassergefährdung mit Hilfe künstlicher neuronaler Netze. Lehrstuhl für Angewandte Geologie der Univ., 2006.

[LK17] Liang, S; Khalafbeigi, T: , SensorThings API Part 2: Tasking Core, OGC Doc. No. 17-079r1, 2017.

[Ma15] Makantasis, Konstantinos; Karantzalos, Konstantinos; Doulamis, Anastasios; Doulamis, Nikolaos: Deep supervised learning for hyperspectral data classification through convolutional neural networks. In: 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). IEEE, S. 4959–4962, 2015.

[MD00] Maier, Holger R; Dandy, Graeme C: Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. Environmental modelling & software, 15(1):101–124, 2000.

[Mo18] Moßgraber, Jürgen; Hilbring, Désirée; van der Schaaf, Hylke; Hertweck, Philipp; Kontopoulos, Efstratios; Mitzias, Panagiotis; Kompatsiaris, Ioannis; Vrochidis, Stefanos; Karakostas, Anastasios: The sensor to decision chain in crisis management. In: ISCRAM. 2018.

[Mo19] Mo, Shaoxing; Zabaras, Nicholas; Shi, Xiaoqing; Wu, Jichun: Deep autoregressive neural networks for high-dimensional inverse problems in groundwater contaminant source identification. Water Resources Research, 55(5):3856–3881, 2019.

[NFL15] Nolan, Bernard T.; Fienen, Michael N.; Lorenz, David L.: A statistical learning framework for groundwater nitrate models of the Central Valley, California, USA. Journal of Hydrology, 531:902 – 911, 2015.

[PW19] Pete Warden, Daniel Situnayake: TinyML. O'Reilly Media, Inc.., 2019.

[Ra17] Ransom, Katherine M.; Nolan, Bernard T.; A. Traum, Jonathan; Faunt, Claudia C.; Bell, Andrew M.; Gronberg, Jo Ann M.; Wheeler, David C.; Z. Rosecrans, Celia; Jurgens, Bryant; Schwarz, Gregory E.; Belitz, Kenneth; M. Eberts, Sandra; Kourakos, George; Harter, Thomas: A hybrid machine learning model to predict and visualize nitrate concentration throughout the Central Valley aquifer, California, USA. Science of The Total Environment, 601-602:1160 – 1172, 2017.

[Ra19] Rahmati, Omid; Choubin, Bahram; Fathabadi, Abolhasan; Coulon, Frederic; Soltani, Elinaz; Shahabi, Himan; Mollaefar, Eisa; Tiefenbacher, John; Cipullo, Sabrina; Ahmad, Baharin Bin; Tien Bui, Dieu: Predicting uncertainty of machine learning models for modelling nitrate pollution of groundwater using quantile regression and UNEEC methods. Science of The Total Environment, 688:855 – 866, 2019.

[REN19] Rajaee, Taher; Ebrahimi, Hadi; Nourani, Vahid: A review of the artificial intelligence methods in groundwater level modeling. Journal of Hydrology, 572:336–351, Mai 2019.

[Ro14] Rodriguez-Galiano, Victor; Mendes, Maria Paula; Garcia-Soldado, Maria Jose; Chica-Olmo, Mario; Ribeiro, Luis: Predictive modeling of groundwater nitrate pollution using Random Forest and multisource variables related to intrinsic and specific vulnerability: A case study in an agricultural setting (Southern Spain). Science of The Total Environment, 476-477:189 – 206, 2014.

[Ro18] Rodriguez-Galiano, V.F.; Luque-Espinar, J.A.; Chica-Olmo, M.; Mendes, M.P.: Feature selection approaches for predictive modelling of groundwater nitrate pollution: An evaluation of filters, embedded and wrapper methods. Science of The Total Environment, 624:661 – 672, 2018.

[Sa18] Sajedi-Hosseini, Farzaneh; Malekian, Arash; Choubin, Bahram; Rahmati, Omid; Cipullo, Sabrina; Coulon, Frederic; Pradhan, Biswajeet: A novel machine learning-based approach for the risk assessment of nitrate groundwater contamination. Science of The Total Environment, 644:954 – 962, 2018.

[Si60] Simon, Herbert A: The new science of management decision. 1960.

[Su13] Suakanto, Sinung; Supangkat, Suhono H; Saragih, Roberd et al.: Smart city dashboard for integrating various data of sensor networks. In: International Conference on ICT for Smart Society. IEEE, S. 1–5, 2013.

[Vo18] VoPham, Trang; Hart, Jaime E; Laden, Francine; Chiang, Yao-Yi: Emerging trends in geospatial artificial intelligence (geoAI): potential applications for environmental epidemiology. Environmental Health, 17(1):40, 2018.

[We15] Weber, Frank-Andreas; Bergmann, Axel; Kämpf, Markus; Spinola, Anette; Gerdes, Heiko; Kludt, Christoph; Schüth, Christoph; Allendorf, Arnd; Mikat, Hermann; Berthold, Georg: Quantifizierung des Nitratabbauvermögens in den Grundwasserkörpern des Hessischen Rieds und Lokalisierung von Risikogebieten. 09 2015.

[Wi16] Wiener, Patrick; Stein, Manuel; Seebacher, Daniel; Bruns, Julian; Frank, Matthias; Simko, Viliam; Zander, Stefan; Nimis, Jens: Biggis: a continuous refinement approach to master heterogeneity and uncertainty in spatio-temporal big data (vision paper). In: Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. S. 1–4, 2016.

[Wi20] Wiener, Patrick; Zehnder, Philipp; Heyden, Marco; Philipp, Patrick; Riemer, Dominik: Fogsy: Towards Holistic Industrial AI Management in Fog and Edge Environments. KuVS-Fachgespräch Fog Computing 2020, S. 16, 2020.

[WLB18] Wunsch, Andreas; Liesch, Tanja; Broda, Stefan: Forecasting groundwater levels using nonlinear autoregressive networks with exogenous input (NARX). Journal of Hydrology, 567:743–758, 2018.

[Xi15] Xingjian, SHI; Chen, Zhourong; Wang, Hao; Yeung, Dit-Yan; Wong, Wai-Kin; Woo, Wang-chun: Convolutional LSTM network: A machine learning approach for precipitation nowcasting. In: Advances in neural information processing systems. S. 802–810, 2015.

[YJX17] Yu, Shiqi; Jia, Sen; Xu, Chunyan: Convolutional neural networks for hyperspectral image classification. Neurocomputing, 219:88–98, 2017.

R. Reussner, A. Koziolek, R. Heinrich (Hrsg.): INFORMATIK 2020,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2021 1081

Online-Überwachung von Chlor und Chlordioxid mittels optischer Spektroskopie

Machine Learning in der Überwachung der Trinkwasserqualität

Martin Wagner[1] Averil Fernandes[2] Gabriele Nüske[3]

Abstract: In der vorliegenden Arbeit wird gezeigt, wie UV/VIS-Spektren, die für die Überwachung von Desinfektionsmittelrestgehalten im Trinkwasser mit einem online-Spektrometer aufgenommen werden, mit Methoden des maschinellen Lernens ausgewertet werden. Es wurden Regression-Pipelines für die Bestimmung der Konzentration von freiem Chlor und Chlordioxid in Trinkwasser im Bereich zwischen 0,1 mg/L und 1,0 mg/L erstellt. Der Root Mean Squared Error (RMSE) der Kalibrierung beträgt 0,03 mg/L (Chordioxid) und 0,05 mg/L (freies Chlor). Die Anwendung der Methode wird am Beispiel einer Desinfektionsanlage in Haridwar, Indien, demonstriert.

Keywords: Freies Chlor; Chlordioxid; UV/VIS Spektroskopie; online; Überwachung; Principal Components Analysis; Linear Regression; Neural Network; Machine Learning

1 Einleitung

Die Desinfektion von Wasser ist weltweit ein wichtiges Thema. Vor allem Trinkwasser, das aus Oberflächenwässer gewonnen wird, muss mittels Chlor, Chlordioxid oder UV-Desinfektion behandelt werden, um Bakterien und Viren zu inaktivieren. Ein bedeutender Aspekt ist dabei die Überwachung des Prozesses um jederzeit eine sichere Desinfektion gewährleisten zu können.

Die Trinkwasseraufbereitung erfolgt in mehreren Stufen und variiert von der Art der Wasserquelle (Grundwasser, Flusswasser, Talsperre, Uferfiltrat, etc.) und der Wasserqualität. Die Überwachung der Wasserqualität erfolgt dabei sowohl direkt vor Ort mittels online- und onsite-Messgeräten als auch im Labor. Online werden vor allem einfach zu bestimmende Parameter, wie das Redoxpotential, die elektrische Leitfähigkeit und die Trübung ermittelt. Komplexere Parameter wie organische Spurenstoffe (z. B. Arzneimittelrückstände) und mikrobiologische Parameter werden in regelmäßigen Abständen im Labor bestimmt.

[1] TZW: DVGW-Technologiezentrum Wasser, Außenstelle Dresden, Wasserwerkstr. 2, 01326 Dresden, Deutschland, martin.wagner@tzw.de
[2] TZW: DVGW-Technologiezentrum Wasser, Außenstelle Dresden, Wasserwerkstr. 2, 01326 Dresden, Deutschland
[3] TZW: DVGW-Technologiezentrum Wasser, Außenstelle Dresden, Wasserwerkstr. 2, 01326 Dresden, Deutschland, gabriele.nueske@tzw.de

doi:10.18420/inf2020_102

Die initial zugegebene Konzentration von Desinfektionsmitteln wie Chlor oder Chlordioxid verringert sich über die Zeit durch die Reaktion mit Wasserinhaltsstoffen. Die Dosierung erfolgt in der Praxis daher so, dass eine bestimmte Restkonzentration nach einer Reaktionszeit von 30 min eingehalten wird. Die vorgeschriebene Restkonzentration beträgt in Deutschland zwischen 0,1 mg/L und 0,3 mg/L für Chlor sowie zwischen 0,05 mg/L und 0,20 mg/L für Chlordioxid. In anderen Ländern werden dagegen durchaus höhere Restkonzentrationen gefordert.

Die Online-Überwachung von Desinfektionsmittelrestgehalten, wie Chlor und Chlordioxid, erfolgt in der Regel mittels amperometrischen (elektrochemischen) Sensoren. Die optische Spektroskopie wurde bisher für die Überwachung von Desinfektionsmittelrestgehalten noch nicht verwendet. Der Vorteil ist jedoch, dass mit einem Spektrometer neben dem Desinfektionsmittel auch eine Vielzahl weiterer Wasserqualitätsparameter erfasst werden können. Dazu gehören die Trübung, Nitrat sowie die Summe der gelösten organischen Verbindungen (DOC, Dissolved Organic Carbon). In der vorliegenden Arbeit wird gezeigt, wie Restgehalte von freiem Chlor und Chlordioxid quantitativ ermittelt werden können.

Die Herausforderung für die Bestimmung von Chlor und Chlordioxid mittels UV/VIS-Spektren besteht darin, die schwachen Absorptionssignale der vergleichsweise niedrigen Restgehalte zu erfassen und vom Hintergrund (durch natürliche Wasserinhaltsstoffe hervorgerufene Absorption) zu trennen (siehe Abb. 1). Dies wurde in dieser Arbeit durch die Anwendung von Verfahren des maschinellen Lernens erzielt.

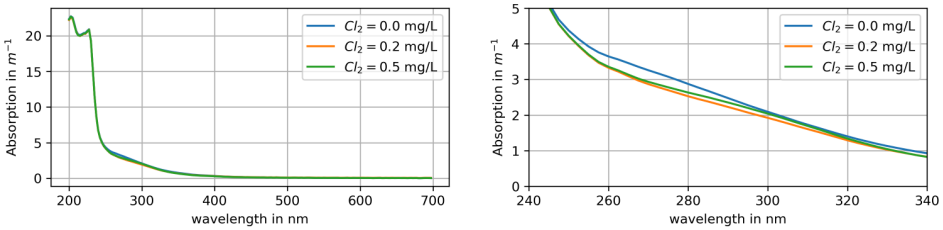


Abb. 1: UV/VIS-Spektren einer Wasserprobe mit unterschiedlichen Chlorkonzentrationen. Links: das vollständige Spektrum, welches hauptsächlich durch Nitrat und DOC geprägt ist. Rechts: Vergrößerung des Wellenlängenbereiches zwischen 240 nm und 340 nm, in welchem Chlor absorbiert.

Chlor liegt im Wasser in Abhängigkeit des pH-Wertes in den drei verschiedenen Formen Cl_2 (sauer), HOCl (neutral) und OCl^- (basisch) vor. Der zulässige pH-Wert Bereich von Trinkwässern in Deutschland liegt zwischen 6,5 und 9,5, so dass nur die beiden Spezies hypochlorige Säure (HOCl) und Hypochlorit (OCl^-) von Bedeutung sind. Beide weisen unterschiedliche Absorptionseigenschaften auf, dargestellt in Abb. 2.

Im Gegensatz zu Chlor ist Chlordioxid im trinkwasserrelevanten Bereich unabhängig vom pH-Wert. Die Absorption ist daher direkt von der Konzentration abhängig.

Die quantitative Analyse von Spektren durch multivariate Datenanalysemethoden ist bereits

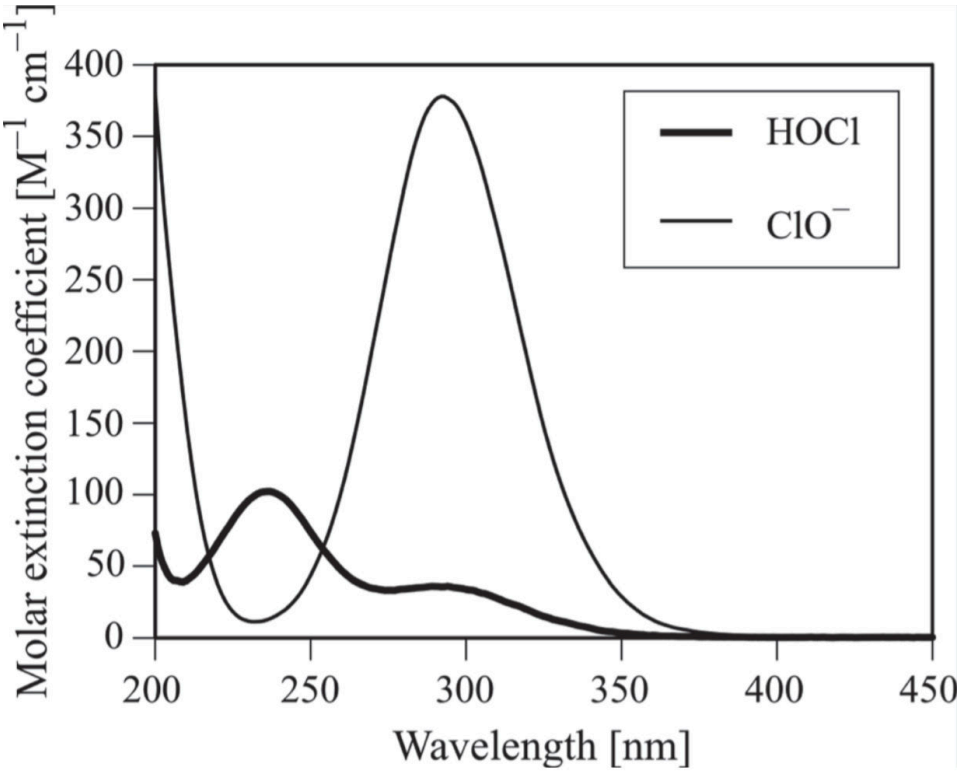


Abb. 2: UV/VIS-Spektren von HOCl und OCl^-. [Ki19].

seit vielen Jahren gängige Praxis. Standardmethoden sind hier die Hauptkomponentenanalyse zur Dimensionsreduktion in Kombination mit einer einfachen multiplen linearen Regression (Principal Components Regression, PCR). Ein weiteres Standardverfahren stellt darüber hinaus die Partial Least Squares Regression (PLSR) dar [Ke06].

Die spektroskopische Bestimmung von Chlor und Chlordioxid erfolgte bisher nur für hochkonzentrierte Stammlösungen im Bereich zwischen 20 mg/L und 100 mg/L (Chlordioxid) bzw. 150 mg/L bis 450 mg/L (Chlor) [Wa11]. Das Ziel dieser Arbeit liegt in der Erfassung von Konzentrationen < 1 mg/L.

2 Material und Methoden

2.1 Datengrundlage

Die Grundlage bilden ca. 300 UV/VIS-Spektren, die mit dem Absorptionsspektrometer Spectro::lyser der Firma s::can im Wellenlängenbereich zwischen 200 nm und 750 nm ($\Delta\lambda = 2,5$ nm) und einer optischen Weglänge von 10 cm aufgenommen wurden.

Der Datensatz umfasst einerseits Kalibriermessreihen von sowohl Chlor und Chlordioxid als auch Mischungen von Chlor und Chlordioxid. Diese wurden bei definierten pH-Werten in Pufferlösungen (7,5 bis 9,5) bzw. synthetischen Modellwässern aufgenommen, die eine vernachlässigbare Eigenabsorption aufweisen. Darüber hinaus wurden auch zahlreiche Versuche durchgeführt, bei denen Realwasserproben mit unterschiedlichen Chlor-/Chlordioxidmengen versetzt wurden und deren Restgehalte nach einer Reaktionszeit von 30 min bestimmt wurden. Eine Übersicht dazu gibt Tab. 1.

Tab. 1: Datengrundlage (Anzahl Spektren) für die Modelle von Chlordioxid und Chlor.

Dataset	Chlordioxid	Freies Chlor
Kalibration Cl2	78	78
Kalibration Cl2/ClO2	7	11
Kalibration ClO2	59	60
Monochloramin	65	3
Realwasser + Cl2	58	58
Realwasser + Cl2/ClO2	5	5
Realwasser + ClO2	21	21
Sonstige	12	3
Summe	305	239

Wie aus Tab. 1 ersichtlich, wurde für jede Verbindung (Chlor, Chlordioxid) ein separates Modell erstellt. Der Datensatz Monochloramin beinhaltet Spektren von monochloraminhaltigen synthetischen Lösungen. Monochloramin entsteht als Nebenprodukt bei der Desinfektion von ammoniumhaltigen Wässern. Der Datensatz Sonstige beinhaltet verschiedene Blindwerte von desinfektionsmittelfreien Pufferlösungen.

Als Referenzverfahren zur Bestimmung von Chlor sowie Chlordioxid diente das DPD-Verfahren.

2.2 Modellierung

Bei der Bestimmung der Chlor-/Chlordioxidkonzentration handelt es sich um ein Regressionsproblem, zur dessen Lösung im Wesentlichen eine Dimensionsreduktion in Kombination mit einem Regressor (Lineare Regression bzw. neuronales Netz) benutzt wurden. Die Dimensionsreduktion erfolgte durch eine Hauptkomponentenanalyse (engl. Principal Components Analysis, PCA). Es handelt sich dabei um ein bewährtes Verfahren zur Entrauschung und Dekorrelation von Daten, bei dem vereinfacht ausgedrückt korrelierende Features zu Hauptkomponenten zusammengefasst werden und somit die Anzahl der Eingangsgrößen reduziert wird. Die Dimensionsreduktion und Dekorrelation sind bei dem vorliegenden Problem zwingende Vorrausetzungen für die Anwendung einer linearen Regression, da

- Absorptionsspektren eine hohe Korrelation zwischen den einzelnen Wellenlängen (Features) aufweisen, so dass eine Multikollinearität vorliegt und
- die Anzahl der Features (256 Werte pro Spektrum) in derselben Größenordnung wie die Anzahl der Spektren liegt, so dass ein unterbestimmtes Gleichungssystem im Fall von Chlordioxid vorliegt.

Für beide Desinfektionsmittel wurde je ein mehrstufiges Modell (Pipeline) erstellt, dargestellt in Abb. 3.

Der erste Schritt besteht in der Berechnung der Ableitung der Spektren. Diese erfolgt durch einen Savitzky-Golay Filter (gleitendes Polynom) und zwar unabhängig für jedes Spektrum. Durch die Bildung der Ableitung können Basislinienveränderungen im Spektrum korrigiert werden und zudem spektrale Eigenschaften besser hervorgehoben werden. Anschließend wird für die weiteren Berechnungen nur der Wellenlängenbereich zwischen 240 nm und 450 nm betrachtet, da Chlor und Chlordioxid nur hier eine Absorption aufweisen. Im nächsten Schritt erfolgt eine Mittenzentrierung mit dem Median und die Skalierung mit dem Interquartilabstand (Robust Scaling) für jedes Feature (Wellenlänge). Mit Hilfe einer Hauptkomponentenanalyse erfolgt nun eine Dimensionsreduktion auf k Komponenten. Im Fall von Chlordioxid erfolgt abschließend eine einfache multiple lineare Regression. Im Fall von Chlor erfolgt erneut eine robuste Skalierung für den mittels PCA auf k Features transformierten Datensatz (Scores), um sicherzustellen, dass alle k Features mit gleicher Wichtung als Eingangsdaten in ein neuronales Netz (Multilayer Perceptron, MLP-Regressor) eingehen. Bei neuronalen Netzen handelt es sich um eine Klasse von Algorithmen, die aufgrund ihrer flexiblen Konfigurierbarkeit (Architektur) sehr leistungsstark sind und auch nichtlineare Probleme sehr gut approximieren können.

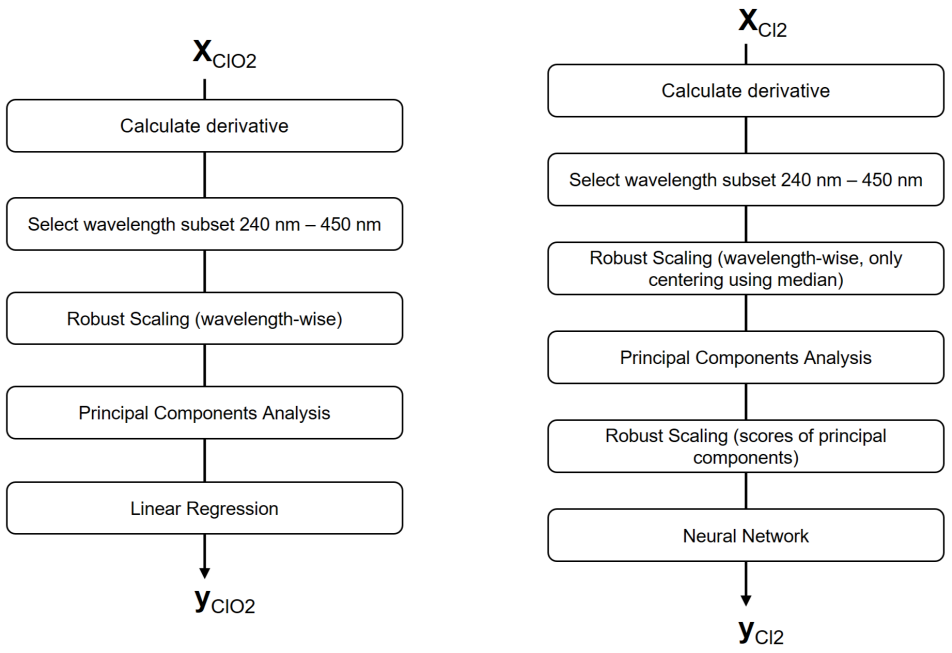


Abb. 3: Pipelines für Chlordioxid (links) und Chlor (rechts).

Für beide Modelle wurde ein Hyperparametertuning durchgeführt, mit den in Tab. 2 und Tab. 3 dargestellten Parametern.

Tab. 2: Hyperparametertuning Chlordioxid.

Schritt	Parameter	Werte	Bester Wert
Derivative	Window size	15, 25, 35	35
Derivative	Order	0, 1, 2	1
PCA	Number of components	5, 6, 7, 8, 9, 10, 11, 12	12

Tab. 3: Hyperparametertuning Chlor.

Schritt	Parameter	Werte	Bester Wert
Derivative	Window size	25, 35	35
Derivative	Order	0, 1, 2	2
PCA	Number of components	5, 6, 7, 8, 9, 10, 11, 12	12
Neural Network	Number of units of hidden layer	1-layer: 2, 5, 10	10

Unveränderliche Parameter des MLP-Regressor, die nicht im Hyperparameter Tuning verändert wurden, sind in Tab. 4 aufgeführt.

Tab. 4: Hyperparameter des MLP-Regressors.

Parameter	Wert
Aktivierungsfunktion Hidden Layer	tanh
Aktivierungsfunktion Output Layer	softplus
Solver	Stochastic Gradient Descent
Learning Rate	1e-3
Regularization	0
Momentum	0
Weights Initialization	Xavier

Das Training der Modelle erfolgte mittels einer 5fachen Kreuzvalidierung, wobei 20 % der Spektren als unabhängiger Testdatensatz verwendet wurden und 80 % der Spektren für 5-fache Kalibrierung/Validierung. Die Zuordnung erfolgte dabei durch eine geschichtete Zufallsstichprobe wobei als Schichtung zwei Konzentrationsklassen (Desinfektionsmittelgehalt $\leq 0,1$ mg/L, Desinfektionsmittelgehalt $> 0,1$ mg/L) verwendet wurde. Als Metrik wurde der Root Mean Squared Error (RMSE) verwendet und auf Basis der Kreuzvalidierung insgesamt drei Maße abgeleitet:

- RMSEC (Calibration),
- RMSEV (Validation),
- RMSEP (Prediction).

2.3 Praxisanwendung in Indien

Zum Nachweis der Anwendbarkeit wurde das Verfahren der spektroskopischen Chlorbestimmung im Rahmen eines BMBF-Projektes in Indien angewandt (Stadt Haridwar am Ganges in Nordindien). Hierzu wurde eine Anlage, die Chlor für die Desinfektion von Uferfiltrat erzeugt, mit einem online-Spektrometer ausgerüstet und die Chlorkonzentration im Trinkwasser unmittelbar nach der Desinfektion gemessen. Der Förderbrunnen wurde aufgrund von Wasserknappheit nur zweimal täglich für jeweils vier Stunden betrieben. Die Spektren wurden in einem Intervall von 15 Minuten aufgenommen. Mindestens einmal die Woche wurde zudem die Chlorkonzentration photometrisch mit einem vor-Ort-Messgerät mittels DPD bestimmt.

Im Rahmen der Praxisanwendung erfolgte zudem eine sogenannte vor-Ort-Kalibrierung. Hier wird ein Modell für ein spezifisches Wasser, in diesem Fall das Uferfiltrat des Ganges, erstellt. Dabei werden ausschließlich Wasserproben des Ganges für das Modelltraining verwendet. Weitere Trainingssamples können erzeugt werden, indem die Wasserproben zusätzlich mit verschiedenen Chlormengen versetzt werden, um den Umfang des Datensatzes zu erhöhen. Der Vorteil ist eine höhere Genauigkeit, jedoch ist eine Übertragbarkeit auf andere Standorte nicht gewährleistet.

3 Ergebnisse und Diskussion

3.1 Modellentwicklung

Die Ergebnisse des Chlordioxidmodells sind in Abb. 4 dargestellt, in der die wahren (gemessenen) Konzentrationen und die auf Basis des Modells berechneten Konzentrationen aufgetragen sind.

Bereits mit einer PCR-Regression wird eine zufriedenstellende Modellgüte erzielt. Die RMSE-Werte (für Calibration, Validation und Prediction) liegen zwischen 0,026 mg/L und 0,037 mg/L, was zufriedenstellend ist. Die Dosierung von Chlordioxid erfolgt in der Praxis mit weitaus höheren Konzentrationen (bis 0,4 mg/L). Erfolgt daher die Überwachung der Chlordioxidkonzentration unmittelbar nach der Dosierung, so besitzt das Modell eine ausreichende Güte. Nur bei der Überwachung der Restkonzentration im Verteilungsnetz muss die Empfindlichkeit des Modells erhöht werden, um auch Restkonzentrationen von 0,05 mg/L sicher nachweisen zu können. Die Ursache für die geringen Restkonzentrationen im Netz ist die Zehrung bzw. Reaktion mit anderen Wasserinhaltsstoffen.

Für die Bestimmung von Chlor wurden zwei Modelle erstellt, deren Güte in Abb. 5 dargestellt ist.

Die PCR-Regression als Benchmark weist eine ungenügende Güte auf. Ursache ist die starke Abhängigkeit des Chlors vom pH-Wert. Die untersuchten Wässer weisen einen pH-Wert

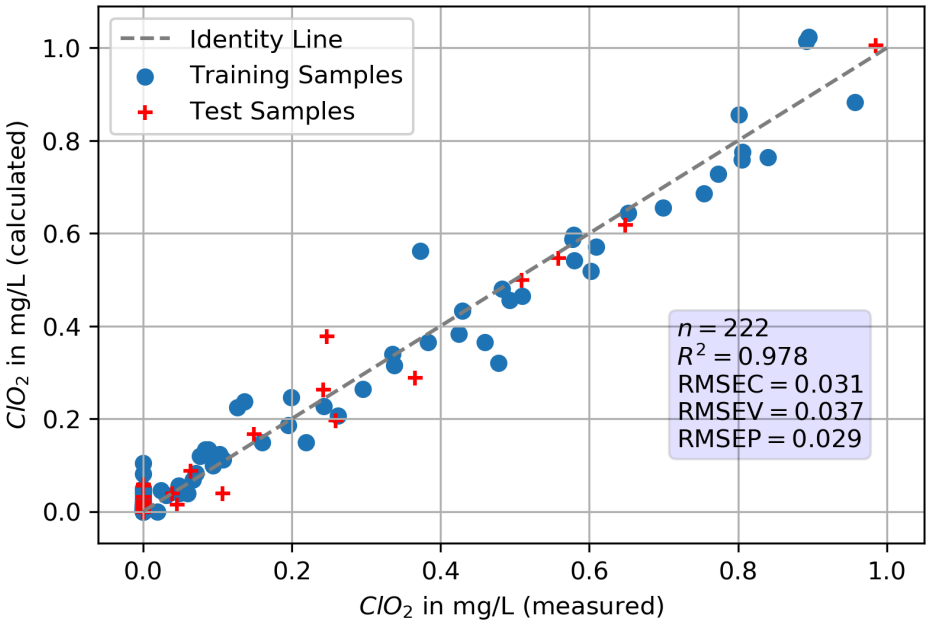


Abb. 4: Modellgüte des PCR-Modells für Chlordioxid.

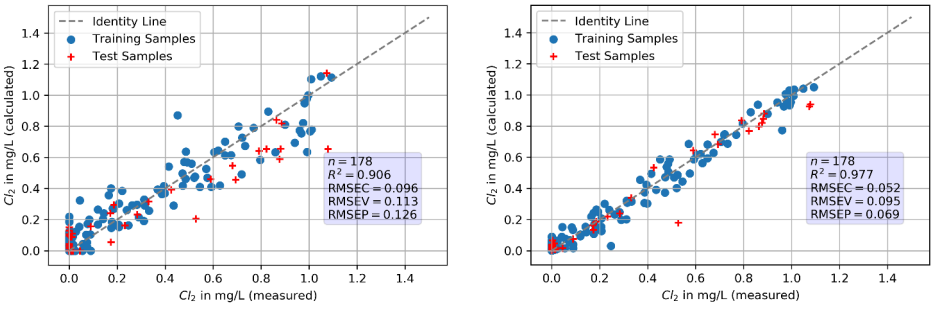


Abb. 5: Modellgüte des Benchmark (PCR-Regression) sowie der PCA-MLPR (rechts) für die Bestimmung von freiem Chlor.

zwischen 7,5 und 9,5 auf (siehe Abb. 6). In diesem Bereich liegt Chlor in den Spezies hypochlorige Säure (HOCl) und Hypochlorit (OCl^-) vor. Beide Spezies absorbieren jedoch UV-Licht unterschiedlich stark. Als Maß für die Absorption wird der Extinktionskoeffizient ϵ (Einheit L mol^{-1} cm^{-1}) verwendet; je höher ϵ, desto stärker ist das Absorptionssignal. Hierbei zeigt vor allem Hypochlorit eine starke Absorption mit einem ϵ von ca. 350 L mol^{-1} cm^{-1}, während die hypochlorige Säure mit einem ϵ von 100 L mol^{-1} cm^{-1} einen wesentlich geringeren Extinktionskoeffizienten aufweist [Ki19].

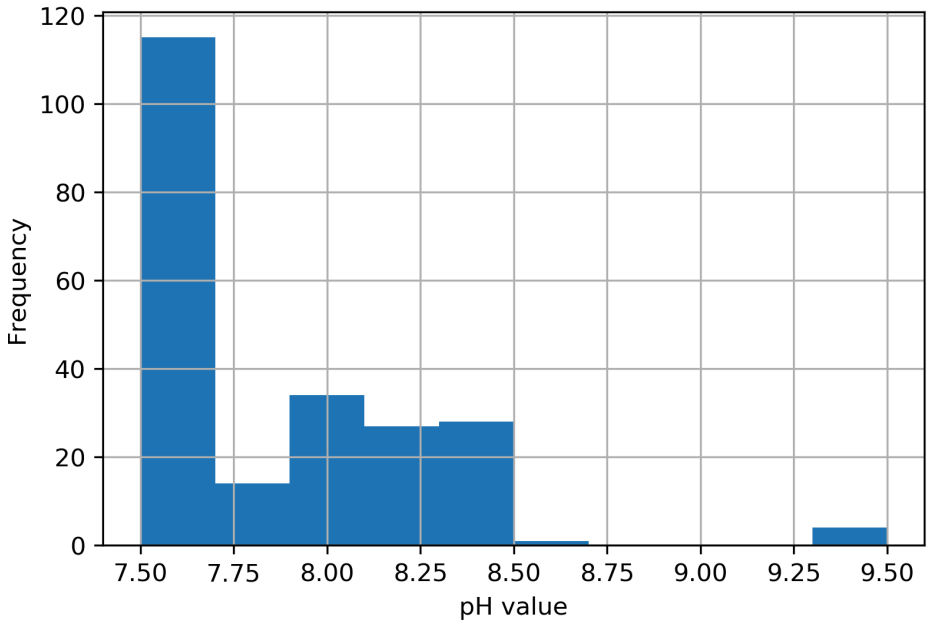


Abb. 6: Histogramm des pH-Wertes der analysierten Wasserproben (n = 223).

Die Modellgüte der PCA-MLPR ist dagegen wesentlich besser als die des PCR-Benchmark. Offenbar besteht zwischen den Scores der PCA und der Chlorkonzentration eine nichtlineare Beziehung, die durch den MLP-Regressor besser dargestellt werden kann als durch eine lineare Regression.

Aus der Modellgüte lassen sich Nachweis- und Bestimmungsgrenzen berechnen [RE11], die in der Analytik wichtige Parameter für die quantitative Bestimmung von Wasserqualitätsparametern darstellen. Die Nachweisgrenze gibt an, ob eine Messgröße (Chlor, Chlordioxid) gerade noch zuverlässig bestimmt werden kann (ja/nein-Entscheidung). Die Bestimmungsgrenze ist diejenige Konzentration, ab der eine Konzentration für die Messgröße angegeben werden kann. Die erzielten Nachweis- und Bestimmungsgrenzen für Chlor und Chlordioxid sind in Tab. 5 zusammengefasst.

Tab. 5: Nachweis- und Bestimmungsgrenzen des Chlor- und Chlordioxidmodells.

Messgröße	Nachweisgrenze in mg/L	Bestimmungsgrenze in mg/L
Chlor (PCA-MLPR)	0,084	0,302
Chlordioxid (PCR)	0,054	0,196

3.2 Praxisanwendung

Für eine Anwendung des Verfahrens in Deutschland wäre mindestens eine Bestimmungsgrenze von 0,1 mg/L für Chlor bzw. 0,05 mg/L für Chlordioxid wünschenswert. Dies ist mit 0,3 mg/L bzw. 0,2 mg/L noch nicht gegeben. Es ist jedoch davon auszugehen, dass mit einer lokalen Kalibrierung weitaus geringe Bestimmungsgrenzen erzielt werden können. Die in dieser Arbeit gezeigten Modelle basieren auf einer Vielzahl verschiedener Wasserproben unterschiedlicher Herkunft (Oberflächenwasser, Grundwasser). Bei einer lokalen Kalibrierung wird dagegen ein Modell für ein spezifisches Wasser, also beispielsweise für ein konkretes Wasserwerk erstellt. Nach diesem Prinzip wurde das Verfahren an einem Wasserwerk in Indien eingesetzt. Im Rahmen eines BMBF-Vorhabens wurde dort eine Anlage zur Desinfektion von Uferfiltrat mit Chlor installiert und mit einer online-Spektrometersonde zur Überwachung der Chlorkonzentration ausgestattet. Dabei konnte das in dieser Arbeit vorgestellte Verfahren zum Nachweis von Chlor mittels optischer Spektroskopie im realen Praxisbetrieb angewendet werden. Hierzu wurde das Spektrometer über USB an einem Laptop angeschlossen, so dass Datenübertragung und –auswertung in Echtzeit erfolgte. Für die grafische Darstellung wurde ein Dashboard basierend auf der Pythonbibliothek *bokeh* entwickelt. Somit konnte die aktuelle Chlorkonzentration während des Betriebes jederzeit vom Anlagenbetreiber abgelesen werden. Die Ergebnisse sind in Abb. 7 am Beispiel eines Tages (09.02.2018) dargestellt.

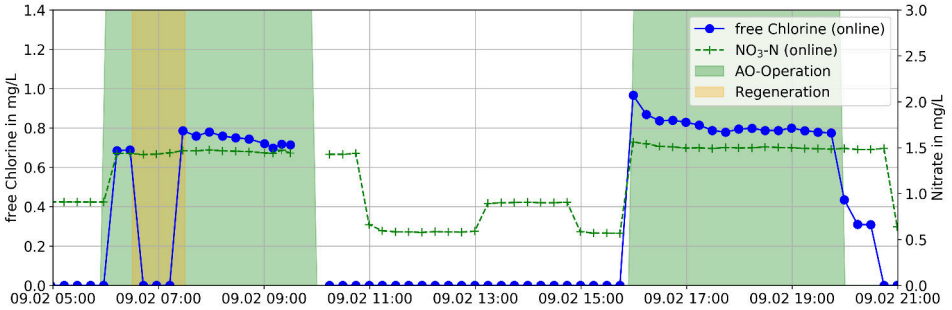


Abb. 7: Bestimmung von Chlor mittels optischer Spektroskopie im realen Praxisbetrieb an einer Anlage in Indien.

Die Anlage (AO) war aufgrund von Wassermangel nur zweimal täglich für jeweils vier Stunden im Betrieb (grün dargestellt). Mit Hilfe der online-Spektrometersonde wurde sowohl freies Chlor als auch Nitrat ermittelt. Die Anlage wurde zusammen mit der Spektrometersonde über einen Zeitraum von drei Monaten betrieben. In dieser Zeit erfolgten

Vergleichsmessungen (Bestimmung der Chlorkonzentration mit dem Referenzverfahren DPD). Der erzielte RMSEP beträgt 0,051 mg/L (siehe Abb. 8) und ist damit niedriger als der RMSEP des in Abb. 5 dargestellten Modells. Dies zeigt, dass mittels einer lokale Kalibrierung eine höhere Modellgüte erzielt werden kann. Dies liegt unter anderem daran, dass der pH-Wert des Wassers nahezu konstant 7,5 betrug.

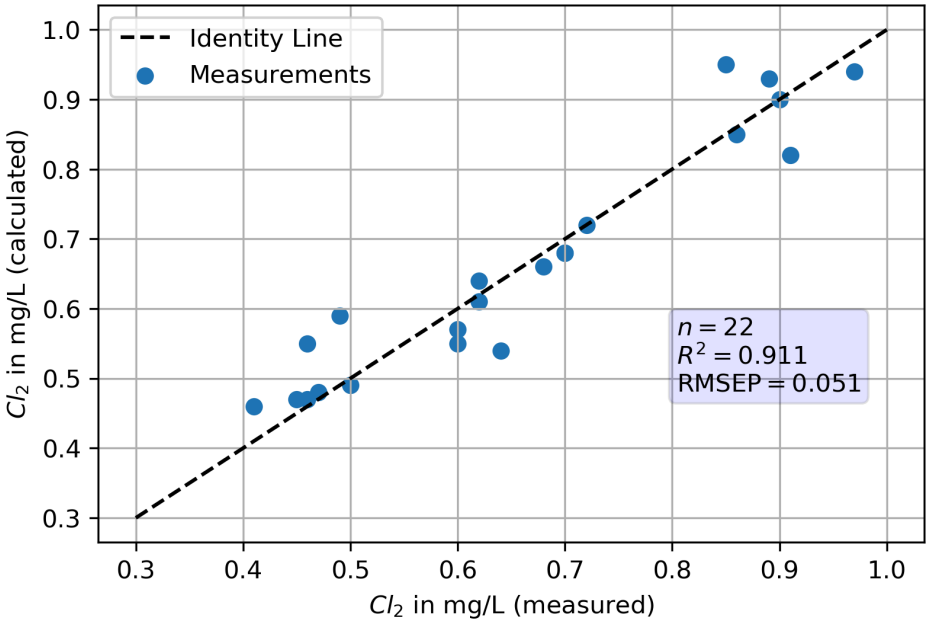


Abb. 8: Güte der Chlorbestimmung unter Praxisbedingungen in Indien.

4 Zusammenfassung und Ausblick

In der Arbeit wurde eine Methode gezeigt, wie mit Hilfe von maschinellem Lernen eine quantitative Spektrenanalyse zur Bestimmung von Chlor als auch Chlordioxid auf Basis von UV/VIS-Absorptionsspektren erzielt werden kann. Dabei wurden Nachweis- und Bestimmungsgrenzen von 0,08 mg/L und 0,30 mg/L für Chlor sowie 0,05 mg/L und 0,20 mg/L erreicht.

Eine grundsätzliche Herausforderung bei der Modellierung besteht darin, dass die Absorption von Chlor und Chlordioxid im Konzentrationsbereich von kleiner einem mg/L im Vergleich zur Absorption anderer Wasserinhaltsstoffe (natürliche organische Kohlenstoffverbindungen, DOC) sehr gering ist. Darüber hinaus zeigt Chlor eine ausgeprägte pH-Wert-Abhängigkeit, was eine exakte Quantifizierung zusätzlich erschwert. Dennoch

konnte mit einer Hauptkomponentenanalyse und einem neuronalen Netz (Multilayer Perceptron) eine zufriedenstellende Modellgüte erzielt werden. Die Anwendbarkeit der Methode wurde im Realbetrieb an einer Desinfektionsanalage in Haridwar, Nordindien, nachgewiesen.

Der Vorteil der spektroskopischen Bestimmung von Desinfektionsmittelrestgehalten (Chlor, Chlordioxid) besteht darin, dass die Methode einfach zu benutzen und zudem wartungsarm ist. Darüber hinaus können mit einem online-Spektrometer auch weitere wichtige Wasserqualitätsparameter wie der DOC, die Färbung, Trübung und der Nitratgehalt bestimmt werden.

Bereits jetzt besitzt die Methode damit vor allem in Entwicklungsländern ein hohes Potential, da dort nur unzureichendes Personal und eingeschränkte Laborkapazitäten für die Überwachung von Wasseraufbereitungsprozessen zur Verfügung stehen.

Dennoch besteht Verbesserungspotential, indem mehr Daten sowie der pH-Wert des Wassers als zusätzliche Eingangsgröße verwendet wird. Damit ist die Methode perspektivisch auch für Deutschland geeignet, wenn entweder höhere Restkonzentrationen im Wasser vorliegen oder die Empfindlichkeit gesteigert wird.

5 Danksagung

Diese Arbeit wurde vom deutschen Bundesministerium für Bildung und Forschung (BMBF) im Rahmen der Projekte *„Sichere und nachhaltige Trinkwassergewinnung in Indien durch Kopplung von naturnahen und innovativen Verfahren* (NIRWINDU, (Förderkennzeichen 02WCL1356B) sowie *Wissenschaftlich begründeter Masterplan Uferfiltration zur Trinkwasserversorgung in Vietnam* (AquaViet, Förderkennzeichen 02WCL1472C) unterstützt.

Literaturverzeichnis

[Ke06] Kessler, Waltraud: Multivariate Datenanalyse für die Pharma-, Bio- und Prozessanalytik. Wiley-VCH, 2006.

[Ki19] Kishimoto, Naoyuki: State of the Art of UV/Chlorine Advanced Oxidation Processes: Their Mechanism, Byproducts Formation, Process Variation, and Applications. Journal of Water and Environment Technology, 17(5):302–335, 2019.

[RE11] Reichenbächer, Manfred; Einax, Jürgen W.: Challenges in Analytical Quality Assurance. Springer, 2011.

[Wa11] Wang, Qiang; Chen, Kefu; Li, Jun; Xu, Jun; Liu, Shanshan: Simultaneous determination of chlorine dixoide and hypochlorous acid in bleaching systems. BioResources, 6(2):1868–1879, 2011.

R. Reussner, A. Koziolek, R. Heinrich (Hrsg.): INFORMATIK 2020,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2021 1095

Quantifizierung von Zielkonflikten globaler Landnutzung mit Hilfe mehrdimensionaler Optimierung und LPJ-GUESS

Sven Lautenbach[1], Anita D. Bayer[2], Almuth Arneth[3]

Abstract: Die Art und Weise, wie Land genutzt wird hat entscheidenden Einfluss darauf, welche Güter und Dienstleistungen aus der Natur dem Menschen zur Verfügung stehen. Landwirtschaftliche Produktion nimmt weltweit große Flächen in Anspruch, um die wachsende Weltbevölkerung mit Nahrungsmitteln zu versorgen. Die damit einhergehenden Veränderungen der Vegetationsbedeckung haben bedeutende Auswirkungen auf vielfältige ökologische Dienstleistungen. Angesichts der globalen Herausforderungen scheint es daher angebracht, darüber nachzudenken, welches Optimierungspotential in der Verteilung der landwirtschaftlichen Flächen besteht. Wir untersuchen hier, welche Zielkonflikte zwischen Nahrungsmittelproduktion, Kohlenstoffspeicherung und Wasserverfügbarkeit auf globaler Ebene bestehen. Hierzu wurden Szenariorechnungen des dynamischen globalen Vegetationsmodells LPJ-GUESS genutzt, um Pareto-optimale Lösungen verschiedener Landnutzungszuweisungen hinsichtlich der drei Zieldimensionen zu schätzen. Als Optimierungsverfahren kam ein genetischer Algorithmus (NSGA-II) zum Einsatz. Die Ergebnisse zeigen deutliche Optimierungspotentiale gegenüber der heutigen Landutzungskonfiguration, die jedoch für einzelne Weltregionen nicht unerhebliche Veränderungen nach sich ziehen würden.

Keywords: Optimierung; Genetischer Algorithmus; Zielkonflikte; Landnutzung; Ökologische Dienstleistungen

1 Einleitung

Die Nutzung der knappen Ressource Land führt zwangsläufig zu Zielkonflikten im Bezug auf ökologische Dienstleistungen [LL16, BPG09]. Unterschiedliche Nutzungen führen dazu, dass die Produktion einzelner essentieller Güter und Dienstleistungen auf Kosten anderer ausgeweitet wird [Fo05]. Das wichtige Ziel der Nahrungsmittelsicherheit für die wachsende Weltbevölkerung steht beispielsweise im prominenten Zielkonflikt mit der Wasserversorgung und der Speicherung von Kohlenstoff in der Vegetation und im Boden. Als mögliche Lösungen werden das Schließen von Ertragslücken, einer Umstellung der Ernährung auf vegane Kost und das Vermeiden von Nahrungsmittelverlusten diskutiert [Fo11]. Es stellt sich aber auch die Frage, ob Alternativen zur historisch gewachsenen Verteilung

[1] HeiGIT an der Universität Heidelberg, Schloss-Wolfsbrunnenweg 33, D-69118 Heidelberg, Germany, sven.lautenbach@heigit.org

[2] Karlsruhe Institute of Technology, Institute of Meteorology and Climate Research (IMK-IFU), 82467 Garmisch-Partenkirchen, Germany, anita.bayer@kit.edu

[3] Karlsruhe Institute of Technology, Institute of Meteorology and Climate Research (IMK-IFU), 82467 Garmisch-Partenkirchen, Germany, Almut.arneth@kit.edu

doi:10.18420/inf2020_103

der Anbaugebiete, die unter anderem durch Agrarsubventionen und Handelsabkommen stabilisiert wird, existieren. Es gibt zahlreiche Hinweise, dass Zielkonflikte der Landnutzung räumlich variieren [Sc18b, La17, La13, La14, Gr14], woraus sich die Frage ergibt, ob die Möglichkeiten besteht, die Verteilung der Landwirtschaft auf globaler Ebene so zu modifizieren, dass mehr Nahrungsmittel und mehr Wasser verfügbar sind und gleichzeitig mehr Kohlenstoff in terrestrischen Senken gespeichert werden kann.

Bisherige Forschungen nähern sich dieser Frage auf der globalen Skala in Form von Szenarioanalysen, bei denen die Konsequenzen unterschiedlicher möglicher Entwicklungen untersucht werden [Kr17]. Veränderungen der Landnutzung werden dabei in der Regel aus abstrakten sozio-ökonomischen Szenarienannahme abgeleitet oder anhand von Landnutzungsmodellen wie Clumondo [vAV12] oder MAgPIE [Po17] bzw. integrierten Bewertungsmodellen wie IMAGE [St14] modelliert. Dies bilden dann die Grundlage für Simulationsmodelle, welche die Auswirkung der Landnutzungsveränderungen auf unterschiedliche Zielgrößen wie ökologische Dienstleistungen bewerten. Analysen, die versuchen Handlungsspielräume jenseits einer beschränkten Anzahl von Szanarien zu erkunden sind selten [SLV13]. Als zielführend hierfür für die Analyse von Zielkonflikten zwischen Landnutzungsoptionen und dem Aufdecken von Handlungsoptionen hat sich die Simulation von Pareto-Fronten mittels Optimierungsansätzen erwiesen. Auf der lokalen bis regionalen Skala finden sich einige Studien die sich mit diesen Analysen beschäftigt haben [Sc18b, La13, La14, Ve18, Br15], jedoch fehlte bisher eine Betrachtung auf globaler Ebene.

Hier setzt die vorliegende Arbeit an, die folgenden Forschungsfragen nachgeht:

1. Wie groß ist das Potential, durch Veränderungen der räumlichen Allokation von landwirtschaftlichen Anbaugebieten Verbesserungen im Bezug auf Nahrungsmittelproduktion, Wasserverfügbarkeit und Kohlenstoffspeicherung zu erziehlen?
2. Welche Verschiebungen in den Anbauregionen wären dafür notwendig? Welche Regionen sind besonders stark von Zielkonflikten betroffen? Lassen sich Regionen hinsichtlich der Zielkonflikte eingruppieren?
3. Wie groß sind die Unterschiede auf die vorherigen Fragen, wenn verschiedene Planungshorizonte und Emissionsszenarien betrachtet werden? Verschieben sich dadurch die regional dominierenden Zielkonflikte?

Um dieser Frage nachzugehen, wurden Pareto-optimale Landnutzungskonfigurationen hinsichtlich der drei genannten ökologischen Dienstleistungen durch Kombination eines dynamischen Vegetationsmodells und eines genetischen Algorithmus geschätzt. Die Lösungen der Pareto-Front wurden dann hinsichtlich der räumlichen Verteilung der Landnutzung und der Produktion der ökologischen Dienstleistungen untersucht. Die Analyse berücksichtigt die Auswirkung des Klimawandels in Form von Klimaszenarien für zwei Treibhausgas-Emissionsszenarien. Die OPtimierung der Landnutzung erfolgte jeweils für einen kurzen bis mittleren (in cirka 20 Jahren) und einen langen (Ende des 21. Jhds.) Planungshorizont.

2 Methoden

Die Optimierung, auf deren Grundlage die Pareto-Front geschätzt wurde, verwendete eine dreidimensionale Zielfunktion, die sich aus den globalen Summen der Nahrungsmittelproduktion (kcal), Wasserbereitstellung (m^3) und Kohlenstoffspeicherung [t Kohlenstoff] zusammensetzt. Als Kontrollvariablen wurden verschiedene Landnutzungen verwendet. Potentielle Natürliche Vegetation (PNV) sowie 4 funktionale Furchtartengruppen (C3 Getreide, sonstige C3 Anbaufrüchte, C4 Anbaufrüchte, Reis) . Die Fruchtartengruppen wurden jeweils in der Variante Regenfeldanbau und Bewässerung implementiert. Da nicht davon ausgegangen werden kann, dass eine Umstellung auf vegane Ernährung eine reale Option ist, wurde Weideland als weitere Landnutzung mit aufgenommen, so dass in Summe je Entscheidungseinheit 10 Landnutzungsoptionen zur Verfügung stehen: PNV, Weideland und vier Anbaufruchtgruppen in je zwei Anbauvarianten.

Die Bereitstellung von Nahrungsmitteln, Wasser, Kohlenstoffspeicherung sowie Grünfutter wurde auf Basis einer 1x1° großen Zelldiskretisierung der Landoberfläche der Erde geschätzt. Für die Berechnung wurde das globale dynamische Vegetationsmodell LPJ-GUESS [Sm11] eingesetzt. LPJ-GUESS simuliert die Vegetationsentwicklung in Abhängigkeit von Klima, Bodeneigenschaften, athmosphärischer CO_2-Konzentration, Landnutzung [Li13] und Stickstoffdynamik [Ol15, Sm14]. LPJ-GUESS arbeitet hinsichtlich der Vegetationsdynamik als Punktmodell, d.h. modelliert eine räumliche Einheit, ohne Interaktionen mit benachbarten räumliche Einheiten zu berücksichtigen. Die Modellierung unterscheidet sich dabei für Flächen natürlicher Vegetation, Acker- und Weideland. Die Vegetationsdynamik auf natürlichen Flächen wird anhand von zwölf funktionalen Pflanzengruppen (*plant functional groups*) modelliert, die sich hinsichtlich ihrer bioklimatischen Präferenzen, ihreres Photosynthesesweges und Wachstumsstrategie unterscheiden [Sm14]. In Weideländern wurde die Vegetationsdynamik anhand konkurrierender funktionaler C3 und C4 Gräsergruppen abgebildet. Hierbei wurde eine Entnahme von 50 Prozent der Biomasse als Folge der Beweidung angenommen. Die Modellierung der Dynamik der funktionalen Fruchtartengruppen auf Ackerland berücksichtigt Eingriffe wie Aussaht und Ernte, Bewässerung, Düngung und Zwischensaaten [Ol15]. Aus der Vielzahl der im Modell repräsentierten Zustands- und Flussgrößen wurde der Ernteertrag, die Kohlenstoffsequestrierung und die Wasserverfügbarkeit im weiteren verwendet. Die Wasserverfügbarkeit einer räumlichen Einheit berücksichtigt dabei laterale Flüsse zwischen Zellen anhand des Fliessgewässernetzwerkes.

Das Modell wurde anhand historischer Klimadaten mit entsprechender Spin-off Periode gestartet und dann mit zwei Klimaszenarien angetrieben. Die Spin-off oder Einschwing-Phase dient dazu die Zustandsgrößen des Modells auf realistische Startwert zu initialisieren, bevor die eigentliche Simulation beginnt [Sm14]. Die Einschwingphase erfolgte mit der CO_2 Konzentration von 1861 und trend-bereinigten Klimadaten der Periode 1850-1879. Die Einschwingphase von 500 Jahren verwendete durchgehend die Potentielle Natürliche Vegetation als Landnutzung. An die Einschwingphase schloss sich für alle Szenarien die Simulation mit historischen Daten für den Zeitraum 1850-2017 an. Dieser folgten dann die Klimaszanarien für die beiden Emissionsszenarien. Landnutzungsinformationen für

die historische Entwicklung stammten aus HYDE 3.2.1 [Kl17]. Klimazeitreihen von vier Globalen Klimamodellen (IPSL-CM5A-LR, GFDL-ESM2M, MIROC5, HadGEM2-ES) estammten dem Inter-Sectoral Impact ModelIntercomparison Project (ISI–MIP), Phase 2b [Wa].

Aus der Optimierung herausgenommen wurden alle Flächen in Nationalparks oder ähnlicher Schutzstufe sowie alle Flächen mit niedrigem landwirtschaftlichen Ertrag oder solche, wo eine hohe Hangneigung die landwirtschaftliche Nutzung ausschließt. Die Optimierung fand damit auf insgesamt 9412 1x1° Zellen statt. Um die Varianz zwischen einzelnen Jahren zu berücksichtigen, wurden die Vorhersagen von LPJ-GUESS stets über mehrere Jahre gemittelt: für die Zeiträume 2033-2042 und 2090-2099. Es wurden weiterhin zwei unterschiedliche Repräsentative Konzentrationspfade (RCP) verwendet: RCP2.6 entspricht einem Szenario mit deutlichen Anstrengungen beim Klimaschutz, während RCP6.0 von einem deutlichen Anstieg der Treibhausgas-Emissionen mit Höhepunkt in 2080 ausgeht [IP20].

Für die Optimierung wurde der genetische Algorithmus NSGA-II [De02] eingesetzt. Die Optimierung erfolgte hierarchisch in drei Stufen: i) auf Ebene der 8 Biome, ii) auf Grundlage von 300 food producing units (FPU) [El14], die durch Verschneidung mit den Biomgrenzen 714 räumliche Einheiten ergeben, iii) auf Ebene der 9412 Zellen. Hierbei wurden jeweils alle zu einer Einheit gehörenden Zellen im Genom als eine Einheit abgelidet, Während die Auswertung der Zielfunktion auf der Zellebene erfolgte. Die Ergebnisse der übergeordneten hierarchischen Ebene wurden stets als Startwerte für die nachgelagerte Optimierungsstufe verwendet. Ab Stufe der FPUs wurden zu erreichende Grenzwerte (Constraints) auf Grundlage des heutigen Niveaus der Zielgrößen verwendet. Futterproduktion auf Grundlage des Weidelandes wurde stets nur in Form eines zu erreichenden Grenzwertes, aber nicht als Zielgröße in der Pareto-Optimierung verwendet. In der finalen Front finden sich stets nur Pareto-optimale Lösungen, die in allen drei Zieldimensionen und der Futterproduktion mindestens so gut wie die Ergebnisse der heutigen Landnutzung unter den jeweiligen Klimabedingungen sind. Die Berücksichtigung der Grenzwerte in der Auswertung der Zielfunktion wurden nach Deb (2002)[De02] in den Optimierungsalgorithmus implementiert. Der Vergleichsoperator berücksichtigt die Anzahl der Zieldimensionen, in denen ein Grenzwert verletzt wurde und ggf. die Stärke der Abweichung vom Grenzwert.

In der Optimierung kam auf der FPU-Ebene ein einfacher Single Point Crossover und eine Integer Mutation Operator zum Einsatz, während auf der Zellebenen-Optimierung die räumliche Topologie berücksichtigt wurde, was sich im Rahmen einer intensiven Parameterstudie in einem vergleichbaren Landnutzungsproblem als vorteilhaft erwiesen hat [Sc18a]. So erfolgte der Crossover entlang einer Geraden zwischen zwei zufällig gezogenen geographischen Koordinaten und die Mutation in Form von kreisförmiger Zonen mit 1-5 Zellen Radius. Die Konvergenz der Optimierung wurde anhand des Dominated Hypervolume [BNE07] beurteilt, welches den Teil des Lösungsraum angibt, der durch die Pareto-Front dominiert wird. Die Optimierung erfolgte auf der FPU-Ebene parallel in 5 unabhängigen Läufen, deren Pareto-Archive nach 100 Generationen nach einer Pareto

Filterung kombiniert wurden und als Startlösungen für die Zellebenen-Optimierung dienten. Auf der Zellebene wurden jeweils 5 Läufe parallel durchgeführt, die Pareto-Archive nach 100 Generationen kombiniert, durch einen Epsilon-Pareto Filter ausgedünnt und wieder in fünf parallelen Läufen für weitere 70 Generationen optimiert. Die Populationsgröße differierte zwischen FPU- und Zell-Ebene und zwischen den Zeithorizonten und RCPs. Auf Zellebene umfassten die Populationen zwischen 3000 und 11.000 Individuen, auf FPU-Ebene zwischen 1800 und 3500 Individuen.

Die Lösungen wurden nach Abschluss der Optimierung hinsichtlich Ihrer Landnutzungsmuster analysiert. So wurde der Prozentsatz aller Lösungen berechnet, in denen einer Zelle eine bestimmte Landnutzung zugeordnet wurde. Um die Veränderungen entlang der Front zu erfassen, wurden die Lösungen im nächsten Schritt anhand ihrer Lage in den drei Zieldimensionen in vier Gruppen eingeteilt (Bereiche um die maximal erreichten Zieldimensionen sowie Mittelgruppe als Kompromiss der 3 Zieldimensionen). Für jede Gruppe wurde wiederum der Prozentsatz aller Lösungen berechnet, in denen einer Zelle eine bestimmte Landnutzung zugeordnet wurde, sowie Differenzen zwischen den Prozentwerten der Kompromisslösungen und der Extremlösungen berechnet.

Tab. 1: Verbesserung der Zieldimensionen gegenüber der heutigen Landnutzung für die jeweiligen Klimabedingungen. Die Werte sind relativ zu diesem Referenzwert angegeben.

	Min.	1st Qu.	Median	Mittelwert	3. Qu.	Max
RCP2.6	**2033-2042**					
C02 Speicherung	1.000	1.005	1.010	1.010	1.015	1.023
Nahrungsmittel	1.000	1.326	1.613	1.626	1.920	2.369
Wasser	1.000	1.034	1.052	1.049	1.065	1.091
RCP6.0	**2033-2042**					
C02 Speicherung	1.000	1.006	1.011	1.011	1.016	1.024
Nahrungsmittel	1.000	1.321	1.587	1.608	1.889	2.361
Wasser	1.000	1.034	1.050	1.048	1.063	1.088
RCP2.6	**2090-2099**					
C02 Speicherung	1.000	1.013	1.027	1.029	1.042	1.073
Nahrungsmittel	1.000	1.354	1.755	1.834	2.259	3.106
Wasser	1.000	1.069	1.089	1.084	1.103	1.131
RCP6.0	**2090-2099**					
C02 Speicherung	1.000	1.013	1.026	1.029	1.042	1.076
Nahrungsmittel	1.000	1.354	1.717	1.792	2.162	3.058
Wasser	1.000	1.053	1.074	1.069	1.088	1.116

3 Ergebnisse

Die Ergebnisse (s. Tabelle 1) zeigen, dass für beide RCPs und Planungshorizonte Verbesserungen in allen drei Zieldimensionen möglich sind, ohne dass in einem anderen Ziel eine Verschlechterung gegenüber der Zielerfüllung unter der heutigen Landnutzungskonfiguration eintreten würde. Über alle vier Szenarien hinweg ist der Wertebereich, in dem die Optimierungspotentiale identifiziert wurden, für den Nahrungsmittelanbau über

Feldfrüchte deutlich größer als für Wasserspeicherung und Kohlenstoffspeicherung. Für den Feldfruchtanbau sind für den mittelfristigen Planungshorizont 2,3-fache Ergebnisse möglich, für den langfristigen Planungshorizont sogar Verdreifachungen. Es zeigt sich weiterhin, dass die Optimierungspotentiale gegenüber der heutigen Landnutzungskonfiguration für den längerfristigen Planungshorizont zunehmen.

Will man die Verbesserungen einigermaßen gleichmäßig auf alle drei Zieldimensionen verteilen, wird für RCP2.6 und den mittelfristigen Planungshorizont Potentielle Natürliche Vegetation in der dominierenden Anzahl der Lösungen in den tropischen Regenwäldern und in der borealen Nadelwaldzone verortet (s. Abb. 1). Aber auch weite Teile der USA werden in diesen Lösungen nicht in Landwirtschaftliche Nutzung genommen. Lösungen, die stark auf die Erhöhung der Kohlenstoffspeicherung fokussieren, nehmen vor allem Gebiete in China, Ost-und Mitteleuropa, aber auch Teilen Russlands und in den USA und Kanada aus der Nutzung. Lösungen, die dagegen die Nahrungsmittelproduktion im Fokus haben, nehmen Gebiete im Westen der USA, in Teilen Afrikas aber auch in Ost- und Mitteleuropa sowie zentralasiatischen Steppengebieten in Nutzung. In diesem Fall handelt es sich überwiegend um Bewässerungsfeldbau (s. Abb. 2). Wasserfokussierte Lösungen weisen für dieses Szenario vielfältige Zu- und Abnahmen von PNV gegenüber den Kompromisslösungen auf. Bewässerungsfeldanbau spielte abgesehen von Nahrungsmittel fokussierten Lösungen nur eine marginale Rolle. Regenfeldfruchtanbau stellt den wichtigsten Teil der Kompromisslösungen dar (s. Abb. 3).

Mit zunehmender Veränderung des Klimas bis Ende des Jahrhunderts gehen Veränderungen der Landnutzungsallozierung einher, die sich beispielhaft anhand der Gebiete zeigen lassen, die für Regenfeldanbau in den Kompromisslösungen bedeutend sind (s. Abb. 3 und 4). Für RCP2.6 wurden die Anbaugebiete deutlich ausgeweitet - da die höhere CO_2-Düngung und das wärmere Klima für Pflanzen in vielen Gebieten höhere Erträge bedeutet, besteht mehr Flexibilität, um Verbesserungen gegenüber der heutigen Landnutzungskonfiguration zu erzielen. Gebiete mit hohem Anteil Feldfruchtanbau in 2033-2042 bleiben überwiegend erhalten und werden durch benachbarte Gebiete ergänzt. Besonders dominant zeigte dies für den Süden der USA und Mexico aber auch in Südafrika und Europa.

Weidewirtschaft (s. Abb. 5) wurde vor allem Zellen in Grasländern und Savannen sowie in Mediterranen Klimazonen zugewiesen. Auffällig ist, dass quasi der gesamte indische Subkontinent für die Verortung von Weideland verwendet wurde. Diese Zuordnung bleibt auch für den langfristigen Zeithorizont und RCP6.0 bestehen (Ergebnisse nicht gezeigt).

4 Diskussion

Die Ergebnisse zeigen, dass es ein nicht unerhebliches Optimierungspotential hinsichtlich der räumlichen Verortung von Landnutzung gibt. Es scheint eine Vielzahl von Lösungsmöglichkeiten zu geben, um sowohl mehr (pflanzliche) Nahrungsmittel zu produzieren, als auch mehr Wasser zur Verfügung zu haben und mehr Kohlenstoff in der Vegetation und der Pedosphäre zu speichern.

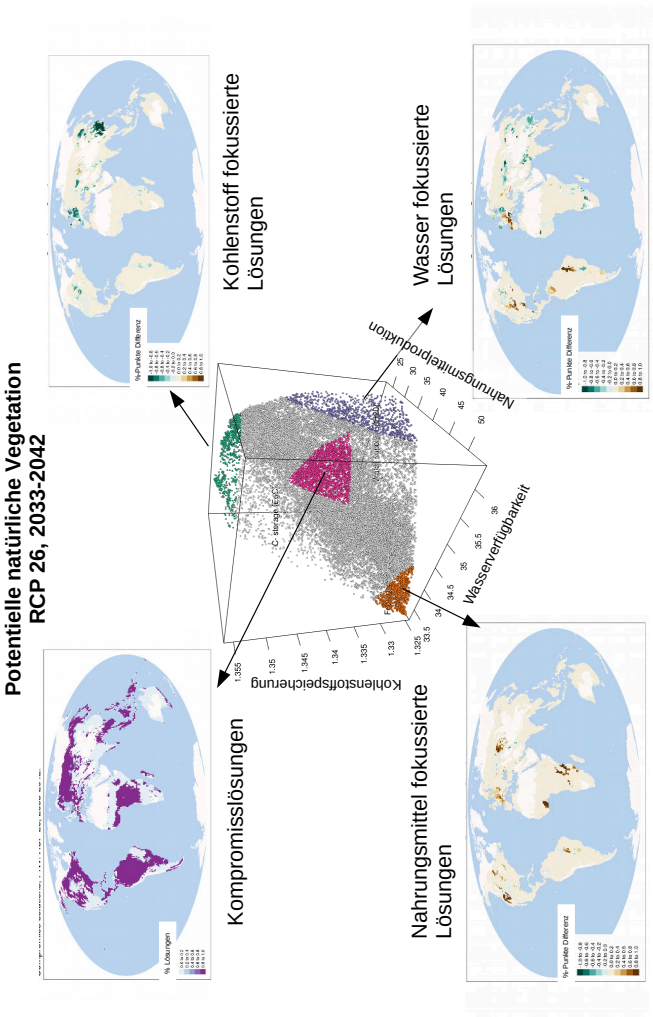


Abb. 1: Optimierungsergebnisse für RCP2.6, mittlerer Planungshorizont (2033-2042). Die 3D-Abbildung zeigt die geschätzte Lage der Pareto-Front. Die Werte sind optimal hinsichtlich der drei Zieldimensionen und besser als die heutige Landnutzung unter den gegebenen Klimabedingungen. Zudem wird die Futtermittelproduktion in Grasländern vergleichbar zum heutigen (2008-2017) Niveau erreicht. Die eingefärbten Bereiche kennzeichen a) den Bereich, in dem alle drei Ziele zu mindestens 40% des maximalen Wertes erfüllt (pink) sind sowie die Bereiche, in denen mindestens 95% eines Zieles erreicht sind (orangene, grüne und lilane Gruppen). Die der mittleren Gruppe zugeordnete Karte zeigt den Prozentsatz der Lösungen, bei denen PNV für die jeweilige Zelle gewählt wurde. Die anderen drei Karten zeigen Abweichungen der Extremgruppen gegenüber der mittleren Gruppe: negative Werte (grüner Farbton) korrespondieren mit Situationen, in denen die Zelle öfter als in der mittleren Gruppe gewählt wurde.

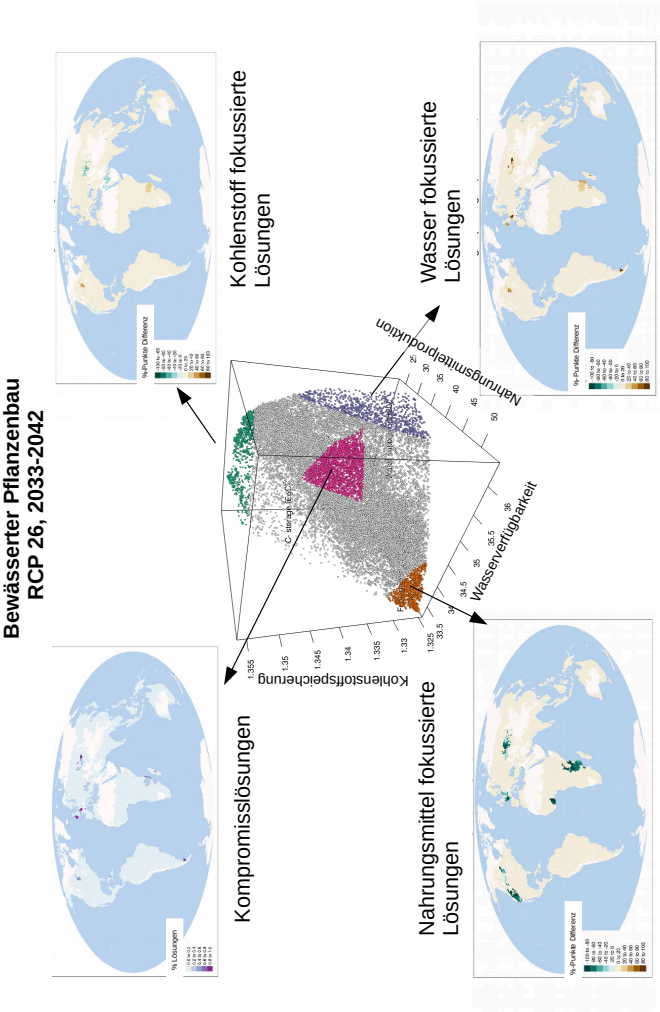


Abb. 2: Optimierungsergebnisse für RCP2.6, mittlerer Planungshorizont (2033-2042). Im Unterschied zu Abbildung 1 wird hier der Anteil der Lösungen dargestellt, in denen eine Zelle für Bewässerungsfeldbau alloziert wurde.

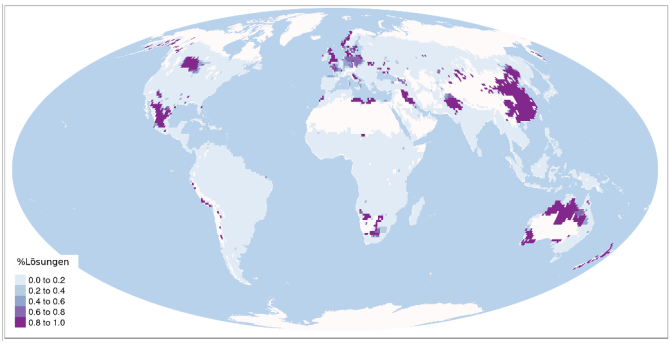


Abb. 3: Optimierungsergebnisse für RCP2.6, mittlerer Planungshorizont (2033-2042). Anteil der Kompromisslösungen, in denen einer Zelle Regenfeldfruchtanbau zugewiesen wurde.

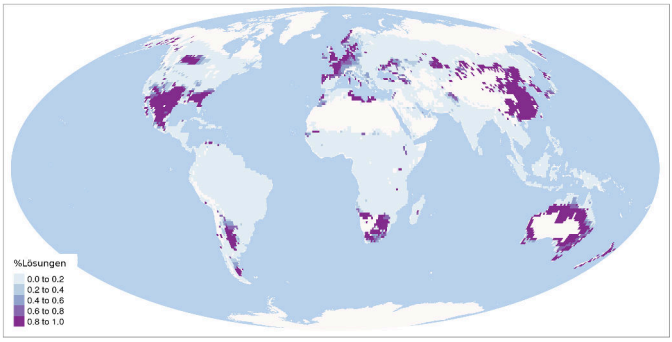


Abb. 4: Optimierungsergebnisse für RCP2.6, langfristiger Planungshorizont (2090-2099). Anteil der Kompromisslösungen, in denen einer Zelle Regenfeldfruchtanbau zugewiesen wurde.

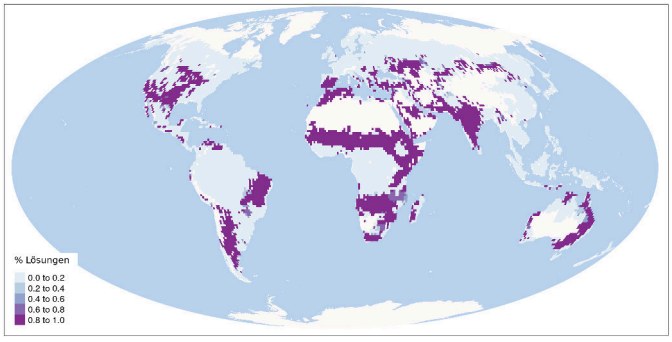


Abb. 5: Optimierungsergebnisse für RCP2.6, mittlerer Planungshorizont (2033-2042). Anteil der Kompromisslösungen, in denen einer Zelle Weidewirtschaft zugewiesen wurde.

Kritisch zu hinterfragen ist natürlich die politische Durchsetzbarkeit solcher Lösungen. Aus Sicht einer hypothetischen Weltregierung oder von internationale Organisation wie UNEP oder FAO sehen die Lösungen möglicherweise interessant aus, aus Sicht der betroffenen Staaten und Regionen sicherlich weniger. Im Sinne des oft geforderten out-of-the-box Denkens ist es dennoch wichtig, die Potentiale, die in der räumlichen Verteilung der Landnutzung liegen, aufzuzeigen.

Unsicherheiten der Modellierung bestehen einerseits in den nur unvollständig abgedeckten Rückkopplungen zwischen Landnutzungsveränderungen und Klima, andererseits in der Vielfalt an Managementoptionen im Pflanzenbau, die unterschiedliche Anbausorten, Bewässerungsvarianten und Düngung betreffen. Hinsichtlich der Annahmen der Nachfrageseite bestehen ebenfalls große Unsicherheiten. Eine Umstellung auf vegane Ernährung würde enorme Einsparungen bei der notwendigen Kalorienmenge bedeuten, was neue Optimierungsspielräume eröffnen würde. Eine Substitution von Fleischprodukten durch insektenbasierte Fleischersatzprodukte könnte ebenfalls - allerdings geringere - Einsparungen ermöglichen. Aktuelle Entwicklungen zeigen allerdings in die entgegengesetzte Richtung, so dass aufgrund des weltweit steigenden Konsums von Fleisch und Milchprodukten eher von einer deutlichen Steigerung des benötigten Kalorienverbrauchs auszugehen ist. Ebenso wären durch den Verzicht auf Bioenergiekraftstoffe eine deutliche Reduzierung der weltweit benötigen Pflanzenproduktion möglich. Nicht berücksichtigt wurden durch die Veränderung der Landnutzungsallokation notwendige Transportvorgänge und deren Auswirkungen auf die Treibhausgasemissionen. Ergebnisse aus einer life-cylce-assessment Studie unter Einbeziehung einer, hinsichtlich der Minimierung von Treibhausgasemissionen, optimierten Landnutzungsallokation lassen jedoch vermuten, dass die Emissionen durch den Transport für ungekühlte Lebensmittel, gegenüber den anderen genannten Faktoren, eher nachrangig sein dürften [KLK16]. Letzendlich findet die Optimierung stets unter den Randbedingungen von Szenario-Annahmen statt. Für die dargestellte Beispielanwendung wurde von relativ konstanten Randbedingunen ausgegangen, um die Interpretation zu vereinfachen. Ziel war es, die Veränderungen der Zielgrößen zwischen Optimierungslösungen und dem Referenzzustand auf Veränderungen der Landnutzungsallokation zurückzuführen und die Zuordnung von Wirkung und Ursache nicht durch zusätzliche Veränderungen zu verkomplizieren.

Die Analyse ist aus unserer Sicht dennoch auf der betrachteten globalen Skala und angesichts fehlender vergleichbarer Arbeiten adäquat, um die Größenordnung des Optimierungspotentials und mögliche räumliche Muster aufzuzeigen. Die Ergebnisse bieten deutliche Hinweise auf nicht unerhebliche Optimierungspotentiale durch eine modifizierte Landnutzungsallokation und bieten damit eine zusätzliche Perspektive bei der Suche nach Möglichkeiten, den Klimawandel abzumildern.

Zukünftige Arbeiten sollten sich mit der Identifikation robuster Lösungen beschäftigen, die auch hinsichtlich unterschiedlicher Zeithorizonte und Emissionsszenarien gute Lösungen darstellen. Eine weitere interessante Analysestellt die Frage dar, wie stark die Optimierungspotentiale zurückgehen, falls nicht auf globaler Ebene sondern getrennt für einzelne politische Einheiten (USA, EU, China, Indien,...) Anbaumuster optimiert werden.

Literaturverzeichnis

[BNE07] Beume, Nicola; Naujoks, Boris; Emmerich, Michael: SMS-EMOA: Multiobjective selection based on dominated hypervolume. European Journal of Operational Research, 181(3):1653–1669, September 2007.

[BPG09] Bennett, E M; Peterson, Garry D; Gordon, Line J: Understanding relationships among multiple ecosystem services. Ecology letters, 12(12):1394–1404, Dezember 2009.

[Br15] Bryan, Brett A.; Crossman, Neville D.; Nolan, Martin; Li, Jing; Navarro, Javier; Connor, Jeffery D.: Land use efficiency: Anticipating future demand for land-sector greenhouse gas emissions abatement and managing trade-offs with agriculture, water, and biodiversity. Global Change Biology, 21(11):4098–4114, 2015. ISBN: 13541013 (ISSN).

[De02] Deb, K.; Pratap, A.; Agarwal, S.; Meyarivan, T.: A fast and elitist Multi-Objective Genetic Algorithm: NSGA-II. IEEE Transactions on Evolutionary Computation, 6(2):182–197, 2002. Publisher: Kanpur Genetic Algorithms Laboratory.

[El14] Elliott, Joshua; Deryng, Delphine; Müller, Christoph; Frieler, Katja; Konzmann, Markus; Gerten, Dieter; Glotter, Michael; Flörke, Martina; Wada, Yoshihide; Best, Neil; Eisner, Stephanie; Fekete, Balázs M; Folberth, Christian; Foster, Ian; Gosling, Simon N; Haddeland, Ingjerd; Khabarov, Nikolay; Ludwig, Fulco; Masaki, Yoshimitsu; Olin, Stefan; Rosenzweig, Cynthia; Ruane, Alex C; Satoh, Yusuke; Schmid, Erwin; Stacke, Tobias; Tang, Qiuhong; Wisser, Dominik: Constraints and potentials of future irrigation water availability on agricultural production under climate change. Proceedings of the National Academy of Sciences of the United States of America, 111(9):3239–44, 2014. ISBN: 0027-8424, 1091-6490.

[Fo05] Foley, Jonathan A; Defries, Ruth; Asner, Gregory P; Barford, Carol; Bonan, Gordon; Carpenter, Stephen R; Chapin, F Stuart; Coe, Michael T; Daily, Gretchen C; Gibbs, Holly K; Helkowski, Joseph H; Holloway, Tracey; Howard, Erica A; Kucharik, Christopher J; Monfreda, Chad; Patz, Jonathan A; Prentice, I Colin; Ramankutty, Navin; Snyder, Peter K: Global consequences of land use. Science, 309:570–574, Juli 2005.

[Fo11] Foley, Jonathan a.; Ramankutty, Navin; Brauman, Kate a.; Cassidy, Emily S.; Gerber, James S.; Johnston, Matt; Mueller, Nathaniel D.; O'Connell, Christine; Ray, Deepak K.; West, Paul C.; Balzer, Christian; Bennett, Elena M.; Carpenter, Stephen R.; Hill, Jason; Monfreda, Chad; Polasky, Stephen; Rockström, Johan; Sheehan, John; Siebert, Stefan; Tilman, David; Zaks, David P. M.: Solutions for a cultivated planet. Nature, 478:337–342, Oktober 2011.

[Gr14] Grêt-Regamey, Adrienne; Rabe, Sven-Erik; Crespo, Ricardo; Lautenbach, Sven; Ryffel, Andrea; Schlup, Barbara: On the importance of non-linear relationships between landscape patterns and the sustainable provision of ecosystem services. Landscape Ecology, 29(2):201–212, November 2014.

[IP20] IPCC: , Representative concentration pathways. `https://sedac.ciesin.columbia.edu/ddc/ar5_scenario_process/RCPs.html`, 2020. Accessed: 2020-05-09.

[Kl17] Klein Goldewijk, Kees; Beusen, Arthur; Doelman, Jonathan; Stehfest, Elke: New anthropogenic land use estimates for the Holocene: HYDE 3.2. Earth System Science Data, 9(2):927–953, 2017.

[KLK16] Kreidenweis, U.; Lautenbach, S.; Koellner, T.: Regional or global? The question of low-emission food sourcing addressed with spatial optimization modelling. Environmental Modelling and Software, 82:128–141, 2016.

[Kr17] Krause, Andreas; Pugh, Thomas A. M.; Bayer, Anita D.; Doelman, Jonathan C.; Humpenöder, Florian; Anthoni, Peter; Olin, Stefan; Bodirsky, Benjamin L.; Popp, Alexander; Stehfest, Elke; Arneth, Almut: Global consequences of afforestation and bioenergy cultivation on ecosystem service indicators. Biogeosciences, 14(21):4829–4850, November 2017.

[La13] Lautenbach, Sven; Volk, Martin; Strauch, Michael; Whittaker, Gerald; Seppelt, Ralf: Optimization-based trade-off analysis of biodiesel crop production for managing an agricultural catchment. Environmental Modelling & Software, 48:98–112, Oktober 2013.

[La14] Lautenbach, Sven; Volk, Martin; Strauch, Michael; Whittaker, Gerald; Seppelt, Ralf: Management support by optimization-based trade-off analysis – the example of biodiesel crop production. Geo-Öko, 1-2:39–77, 2014.

[La17] Lautenbach, Sven; Jungandreas, Anne; Blanke, Jan; Lehsten, Veiko; Mühlner, Susanne; Kühn, Ingolf; Volk, Martin: Trade-offs between plant species richness and carbon storage in the context of afforestation – Examples from afforestation scenarios in the Mulde Basin , Germany. Ecological Indicators, 73:139–155, 2017. Publisher: Elsevier Ltd.

[Li13] Lindeskog, M.; Arneth, A; Bondeau, A; Waha, K.; Seaquist, J.; Olin, S.; Smith, B.: Implications of accounting for land use in simulations of ecosystem carbon cycling in Africa. Earth System Dynamics, 4(2):385–407, 2013. ISBN: 2190-4979.

[LL16] Lee, Heera; Lautenbach, Sven: A quantitative review of relationships between ecosystem services. Ecological Indicators, 66:340–351, 2016. Publisher: Elsevier Ltd ISBN: 1470-160X.

[Ol15] Olin, S.; Schurgers, G.; Lindeskog, M.; Wårlind, D.; Smith, B.; Bodin, P.; Holmér, J.; Arneth, A.: The impact of atmospheric CO $_2$ and N management on simulated yields and tissue C : N in the main wheat regions of Western Europe. Biogeosciences Discussions, 12(2):1047–1111, Januar 2015.

[Po17] Popp, Alexander; Calvin, Katherine; Fujimori, Shinichiro; Havlik, Petr; Humpenöder, Florian; Stehfest, Elke; Bodirsky, Benjamin Leon; Dietrich, Jan Philipp; Doelmann, Jonathan C.; Gusti, Mykola; Hasegawa, Tomoko; Kyle, Page; Obersteiner, Michael; Tabeau, Andrzej; Takahashi, Kiyoshi; Valin, Hugo; Waldhoff, Stephanie; Weindl, Isabelle; Wise, Marshall; Kriegler, Elmar; Lotze-Campen, Hermann; Fricko, Oliver; Riahi, Keywan; Vuuren, Detlef P. van: Land-use futures in the shared socio-economic pathways. Global Environmental Change, 42:331–345, Januar 2017.

[Sc18a] Schwaab, J.; Deb, K.; Goodman, E.; Lautenbach, S.; van Strien, M.J.; Grêt-Regamey, A.: Improving the performance of genetic algorithms for land-use allocation problems. International Journal of Geographical Information Science, 32(5):907–930, 2018.

[Sc18b] Schwaab, Jonas; Deb, Kalyanmoy; Goodman, Erik; Kool, Sander; Lautenbach, Sven; Ry, Andrea; Strien, Maarten J Van; Grêt-regamey, Adrienne: Using multi-objective optimization to secure fertile soils across municipalities. Applied Geography, 97(June):75–84, 2018.

[SLV13] Seppelt, Ralf; Lautenbach, Sven; Volk, Martin: Identifying trade-offs between ecosystem services, land use, and biodiversity: a plea for combining scenario analysis and optimization on different spatial scales. Current Opinion in Environmental Sustainability, 5:1–6, Juni 2013. Publisher: Elsevier B.V.

[Sm11] Smith, Benjamin; Samuelsson, Patrick; Wramneby, Anna; Rummukainen, Markku: A model of the coupled dynamics of climate, vegetation and terrestrial ecosystem biogeochemistry for regional applications. Tellus, Series A: Dynamic Meteorology and Oceanography, 63(1):87–106, 2011. ISBN: 0280-6495.

[Sm14] Smith, B.; Wårlind, D.; Arneth, A.; Hickler, T.; Leadley, P.; Siltberg, J.; Zaehle, S.: Implications of incorporating N cycling and N limitations on primary production in an individual-based dynamic vegetation model. Biogeosciences, 11(7):2027–2054, April 2014.

[St14] Stehfest, E; an Vuuren, D; Kram, T; Bouwman, L; Alkemade, R; Bakkenes, M; Biemans, H; Bouwman, A; den Elzen, M; Janse, J; Lucas, P; van Minnen, J; Müller, C; Prins, A: IntegratedAssessment of Global Environmental Change with IMAGE 3.0:Model description and policy applications. Bericht, The Hague, PBLNetherlands Environmental Assessment Agency, The Hague, theNetherlands, 2014.

[vAV12] van Asselen, Sanneke; Verburg, Peter H.: A Land System representation for global assessments and land-use modeling. Global Change Biology, 18(10):3125–3148, Oktober 2012.

[Ve18] Verhagen, Willem; van der Zanden, Emma H.; Strauch, Michael; van Teeffelen, Astrid J.A.; Verburg, Peter H.: Optimizing the allocation of agri-environment measures to navigate the trade-offs between ecosystem services, biodiversity and agricultural production. Environmental Science & Policy, 84:186–196, 2018.

[Wa] Warszawski, Lila; Frieler, Katja; Huber, Veronika; Piontek, Franziska; Serdeczny, Olivia; Schewe, Jacob: The Inter-Sectoral Impact Model Intercomparison Project (ISI–MIP): Project framework. 111(9):3228–3232.

1st Workshop on Evaluating Intelligent and Ubiquitous Mobility Systems

R. Reussner, A. Koziolek, R. Heinrich (Hrsg.): INFORMATIK 2020,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2021 1111

1st Workshop on Evaluating Intelligent and Ubiquitous Mobility Systems – EvalIUMS

Waldemar Titov[1] Mathias Trefzger[1] Thomas Schlegel[1]

Abstract: Evaluation of peoples' mobility and intermodal transport is crucial for understanding mobility and traffic in general and leads to new challenges from the perspective of computer science and especially Human-Computer Interaction.

Keywords: Evaluation; Mobility Experience; Mobility Infrastructure

Ubiquitous Mobility Systems: Complex and Highly Interactive Systems of Systems in Mobility

Huge amounts and varieties of travelers move in different travel chains all over the globe. The interplay of such different systems, like car and bike sharing, local and long-distance public transport and individual transport, must be adapted to the needs of the travelers and their situations, which also lead to different information needs, different interaction possibilities and different behavior.

Intelligent traveler information systems must be created and developed in a way that makes it easier for travelers to plan, book, execute and adapt an intermodal travel chain and to interact with the different systems, i.e. systems of systems. Innovative means of transport are developed, such as electric vehicles and autonomous vehicles as well as the sharing economy and intermodal apps. To achieve the acceptance of these systems, human-machine interaction must be fundamentally redesigned to work in mobility.

Focus on Mobility Modes and Ubiquitous Systems

The first Workshop on Evaluating Intelligent and Ubiquitous Mobility Systems (EvalIUMS) accumulates the different approaches for analyzing and evaluating aspects of the four different mobility modes: walking, cycling, driving and using public transport. The objectives of the Workshop were to exchange findings and lessons learned when applying analysis methods and tools for evaluation of mobility systems as well as the participation of users and represent an opportunity to connect these approaches.

[1] Institute of Ubiquitous Mobility Systems (http://iums.eu), Karlsruhe University of Applied Sciences, Moltkestrasse. 30, 76133 Karlsruhe, Germany, iums@hs-karlsruhe.de

doi:10.18420/inf2020_104

EvalIUMS aimed at giving an insight into modern evaluation techniques of IT-based mobility systems with a focus on intelligent and ubiquitous systems that can improve planning and the evaluation of implemented measures. The overall goal of the evaluation methods is to increase mobility experience and quality in dimensions such as user comfort, usability, safety and the overall user acceptance in urban mobility.

A New Field of Research for Equipping Mobility Researchers and Professionals Evaluation Tools

The concept of EvalIUMS considered the target groups of young researchers in the field of mobility and evaluation, urban and city planners, planners of public transport as well as developers of intelligent transport systems, including Intermodal Transport Control Systems (ITCS), sharing and mobility apps.

Therefore, EvalIUMS was designed for exchanging experiences and collectively identifying points for future joint work as well as starting the definition of a research agenda. To realize the proposed concept, even in the difficult conditions of a virtually held workshop, the workshop was split into an impulse session, the paper presentation session and the extensive discussion session. To ensure the quality and novelty of the submissions the program committee reviewed the applicants' papers and, if necessary, requested required editing and improvements. The members of the program committee were researchers, practitioners from the economy as well as interested persons active in relevant areas of expertise of evaluating mobility systems.

Focus on Evaluating Systems in Non-motorized Transport

EvalIUMS was successful in uncovering current challenges in the evaluation and assessment of modern mobility systems. The workshop focused on the evaluation of the infrastructure of non-motorized transport users such as pedestrians, cyclists as well as public transport users.

In the four position papers presented, different approaches of participation and evaluation were described and discussed. The summaries of which can be found below. A discussion of upcoming challenges and future cooperation completed the workshop.

Sensor Data Collection of Comfort-related Influences on Bicycle Traffic

In this work an objective data basis for the comfort of bicycle paths is created. Referring on conducted studies, precise conclusions about the comfort on partial routes are drawn based on this data and made available to the user on an individualized basis. In addition, made

available to the user on an individualized basis. For this purpose, the factors that influence the comfort of bicycle routes, both positively and negatively, will first be considered. Subsequently, it will be determined how these factors can be can be automatically recorded with the help of a sensor apparatus. Design and programming of the sensor module, as well as the consistency of the collected data will be evaluated based on a field test, in order to determine their suitability for the creation of such a "bicycle information system".

Analysis of Cycling Traffic with a SensorBike with Ubiquitous Sensors

An approach to survey cycling traffic using ubiquitous sensors in the form of a SensorBike is presented. The aim is to record the cycling traffic from the perspective of the cyclists in order to gain new approaches for the promotion of cycling. Based on first experiences from surveys with the SensorBike, their use cases as well as the challenges for data collection and evaluation are presented. The possibilities but also the limits and challenges of the survey with a SensorBike will be discussed.

Active, Mobile Travel Companionship in Public Transport – Connection Establishment of Mobile Devices and Mobile Public Displays

This paper will discuss the advantages of a continuous, stable connection between a public display in a public space and a private smartphone. It will focus on active travel guidance and the flow of communication from start to finish. It explains what happens to the data during a trip and the role of the platform for route calculation. In this paper concepts for establishing connections between private mobile devices and mobile public displays, currently investigated in the research project SmartMMI[2]. Based on requirements determined by personas and scenarios, a first prototype in the form of a smartphone application will be developed to simulate the connection setup. This will allow initial statements to be made about the demands that need to be placed on such a system.

Analysis and Comparison of the Gaze Behavior of E-Scooter Drivers and Cyclists – Depending on Road Surface Quality in a Real Test Environment

In this paper, an eye tracking study to evaluate the gaze behavior of e-scooter drivers and cyclists on high and low quality road surfaces were contributed. Therefore, the surface quality were recorded with sensors and the different surfaces put in relation to the gaze behavior. The eye movements of the participants were analyzed to identify gaze patterns. A significant difference in the attention distribution of the two investigate means of transport.

[2] Research project Model-and Context-based Mobility Information on Smart Public Displays and Mobile Devices in Public Transport: https://smartmmi.de/project/

Program Committee:

Prof. Dr.-Ing. Thomas Schlegel, Hochschule Karlsruhe – Technik und Wirtschaft

Prof. Dr. Michael Koch, Universität der Bundeswehr München

Dipl.-Ing. André Dettmann, Technische Universität Chemnitz

Jan Schwarzer, Hochschule für Angewandte Wissenschaften Hamburg

Dr.-Ing. Matthias Pfriem, Profilregion Mobilitätssysteme Karlsruhe

Dirk Weißer, Head of Research @INIT Group

Dr. Anja Exler, Karlsruhe Institute of Technology

Dr. Wolf Engelbach, Ministry for Mobility Baden Württemberg

Dr. Karl-Heinz Krempels, Fraunhofer FIT/RWTH Aachen

R. Reussner, A. Koziolek, R. Heinrich (Hrsg.): INFORMATIK 2020,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2021 1115

Sensordatenerhebung Komfortbezogener Einflüsse auf den Radverkehr

Lars Badde[1], Waldemar Titov[2], Thomas Schlegel[2]

Abstract: Ziel dieser wissenschaftlichen Arbeit ist es eine objektive Datengrundlage zum Komfort von Fahrradwegen zu schaffen. In nachstehenden Untersuchungen sollen anhand dieser Daten präzise Rückschlüsse zum Komfort auf Teilstrecken gezogen werden können und dem Nutzer individualisiert verfügbar gemacht werden. Zu diesem Zwecke wird zunächst betrachtet welche Faktoren den Komfort von Fahrradstrecken, sowohl positiv als auch negativ, beeinflussen. Anschließend soll ermittelt werden, wie diese Faktoren mit Hilfe einer Sensorapparatur automatisiert erfasst werden können. Aufbau und Programmierung des Sensormoduls, sowie die Konsistenz der erhobenen Daten werden auf Basis eines Feldtestes evaluiert, um deren Tauglichkeit für die Schaffung eines solchen „Fahrrad-Informations-Systems“ zu gewährleisten.

Keywords: Sensorik; Fahrrad; Komfort; Datenerhebung; Tinkerforge

1 Einführung

Im Rahmen mehrerer zusammengehöriger Untersuchungen sollen Wege gefunden werden, Daten bezüglich des Komforts von Fahrradteilstrecken zu erheben, auszuwerten und darzustellen. Ohne Ortskenntnis ist es für einen Radfahrer schwierig einen komfortablen Fahrradweg zu erkennen. Mit Hilfe von Tools wie GoogleMaps lassen sich oft nur die schnellsten Wege finden. Ob diese Wege auch den erwarteten Komfort für einen Fahrradfahrer bieten ist jedoch nicht zu erkennen. Will der Radfahrer nicht auf engen Straßenräumen mit hohem Kfz-Verkehr fahren, oder signalisierte Kreuzungen meiden, so werden spezielle Informationen zu den jeweiligen Teilstrecken benötigt.

Die bestehenden Informationsmöglichkeiten beschränken sich bisher auf spezielle Radfahrkarten, –Bücher oder besondere Apps. Während Printmedien prinzipiell nur subjektive Fremdmeinungen wiedergeben, die oft nicht aktuell sind und nicht die Möglichkeit bieten individuelle Vorlieben zu berücksichtigen, sind Systeme die auf Massendatenerhebungen basieren, bis auf GoogleMaps, nicht vorhanden.

Ziel dieser Untersuchung ist es Faktoren auszumachen, die den Komfort für Fahrradfahrer beeinträchtigen. Mit der Entwicklung eines Sensormodules sollen dann automatisiert Daten

[1] Karlsruhe University of Applied Sciences, Moltkestraße 30, 76133 Karlsruhe, Germany, larsbadde@web.de | https://www.hs-karlsruhe.de/
[2] Institute of Ubiquitous Mobility Systems (IUMS), Karlsruhe University of Applied Sciences, Moltkestraße 30, 76133 Karlsruhe, Germany, iums@hs-karlsruhe.de | http://iums.eu

doi:10.18420/inf2020_105

erfasst werden, die diese Faktoren bemessen und lokalisieren. Auf diese Weise können schnell und effizient massenweise Daten gesammelt werden. In weiteren, nachstehenden Untersuchungen sollen auf Basis dieser Daten Rückschlüsse über den Komfort von Fahrradteilstrecken getroffen werden können und diese mit Hilfe von Mapping-Systemen für den Nutzer verfügbar zu machen.

Das primäre Ziel ist die konsistente, automatisierte Erfassung von Sensordaten, im Hinblick auf deren Aussage zum Komfort von Fahrradteilstrecken. Zu diesem Zweck muss zunächst herausgestellt werden, welche Faktoren einen Fahrradfahrer behindern, stören oder eventuell anreizen können. Anschließend soll untersucht werden in wie fern diese Faktoren durch Sensoren bemessen werden können. Daraufhin kann dann das Sensormodul selbst entwickelt, getestet und evaluiert werden. Die Daten werden mit Hinblick auf Ihre Konsistenz und Aussagekraft betrachtet, um die Tauglichkeit des Sensormoduls bewerten zu können.

2 Themenbezogene Arbeiten

Der Bundesverkehrswegeplan 2030 [IS16] setzt als eines seiner Hauptziele die Verlagerung des Verkehres auf umweltverträglichere Verkehrsträger an. Als „übergeordnete Ziele“ werden unter anderem die Erhöhung der Verkehrssicherheit, sowie die Verbesserung der Lebensqualität einschließlich der Lärmsituation in Regionen und Städten genannt. Im Sinne dessen ist die Förderung des Radverkehrs von gesonderter Bedeutung. Um den Radverkehr zu fördern können sich Länder, Städte und Kommunen zahlreicher Maßnahmen bedienen. Grundsätzlich sind diese Maßnahmen zur Verkehrsträgerverlagerung unterschieden nach „harten“ und „weichen“ Maßnahmen [Sc09]. Harte Maßnahmen stellen Infrastrukturmaßnahmen, wie beispielsweise Straßenaus- oder Neubau dar. Diese sind folglich oft mit hohen Investitionskosten verbunden. Weiche Maßnahmen bestehen aus Information, Kommunikation, Motivation, Koordination und Service [SE14]. Diese weichen Maßnahmen gewinnen im “Informations-Zeitalter“ immer mehr an Bedeutung [Gö11]. Insofern kann ein merkbarer Effekt des forcierten Fahrrad-Informations-Systems auf den Radverkehrsanteil erwartet werden. Des Weiteren können die erhobenen Daten auch als Datengrundlage zu weiterführenden verkehrsplanerischen Analysen genutzt werden. [TS19] verfolgt einen crowdsensing Ansatz. Dabei erfolgt die Datenerfassung mittels einer App auf dem Smartphone welches für die Aufzeichnung starr am Fahrradlenker befestigt wird. Während der Aufzeichnung werden Werte der vier Sensoren (Standort, Rotationsgeschwindigkeit Beschleunigung und lineare Beschleunigung) aufgezeichnet und zur Klassifikation auf einen Server hochgeladen. Mit Hilfe eines maschinell lernenden Algorithmus erfolgt die beschriebene Klassifizierung der Daten in drei Qualitätsstufen.

2.1 Komfortfaktoren im Radverkehr

Um den Komfort eines Fahrradstreckenabschnittes bewerten zu können, muss zunächst ermittelt werden, welche Faktoren diesen beeinflussen und ob, beziehungsweise auf welche Weise, diese durch Sensorik erfasst werden können.

Bisherige Studien haben sich im Zusammenhang von Komfort und Radverkehr vor allem mit Wetterbedingungen und Jahreszeiten befasst [BFG84]. Diese eher zeitlichen Einflussfaktoren sollen nicht im Fokus dieser Untersuchung stehen. Es sollen stattdessen räumlich differenzierte Einflussfaktoren betrachtet werden. Eine Studie aus dem Jahre 2008 bezifferte den Einfluss „lokaler Faktoren“ auf den Radverkehrsanteil auf etwa 70% [TJT08].

Laut dem Allgemeinen Deutschen Fahrrad-Club e.V., kurz ADFC haben Studien gezeigt, dass der Großteil der Menschen, häufiger das Fahrrad benutzen würde, wenn sie den Radverkehr positiver erleben würden [DV14]. Wenn also der Straßenraum und Streckenführung mehr auf den Fahrradfahrer angepasst sind. Die ERA (Empfehlungen für Radverkehrsanlangen) empfiehlt, neben der Ermöglichung von schnellen und direkten Wegen, vor allem die Berücksichtigung einzelner, spezieller Nutzergruppen [FG10]. Dazu gibt die ERA Empfehlungen wie der Straßenraum für spezielle Nutzergruppen zu gestalten ist. Die dort beschriebenen Faktoren können aber nur spärlich auf die allgemeinen Einflussfaktoren einer komfortablen Fahrradstrecke projiziert werden. Prinzipiell empfiehlt die ERA allerdings eine getrennte Führung des Radverkehrs vom Kraftfahrzeugverkehr, soweit möglich. Des Weiteren, zur Minimierung des Kraftaufwandes, Oberflächen mit geringem Rollwiederstand, die Minimierung vermeidbarer Störungen, unnötiger Halte und Zeitverluste. Zudem definiert Sie für nahräumige Radverkehrsanbindungen eine angestrebte Fahrgeschwindigkeit von 20 bis 30 km/h und eine daraus abgeleitete maximale Dauer an Zeitverlusten, bedingt durch Anhalten und Warten, von 35 Sekunden je Kilometer.

Eine Studie an der Universität Maryland bestätigte die Annahme, dass sich Pendler durch Wege abseits des Kfz-Verkehrs ermutigen lassen das Rad zu nutzen [AC09]. Diese Studie führte eine Online-Umfrage mit Studenten der University of Maryland durch, die in einem Radius von maximal fünf Meilen zum Campus wohnen. Die größten Motivationen für Radbesitzer mit dem Fahrrad zu fahren, waren demnach ein eigener Fahrradstreifen, sowie die separate Führung zum Kfz-Verkehr. Aber auch bessere Beleuchtung und eine gute Karte mit lokalen Fahrradwegen motivierten die Probanden das Rad zu nutzen. Auf die Frage, was die Studenten davon abhalte das Fahrrad zu nutzen, antwortete die Mehrheit mit dem fehlenden Gefühl von Sicherheit im Straßenverkehr und dem schlechten Zustand der Straßenräume.

Inwiefern sich Faktoren wie die Luftschadstoffbelastung oder Wartezeiten an Kreuzungen auf das Fahrradverhalten auswirken ist bisher nicht empirisch untersucht worden.

2.2 Sensordatenerhebungen am Fahrrad

In einer vergleichbaren Studie wurde in der chinesischen Millionenstadt Changzhou ein Fahrradsensor entwickelt und getestet, der die Luftverschmutzung an Straßenräumen messen sollte [XBA15]. Verwendet wurden ein Aerosolgassensor, sowie ein Sensor für CO_2, N_2, NO_x, CO. Die Sensoren, sowie ein GPS-Empfänger wurden mit einem Mikroprozessor verbunden. Die Untersuchung zeigte, dass an Straßen mit großen Verkehrsmengen deutlich erhöhte Mengen Feinstaub gefunden wurden. Die Feinstaubwerte stellten sich als besonders guter Indikator für verkehrsbedingte Luftverschmutzung an Straßenräumen heraus.

Bei der Erkennung der Oberflächenbeschaffenheit und –Qualität, sowie zur Erfassung von Hindernissen während des Fahrradfahrens mit Hilfe von einem Ultraschallsensor [TNH15] und Smartphone-Beschleunigungs-Sensorik [HMM13] konnten eingeschränkt Erfolge verzeichnet werden (Abbildung 1) Die Systeme konnten größere Unebenheiten erkennen, erforderten jedoch eine umständliche, komplett starre Anbringung am Fahrrad.

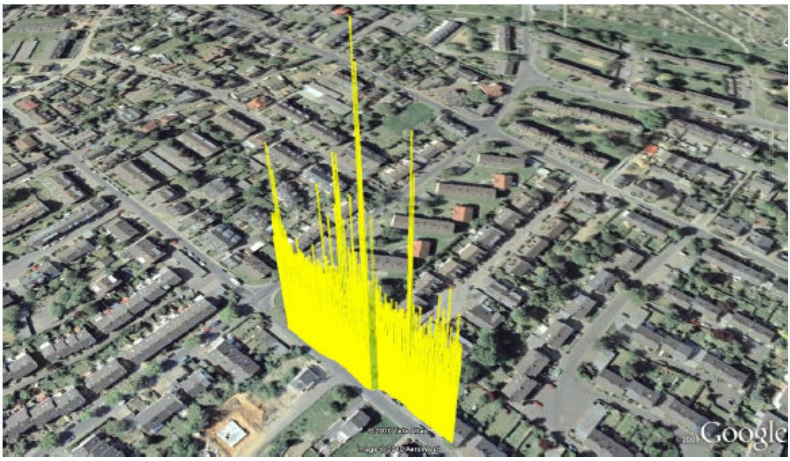


Abb. 1: „Bumps on the road“, [HMM13]

Eine britische Studie beschäftigte sich mit der Tauglichkeit von MEMS-Sensorik an Fahrrädern [MKL15]. Ziel war eine Art Inertialortung mit Hilfe der Beschleunigungs- und Orientierungsdaten zu schaffen. Sie fand heraus, dass weder Beschleunigungssensorik, noch Gyroskope alleine brauchbare Ergebnisse liefern. Eigenständig unterliegen beide Messverfahren hohen Fehlereinflüssen. Durch die Kombination der Daten beider Systeme unter Zuhilfenahme der Kalman-Filter können akkurate Resultate erzielt werden, bei denen Fehlereinflüsse weitestgehend eliminiert wurden.

Auch im Bereich der Höhenmessung im Fahrradverkehr wurde bereits eine Untersuchung durchgeführt. Diese hat mit Hilfe der Sensorik eines Smartphones GPS- und Beschleunigungsdaten erhoben und verknüpft [De15]. Als Ergebnis ließen sich ein Höhenverlauf,

sowie ein Verlauf der einzelnen Achsbeschleunigungen darstellen. Ziel der Untersuchung war es, Radfahrern bessere Informationen über mögliche Fahrradstrecken geben zu können.

Untersuchung mit dem Ziel, die Gesamtheit an Komforteinflüssen für den Fahrradverkehr, anhand eines Sensormodules zu erfassen und auszuwerten sind bisher nicht vorhanden. Erst die Gesamtheit an Einflüssen erlaubt es jedoch Wechselwirkungen und Fehlereinflüsse der Einzelfaktoren zu berücksichtigen.

2.3 Verwandte Datensammlungen

Ähnliche Fahrrad-Informations-Systeme sind in dem Umfang bisher noch nicht vorhanden. Das zum OpenSource-Projekt OpenStreetMap gehörige OpenCylceMap soll Informationen geben, welche Straßen als Fahrradstrecken besonders tauglich sind [GL16]. OpenCycleMap greift auf dieselbe Datengrundlage zu wie OpenStreetMap, stellt diese jedoch angepasst für Fahrradfahrer dar. Dabei beschränkt sich der dargestellte Inhalt jedoch auf Fahrradwege, Fahrradläden, und Unterstände.

Etwas mehr Anpassungsmöglichkeiten bietet BBBike [BB16]. BBBike bietet eine Online-Routensuche mit unterschiedlichen Routenkriterien. Nach Eingabe des Quell- und Zielortes lassen sich bevorzugte Geschwindigkeit, bevorzugter Straßentyp und -Oberfläche auswählen. Zusätzlich gibt es die Möglichkeiten Ampeln und unbeleuchtete Wege zu meiden, sowie „grüne“ Wege zu bevorzugen. Die Datengrundlage basiert dabei auf manuell eingetragenen Straßeneigenschaften. BBBike ist auch als Smartphone-App in begrenzter Funktionalität zu erhalten. Die Input-Daten sind nicht öffentlich zugänglich, sondern dienen nur als Hintergrundinformation für die Routensuche.

Viele Portale wie gps-tour.info oder der ADFC stellen empfohlene Routen zur Verfügung. Diese Routen sind manuell angelegt und nicht individualisierbar. Sie unterliegen subjektiven Unterscheidungskriterien der Routenautoren.

3 Komfortfaktoren

Die bisherigen Untersuchungen lassen eine grobe Kategorisierung der komfortbedingten Einflüsse auf den Radverkehr zu. Dabei lassen sich prinzipiell die Einflussbereiche Straßenraum, Umwelt und Sicherheit abgrenzen. Die drei Einflussbereiche sind dabei sehr eng miteinander verbunden, haben Überschneidungen und stehen miteinander in Wechselwirkung.

3.1 Einflussbereich Straßenraum

Im Bereich Straßenraum sind die wohl am schwierigsten zu erfassenden Faktoren angesiedelt. Der Einflussbereich Straßenraum umfasst alle Faktoren, die durch die vorhandene

Infrastruktur bedingt sind. Als wesentlicher Faktor ist hier die Streckenführung zu nennen. Diese hat vorrangig Einfluss auf die Verkehrssicherheit der Radfahrer, sowie auf deren Wartezeiten. Je sicherer sich der Radfahrer fühlt, desto eher wird er dieses Verkehrsmittel nutzen [DV14]. Mit Hilfe von Sensorik das Gefühl von Sicherheit zu testen, mag eventuell möglich sein, allerdings ist es nicht vorgesehen den Fahrer mit Sensorik zu bestücken, sondern das Fahrrad. Daher kann die direkte Sicherheit im Straßenverkehr nicht bemessen werden, allerdings Indikatoren für Sicherheitsprobleme, wie zum Beispiel starke Abbremsvorgänge. In Gefahrensituationen wird es zumeist zu starkem Abbremsen kommen. Bei solchen Gefahrenbremsungen treten stärkere Kräfte auf, als bei üblichen Abbremsvorgängen. Durch die Aufzeichnung von Beschleunigungswerten in allen drei Freiheitsgraden, in Verbindung mit GPS-Koordinaten können Gefahrenstellen mit Hilfe von Sensorik ermittelt werden.

Normalerweise wird ein Fahrradfahrer strecken meiden, in denen er erheblich durch den Kraftfahrzeugverkehr behindert wird. Einerseits wird dieser Einfluss durch die Erfassung der Abbremsvorgänge ermittelt und zum anderen können Umgebungslärm oder Luftschadstoffe ermittelt werden, die Rückschlüsse über naheliegenden Kraftfahrzeugverkehr zulassen. Die Faktoren Lärm und Luftschadstoffe werden grundsätzlich aber eher in den Bereich Umweltfaktoren fallen. Ein weiterer Überschneidungsbereich mit dem Bereich Umwelt lässt sich aus Lichteinstrahlungsstärken ableiten. Die Sonnenstrahlung. Je nach Situation und Uhrzeit lassen sich aus der Sonneneinstrahlungsstärke, sowie der Farbwerte der Lichtreflexion, eventuelle Rückschlüsse über nebenstehende Bebauungen ziehen.

Die Straßenraumgestaltung ist situativ komplett individuell und es lassen sich daher nur sehr schwer pauschale Aussagen über den positiven, beziehungsweise negativen Einfluss dieser tätigen. Die Wegewahl eines einzelnen Probanden gibt jedoch Rückschlüsse über seine persönlichen Präferenzen. Umfasst die Menge der Probanden, die ein Sensormodul an Ihrem Fahrrad haben, eine ausreichend große Anzahl, so könnten aus der Gesamtmenge an Teilwegen auf einem bestimmten Streckenabschnitt theoretisch Rückschlüsse über den Einfluss der Straßenraumgestaltung gezogen werden. Der Mensch ist hierbei sozusagen der Sensor. Wird eine Strecke überdurchschnittlich oft genutzt, so kann davon ausgegangen werden, dass die Streckenführung und/oder die Straßenraumgestaltung hier vorteilhaft für das Fahrrad ist. Denkbar wäre auch ein direkter Feedback-Knopf am Sensormodul, wenn einem der momentane Streckenabschnitt besonders gefällt oder nicht.

Weitere Einschränkungen für den Komfort einer Fahrradstrecke sind Streckenverlaufsspezifische Eigenschaften wie extreme Steigungen und Gefälle, starke und häufige Kurvenfahrten, sowie häufige Wartezeiten.

Starke Steigungen sind von enthusiastischen Fahrradfahrern zwar mitunter sogar gewollt, stören den normalen Fahrradfahrer aber eher bei seiner Fahrt. Bereits ab Steigungen von 3% ist der Kraftaufwand für die Steigungsbewältigung ebenso hoch, wie der Kraftaufwand für die Bewältigung des Luftwiederstandes [SS07]. Zu einer Datenerhebung zur Komfortabschätzung von Fahrradteilstrecken muss eine Höhenerfassung demnach definitiv dazugehören.

Je nach Winkel der Kurve, kann diese die freie Fahrt erheblich beeinflussen. Vor starken Kurven muss abgebremst und anschließend wieder beschleunigt werden. Ebenso wie starke Steigungen, ist dies wohl eher nur von Radfahrer-Enthusiasten gewollt. Kurvenfahrten sollten deshalb auch erfasst werden.

Oft spielt auch der Fahrbahnbelag eine Rolle für eine komfortable Fahrradfahrt. Je nach Federung können häufige Erschütterungen oder gar Schlaglöcher das Wohlbefinden des Fahrers erheblich beeinträchtigen. Zudem führen ungleichmäßige Beläge zu einer Erhöhung des Kraftaufwandes.

3.2 Einflussbereich Umwelt und Erholung

Wie bereits in einigen vorherigen Untersuchungen beschrieben, ist das Fahrrad als Verkehrsmittel nicht nur aufgrund seiner Effizienz genutzt, sondern oft auch wegen dem Erholungsfaktor. Besonders zum Tragen kommt dieser natürlich im Bereich der Freizeitverkehre. Laut dem MiD sind 36% aller zurückgelegten Wege mit dem Fahrrad Freizeitverkehre [II10].

Die Erholung wird sich beim Fahrradfahren oft stärker ausprägen, wenn die Fahrt in naturnahen Bereichen stattfindet, als in überfüllten Hauptverkehrsstraßen. Direkt messbar ist dieser Einfluss durch Luftschadstoffsensoren. Hierbei ist vor allem die Feinstaubbelastung zu nennen. Wie die Changzhou-Studie [XBA15] zeigte, ist dieser ein besonders guter Indikator für die örtliche Schadstoffbelastung durch Kraftfahrzeuge. Wird eine hohe Feinstaubbelastung im Straßenraum gemessen, kann man davon ausgehen, dass weitere verbrennungsbedingte Luftschadstoffe vermehrt vorhanden sind. Hohe Konzentrationen von Luftschadstoffen treten vor allem an Hauptverkehrsachsen und in Luftaustauscharmen Straßenräumen auf. Diesen Luftschadstoffen ist der Fahrradfahrer im Gegensatz zum Autofahrer direkt ausgesetzt.

Eine weitere starke Beeinträchtigung des Erholungsfaktors stellt eine hohe Lärmbelastung dar. Ständiger Verkehrs- oder Stadtlärm verursacht Stress. Die Lärmbelastung kann direkt mit einem Mikrophon gemessen werden. In dicht bebauten, städtischen Straßenräumen wird der Schall reflektiert und verstärkt sich zusätzlich.

Abgeleitete Indikatoren für die Erholung während der Fahrradfahrt sind die Lichteinstrahlung, sowie das Fernbleiben von kraftfahrzeugverkehrsbedingten Wartezeiten.

3.3 Einflussbereich Sicherheit

Die ERA betont die hohe Bedeutung der Sicherheit des Verkehrssystems Fahrrad [FG10]. Studien, wie jene an der Universität Maryland, bestätigen, dass sich das Gefühl der Sicherheit besonders auf die Radnutzung niederschlägt. Mehr als 60% aller Fahrradbesitzer

befürworteten die Aussage, dass separate Fahrradstreifen sie ermutigen würden das Fahrrad zu nutzen. Beinahe 40% der Fahrradfahrer gaben an, dass eine bessere Straßenbeleuchtung sie zu häufigerer Fahrradnutzung animieren würde [AG09].

Die getrennte Führung des Radverkehrs ist ein Thema, dass sich in allen drei Einflussbereichen wiederfindet. Indikatoren für die gemeinsame Nutzung eines Straßenraumes durch Fahrräder und Kraftfahrzeuge könnten sein; die erhöhte Luftschadstoffanwesenheit, gewissen Geräuschmuster durch vorbeifahrende Kfz und starke Abbremsvorgänge in Gefahrensituationen oder Konfrontationen.

Besonders gefährlich sind unübersichtliche, oder schwer verständliche Verkehrssituationen. Meist sind dies Knotenpunkte. Da der Radfahrer weitestgehend ungeschützt gegen Unfälle ist, können schon kleinere Kollisionen mit dem Kraftfahrzeugverkehr zu großen gesundheitlichen Schäden führen.

Der Einflussfaktor Beleuchtung kommt nachts gesondert zum Tragen. Eine gute Straßenbeleuchtung gibt dem Radfahrer ein gesteigertes Gefühl von Sicherheit in der Nacht [DV14]. Die Straßenbeleuchtung kann in Abend-/Nachtstunden über die Lichteinstrahlung bemessen werden.

4 Datenerhebung

Das entwickelte Sensormodul besteht hauptsächlich aus Mikrocontrollerbausteinen und Sensormodulen des Herstellers Tinkerforge. Die Tinkerforge-Module bieten den Vorteil, dass sie sehr unkompliziert miteinander zu verknüpfen sind und eine einheitliche, gut dokumentierte API bieten. Auf diese Weise lassen sich Verwendung und Einsatzbereich der Sensoren anpassen, ohne zusätzliche Hardware und unterschiedliche Programmiersprachen nutzen zu müssen [TF16].

4.1 Sensorik

- IMU

  Die verwendete IMU (Inertial Measurement Unit) ist das Tinkerforge IMU Brick 2.0. Das in Abbildung 2 dargestellte Brick nutzt den BNO 055-Sensor von Bosch Sensortec [BS14]. Der Sensor vereint die MEMS-Sensoren eines triaxialen Beschleunigungsmessers, eines Gyroskops und eines Triaxial-Magnetometers. Das IMU Brick 2.0 beinhaltet zusätzlich die Komponenten einer AHRS (Attitude and heading reference system). Das Kalman-Filter berechnet Quaternionen, lineare Beschleunigung, Schwerkraftvektor sowie unabhängige Gier-, Roll- und Nick- Winkel. Das Bricklet kalibriert sich zudem selbst während des Betriebes. Fehlereinflüsse sind dadurch bereits weitestgehend minimiert. Die IMU bietet, als Folge dessen, die

Möglichkeit die Orientierung im Raum nicht nur in Eulerwinkeln auszulesen, sondern auch in Quaternionen. Dies ist in diesem Anwendungsfall nicht nötig da die Gefahr eines Gimbal Locks beim Fahrradfahren nicht gegeben ist und die Orientierung des Fahrrades nicht in einem externen Koordinatensystem betrachtet werden muss. Des Weiteren gibt das Bricklet die lineare Beschleunigung aus, die um den Einfluss der Erdbeschleunigung bereinigt ist. Das Kalman-Filter wird bereits durch die API der Sensorik unterstützt und muss nicht nachträglich berechnet werden.

Abb. 2: IMU Brick 2.0, [TF16]

- GPS

  Das GPS Bricklet von Tinkerforge nutzt das PA6H-GPS-Modul von GlobalTop Technology. Laut Hersteller sind die Positionsdaten bis auf 3 Meter genau [GT11]. Das GPS-Modul liefert GGA-Datensätze (Global Positioning System Fix Data). Diese beinhalten die aktuelle UTC-Zeit, Breiten- und Höhengrad, sowie die Anzahl genutzter Satelliten und den „Position Fix Indicator“. Der Position Fix Indicator liefert Information ob aktuell GPS-Daten, oder sogar DGPS (Differential GPS) -Daten verfügbar sind. Genauere Informationen zum Fix-Status und der damit zusammenhängenden Satellitenverfügbarkeit, sowie Informationen zur Positionsgenauigkeit (PDOP, HDOP und VDOP) werden in GSA-Daten übermittelt. Des Weiteren können aus RMC (Recommended Minimum Navigation Information) - und VTG (Course and speed information relative to ground) - Daten auch Informationen zur aktuellen Geschwindigkeit des Objektes auf der Erdoberfläche gewonnen werden.

- Dust Detector

  Das Tinkerforge Dust Detector Bricklet nutzt den Sharp GP2Y1010AU0F-Compact Optical Dust Sensor. Der Sensor bestrahlt mit einer Infrarot-Diode die Luft [SC06]. Das von Staubpartikeln reflektierte Licht wird mit Hilfe eines Fototransistors detektiert. Der Sensor kann so Partikel ab 1µm erkennen. Demnach können die Feinstaube

PM10 und PM2,5 nachgewiesen werden. Ultrafeine Partikel (≤0,1 µm) können nicht gemessen werden. Bis zu einer Partikeldichte von 500 µm/m³ können anhand der Ausgangsspannung sehr genau die Mikropartikel in der Luft gemessen werden (Abbildung 3).

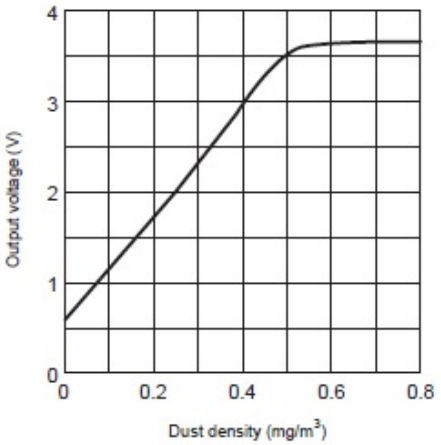


Abb. 3: Output Voltage vs. Dust Density, [SC06]

- Temperatur, Luftfeuchte und Barometer

  Das Tinkerforge Temperatur-Bricklet nutzt den Texas Instruments TMP102 Low Power Digital Temerature Sensor (Abbildung 4). Dieser ist im Bereich zwischen -25°C und +85°C auf 0,5°C genau [TI07]. Die Luftfeuchte wird mit Hilfe des Honeywell HIH-5030 Low Voltage Humidity Sensors gemessen.

  Der Sensor misst in 0,1%-Schritten die relative Luftfeuchte zwischen 0% und 100% [HI10]. Der Hersteller gibt dabei in Bereichen zwischen 11% und 89% Luftfeuchte eine Genauigkeit von ±3% an.

  Das Barometer Bricklet nutzt den MS5611-01BA01 Barometric Pressure Sensor von measurement specialities. Die Messgenauigkeit des Sensors ist von der Temperatur abhängig. Bei 25°C liegt die Messgenauigkeit bei ±1,5 mbar, zwischen 0°C und 50°C ist die Messgenauigkeit bei ±2 mbar: [MS12] Mit Hilfe der simultan gemessenen Temperatur kann durch die barometrische Höhenformel der Luftdruck in Referenz zur Meereshöhe und temperaturkorrigiert berechnet werden.

- Licht-, Farb- und Lautstärkesensorik

  Der TCS3472 Color Light-to-digital Konverter von Taos, der im Color Bricklet von Tinkerforge verbaut ist, misst die Farbstärken RGB, die Farbtemperatur und die Beleuchtungsstärke [TA12]. Die Farbstärken werden für die jeweiligen Farben Rot, Grün, Blau in 16Bit-Werten (0-65535) zurückgegeben. Die Helligkeit in

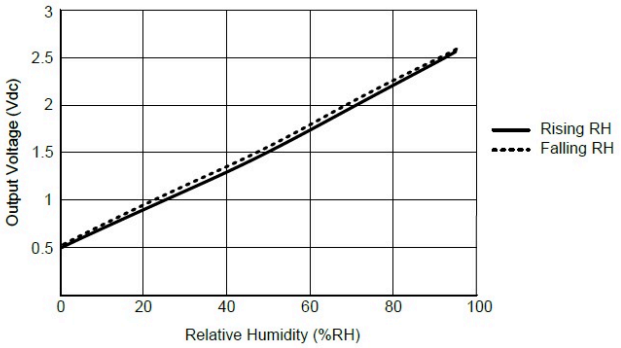


Abb. 4: Output Voltage vs. Relative Humidity, [HI10]

Lux. Der Sensor kann über die Verstärkung (Gain) und die Integrationszeit an die Nutzungsbedingungen angepasst werden.

Das Sound Intensity Bricklet misst die Schallintensität mit einer Mikrofonkapsel. Der zurückgegebene Lautstärkepegel wird zwischen 0 und 4095 zurückgegeben. Inwiefern dieser in Dezibel umrechenbar ist wird nicht angegeben [TF16]. Insofern kann die Lautstärke nur relativ angegeben werden, beziehungsweise muss durch Testmessungen referenziert werden.

- RED Brick

  Das RED (Rapid Embedded Development) -Brick ist das Herzstück des Sensormoduls (Abbildung 5). Es handelt sich hierbei um einen Einplatinencomputer, mit einem 1 Gigahertz (Allwinner A10s) -Prozessor, mit 512 Megabyte DDR3 SDRAM [TF16].

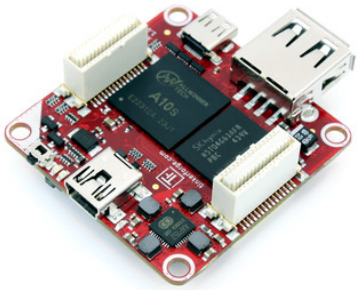


Abb. 5: RED Brick, [TF16]

Als Flash-Speicher wird eine 16 GB-Mirco-SD-Karte genutzt. Das Brick nutzt als Betriebssystem Debian-Linux. Die Stromversorgung, sowie serielle Verbindungen zu einem externen PC können über einen Mini-USB-Anschluss realisiert werden.

Das RED Brick wird dazu genutzt die anderen Bricklets zu steuern, die Sensordaten auslesen zu lassen und zu speichern. Über den Brick Viewer können ausführbare Programme in zahlreichen Programmiersprachen auf den Brick gespielt werden. Zudem kann direkt auf die Linux-Shell des Bricks zugegriffen werden.

Über den Tinkerforge Stack-Connector ist das RED Brick mit dem IMU-Brick und einem Master-Brick verbunden. An diese wiederum sind die weiteren Bricklets angeschlossen.

4.2 Programmierung

Das eigentliche Programm, das die Sensorik und den Auslese- und Speichervorgang steuert ist in Java geschrieben. Das Programm wird automatisch gestartet, wenn der RED Brick gebootet hat.

Vor der eigentlichen Sensorauswertung soll das Programm einige Einstellungen vornehmen. Ist ein USB-Stick angeschlossen, so wird dieser als Export-Verzeichnis der Daten gewählt. Das Exportverzeichnis wird in den globalen Variablen festgelegt. Ist kein USB-Stick angeschlossen werden die Daten in einem Ordner auf dem RED Brick gespeichert.

Anschließend wird im gewählten Verzeichnis nach einer Konfigurationsdatei gesucht. Es handelt sich hierbei um eine Textdatei die eine Erheber-ID, den zugehörigen Namen, sowie die Anzahl der bisher erhobenen Fahrten dieses Erhebers enthält. Die Erheberinformationen werden, falls vorhanden, ausgelesen und die Fahrt-ID um eins erhöht. Falls keine config-Datei vorhanden ist, wird eine neue erstellt und die Werte auf default-Werte gesetzt. Auf diese Weise können Metadaten zu den Erhebungsfahrten gesammelt werden. Durch die Zuweisung eindeutiger Erheber-ID´s und Fahrt-ID können beispielsweise die Anzahl gefahrener Kilometer, Minuten oder Streckenabschnitten je Person und Fahrt erfasst werden. Zudem kann jeder Erheber mit einem USB-Stick und der darauf befindlichen Config-Datei das Sensormodul für sich nutzen und erhält anschließend die Sensordaten-Auswertung auf seinem USB-Stick.

Die Sensordatenauswertung wird in CSV-Dateien exportiert. Die Dateien werden jeweils mit dem aktuellen Datum benannt. Zu Beginn der Erhebung wird deshalb jeweils geprüft ob eine CSV-Datei des aktuellen Datums bereits vorhanden ist, ansonsten wird eine Initiale CSV-Datei erstellt und die Tabellenüberschriften angefügt. Vor der Eigentlichen Erhebung müssen noch einige Sensoren kalibriert werden.

Die Sensordatenerhebung wird prinzipiell durch zwei Kontrollstrukturen realisiert. Hierbei handelt es sich einerseits um eine periodische Auslesung von mehreren Sensorwerten. In einem festgelegten Sekunden-Intervall wird ein Runnable gestartet. Dieses liest Temperatur, Luftfeuchte, Luftdruck, RGB-Farbintensitäten, Lichtstärke, die Partikeldichte, GPS-Koordinaten, sowie den zugehörigen EPE (Estimated Positioning Error), die geschätzte

Geschwindigkeit auf der Erdoberfläche, Euler-Winkel, sowie das Datum und die Uhrzeit aus. Diese Daten werden zusammen mit den Metadaten in die CSV exportiert.

Nebenläufig laufen verschiedene EventListener. Diese werden aktiviert, falls gewisse Sensorwerte über- beziehungsweise unterschritten werden. Diese Sensorwerte sind die linearen Beschleunigungen in der x-, y- und z-Achse, sowie der Schallpegel. Sobald einer dieser Werte eine bestimmte Grenze überschreitet, werden der jeweilige Wert und die Kennzeichnung des auslösenden Events mit dem aktuellen Datum und der Uhrzeit in eine separate CSV-Datei exportiert.

Da die interne Uhr des RED Bricks sich nicht automatisch synchronisiert und daher nicht das korrekte Datum und Uhrzeit ausgibt, wird jeweils pro Fahrt ein Referenzwert vom GPS-Bricklet gespeichert. Somit kann im Nachhinein die tatsächliche Uhrzeit und das Datum bestimmt werden.

4.3 Kalibrierung

Die Kalibrierung der Sensoren dient der Anpassung an Erhebungs- und Anforderungsbedingungen. Hierzu zählen das Setzen von Parameter, Referenzwerten, Grenzen, sowie das Nivellieren.

Zu Beginn einer neuen Erhebung werden die Verstärkung (Gain) und die Integrationszeit des Color Bricklets eingestellt. Die Verstärkung ist prinzipiell davon abhängig, ob am Tag oder in der Nacht gefahren wird. Tagsüber ist die Lichtstärkemessung auch ohne Verstärkung ausreichend. Es hat sich jedoch tagsüber gezeigt, dass bei einem vierfachen Gain Farben besser erkannt werden und auch in der Nacht ein klarer Eindruck der Beleuchtungsstärke zu erkennen ist. Eine erhöhte Integrationszeit verlangsamt zwar die Messung, erhöht jedoch die Genauigkeit. Da in Abständen von mehreren Sekunden gemessen wird, kann die Integrationszeit relativ hoch, auf 101ms gesetzt werden.

Zusätzlich wird zu Beginn der Messung der Referenzluftdruck auf null gesetzt. Die erhobenen Werte stellen dann Luftdruckveränderungen gegenüber der Starthöhe dar. Somit kann keine exakte Höhenangabe getroffen werden, sondern lediglich Aussagen über Höhenveränderungen. Diese sind jedoch merkbar präziser.

Zudem werden die Auslösungsgrenzen für die IMU bestimmt. Durch verschiedene Testversuche wurden die Grenzen für die X-, Y- und Z-Beschleunigung auf 4m/s², 2m/s² und 2,5m/s² gesetzt. Diese sind so gesetzt worden, dass bereits kleinere Bodenunebenheiten, leichte Abbremsvorgänge und starke „Fahrrad-Wackler“ aufgezeichnet werden.

5 Evaluation / Feldtest

Das Sensormodul wurde in und außerhalb der Stadt Karlsruhe getestet um Erkenntnisse über die Tauglichkeit der einzelnen Messgruppen, sowie deren Aussage und Sensitivität zu erhalten. Die erhobenen Daten wurden anschließend mit MS Access nachträglich formatiert und exportiert, sodass erste Analysen und Visualisierungen mit dem Geoinformationssystem QGIS erstellt werden können. Diese dienen der Datenevaluation. Mit Hilfe der GIS-Software werden die jeweiligen Sensordaten gemappt und differenziert dargestellt. Das Ergebnis kann dann hinsichtlich logischer Korrektheit, Erwartungswert und Aussagekraft überprüft werden.

Die Abstände der periodischen Sensordatenauswertung sind auf 5 Sekunden festgelegt. Dadurch ergibt sich bei einer Geschwindigkeit von 20 km/h ein räumlicher Messabstand von etwa 5,5 Metern (Abbildung 6). Die Messdichte ist somit hoch genug um, räumlich sehr differenzierte Aussagen ableiten zu können.

Die Evaluation beruht auf Daten aus vier Testfahrten. Während dieser Testfahrten wurden 2.756 gültige (mit gültigen GPS-Koordinaten) periodische Datensätze und 21.941 Event-Daten erhoben.

Pro periodische erhobenem Datensatz wurden also fast acht Event-Daten erhoben. Da die Event-Daten zur Ermittlung von streckenspezifischen Ausnahmesituationen gedacht sind, sollten die Auslösungsgrenzen der Events erhöht werden. Dies muss ansonsten bei der Auswertung durch zusätzliche Filter geschehen.

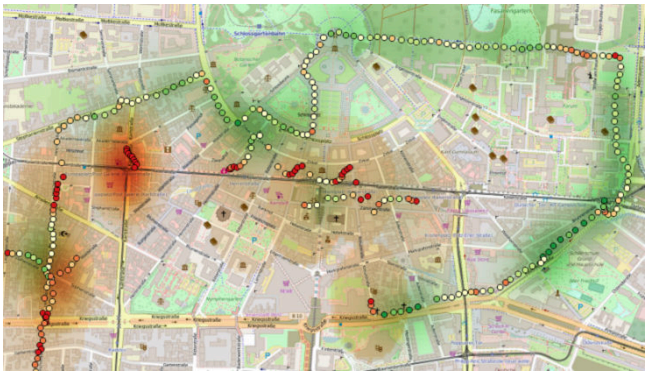


Abb. 6: GPS-Wegpunkte, Kartengrundlage: openstreetmap.org

5.1 Geschwindigkeitsdaten

Der GPS-Sensor gibt die Geschwindigkeit in Knoten aus (Abbildung 7). Die ausgegebenen Werte sind hierbei leicht fehlerbehaftet. Da das Sensormodul jedoch zur massenhaften Datenerhebung gedacht ist, gleichen sich „Ausreißer“ gegenseitig aus, sodass ein plausibles

Ergebnis entsteht, wenn ausreichend Daten vorhanden sind. Die Geschwindigkeiten sind oft zu hoch, sodass Geschwindigkeiten bis 50 km/h gemessen werden. Daher eignen sich die Ergebnisse weniger zu präzisen Einzelaussagen über das Geschwindigkeitsverhalten, allerdings bietet sich die Möglichkeit der prinzipiellen Ermittlung von Standzeiten und Verzögerungen. Die Ergebnisse spiegeln trotz der fehlerhaften Werte, sehr gut wieder ob ein Streckenabschnitt konstant und störungsfrei befahren werden kann, oder ob Verzögerungen aufgetreten sind. Die Daten sind nicht absolut, allerdings relativ sehr aussagekräftig. Standzeiten werden sehr gut erkannt.

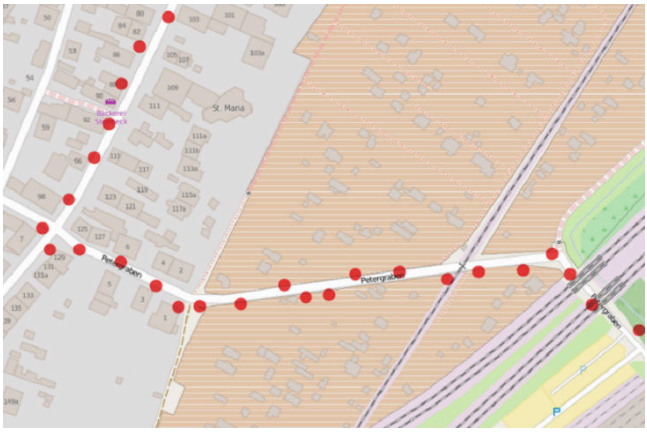


Abb. 7: Geschwindigkeitswerte, Kartengrundlage: openstreetmap.org

5.2 Feinstaubbelastung

Die Auswertungen der Partikeldichte sollen einen direkten Rückschluss über die Feinstaubbelastung zulassen. Nach nur vier Testfahrten kann dies nicht abschließend bestätigt werden. Die Messwerte sind nicht sprunghaft und wirken realistisch. Da die Werte allerdings auch von vielen weiteren Faktoren (Wind, Umwelt, Tageszeit, . . .) beeinflusst werden, können präzise Rückschlüsse erst nach einer sehr hohen Datendichte gezogen werden. Für einen direkten Vergleich sind zwei Vergleichsstrecken unter selben Testbedingung (selber Tag, unmittelbar hintereinander) abgefahren worden. Die Vergleichsstrecken waren die Pulverhausstraße in Karlsruhe und die, nahezu parallele, Fahrradstrecke an dem Fluss „Alb“ (Abbildung 8).

An der vielbefahrenen Hauptstraße (Pulverhausstraße) waren deutlich höhere Feinpartikeldichtewerte vorzufinden, als an der Fahrradstraße an der „Alb“, die entlang des Flusses, abseits des Kfz-Verkehrs führt.

Abb. 8: Feinstaubpartikeldichte, Kartengrundlage: openstreetmap.org

5.3 Temperatur und Luftfeuchte

Die Auswertung der Temperatur zeigt einen klaren Trend. Je urbaner und dichter besiedelt die Erhebungsumgebung ist, desto höher ist die Temperatur. Da jedoch die Temperatur je nach Tages- und Jahreszeit, sowie bei direkter Sonneneinstrahlung noch erheblicher schwankt, als durch die Besiedlung bedingt, sind Rückschlüsse anhand der Temperaturwerte nur durch eine sehr hohe Datenverfügbarkeit oder aufwendige Glättung und Filterung möglich.

Die Messwerte der Luftfeuchtemessungen sind plausibel und ergeben einen klaren Verlauf, lassen allerdings nach ersten Tests keine direkten Rückschlüsse zu. Lediglich eine leicht erhöhte Luftfeuchtigkeit in Wald- und Grünanlagen ist zu verzeichnen. Allerdings dienen Temperatur und Luftfeuchte eher zur Einordnung, beziehungsweise Korrekturrechnung anderer Sensorwerte, weniger als direkter Indikator für räumlich differenzierte Komfortbedingungen des Radverkehrs.

5.4 Höhenverlauf

Das Barometer Bricklet misst die Luftdruckveränderung im Bezug zum Anfangsniveau. Um Höhenänderungen während der Fahrt zu erkennen, müssen zusätzlich die Differenzwerte zwischen den Einzelwerten ermittelt werden. Ist die Differenz $\text{Wert}_x - \text{Wert}_{x+1}$ negativ,

so ist der Luftdruck angestiegen. Steigender Luftdruck bedeutet (bei gleichbleibenden Rahmenbedingungen) sinkende Höhe, also Gefälle. Ist die Differenz positiv, so bedeutet dies Steigung. Betrachtet man die Extremwerte dieser Differenz $|Differenz| \geq 1mbar$ so ergibt sich ein stimmiges Bild: Oft treten starke Luftdruckveränderungen etwas zeitversetzt an Brücken und Unterführungen auf. Luftdruckabfall und -anstieg sind hier paarweise vorhanden.

Der Wert des Pitch-Winkels sollte ebenso Rückschlüsse über die aktuelle Steigung geben. Allerdings zeigte sich, dass an aufgezeichneten Extremstellen und starken Veränderungen des Pitch-Winkels tatsächlich nur sehr vereinzelt große Steigungen oder Gefälle vorhanden waren. Obwohl das Sensormodul fest am Fahrrad befestigt war, sind keine direkten Erkenntnisse aus den Pitch-Winkeln zu gewinnen.

5.5 Abbremsvorgänge

Zur Ermittlung starker Abbremsvorgänge wird mit Hilfe des IMU-Bricks die Y-Achsen-Beschleunigung gemessen. Erwartungsgemäß sollten erhöhte negative Y-Beschleunigungen bei behindertem oder beeinträchtigtem Fahren, beispielsweise durch Kfz-Verkehr oder erhöhtes Fußgängeraufkommen, auftreten. Abbildung 9 zeigt die negativen Y-Beschleunigungswerte im Karlsruher Stadtgebiet als Heat-Map. Die Ergebnisse spiegeln Erwartungswert und das subjektive Empfinden während der Testfahrt wieder: Besonders viele Abbremsvorgänge traten am Karlsruher Marktplatz, bedingt durch den hohen Fußgängerverkehr, sowie an signalisierten Kreuzungen auf.

Abb. 9: Heat-Map: Negative Y-Beschleunigungen, Kartengrundlage: openstreetmap.org

5.6 Bodenunebenheiten

Schlaglöcher, Hindernisse auf dem Boden und weitere potenziell störende Bodenunebenheiten werden ebenso mit Hilfe des IMU-Bricks detektiert. Die Beschleunigung in der Z-Achse

zeichnet eben solche Unebenheiten auf. Starke Ausschläge in der Z-Beschleunigung deuten auf schnelle, ruckartige Höhenveränderungen hin.

Während des Feldtestes wurde dies vor allem auf nicht-asphaltierten Strecken, Feldwegen und an Schlaglöchern getestet. Unasphaltierte Wege wurden mit leichten, kontinuierlichen Ausschlägen aufgezeichnet. Vermerkte Schlaglöcher und starke Wölbungen wurden allesamt mit hohen Z-Beschleunigungs-Beträgen verzeichnet. Die Ergebnisse spiegeln den Erwartungswert wieder.

5.7 Kurvenfahrten

Kurvenfahrten könnten sowohl mit Hilfe der X-Beschleunigung, als auch mit der Veränderung des Head-Winkels erfasst werden. Die Betrachtung der X-Beschleunigungsdaten ergab hierbei allerdings kein logisches Bild. Die Vertikalbeschleunigungen sind eher das Resultat aus unruhigem Fahrverhalten oder Wackeln des Fahrrades.

Die Auswertung der Differenzwerte des Head-Winkels hingegen zeigte sehr präzise und konsistente Ergebnisse zur Detektion von Kurvenfahrten. Besonders starke Kurven konnten auch mit Hilfe von Heat-Mapping der Differenzbeträge nachträglich ermittelt werden. Starke Kurvenfahrten resultieren aus einem großen Differenzbetrag der Head-Winkel zwischen zwei periodischen Messdatensätzen. Diese liegen fünf Sekunden auseinander. Um eine noch präzisere Auflösung von besonders „spitzen“ Kurven zu bekommen, müssten die Messabstände noch geringer gewählt werden. Allerdings konnte auch bei den bisher gewählten Messperioden hinreichend gute und plausible Ergebnisse erzielt werden (Abbildung 10).

Abb. 10: Heat-Map: Kurvenfahrten, Kartengrundlage: openstreetmap.org

5.8 Lichteinfall

Der Lichtsensor erkennt klare Unterschiede zwischen einer Fahrt bei Tageslicht uns einer Fahrt bei Nacht. Ebenso sind Zwischenwerte zu Zeiten der Dämmerung deutlich erkennbar.

Dicht bewaldete Gebiete sind deutlich am Lichteinfallspegel erkennbar. Betrachtet man ausschließlich Nachtfahrten und erhöht die Lichtempfindlichkeit, so ergibt sich auch ein konsistenter Lichtverlauf (Abbildung 11).

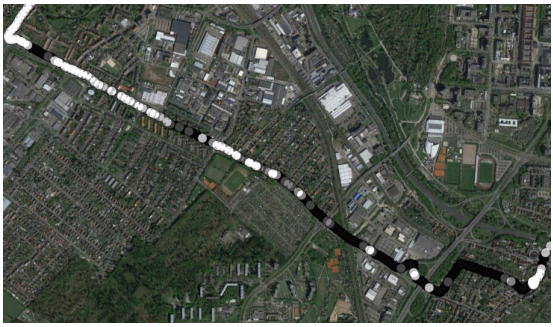


Abb. 11: Lichtintensität bei Nacht Pulverhausstraße, Kartengrundlage: googlemaps.com

Abschnitte mit guter, sowie Abschnitte mit mangelnder/fehlender Beleuchtung sind identifizierbar. Zur Verbesserung der Ergebnisse wäre es sinnvoll, den Gain je nach Jahreszeit und Tageszeit automatisch anzupassen.

Aus den Einzelfarbwerten RGB können, bei aktueller Kalibrierung, keine Rückschlüsse über die Begrünung oder sonstige Umgebungseigenschaften gezogen werden. Bei Tageslicht liegen die Werte meist gemeinsam sehr nahe an Ihrem Maximum und logische Veränderungen sind nicht zu erkennen. Durch Veränderung des Gains, können eventuell repräsentative Daten erhoben werden.

5.9 Metainformationen

Mit Hilfe von einfachen Abfragen können aus den erhobenen Datensätzen Metainformationen zum Fahrradverhalten gewonnen werden. Durch die eindeutige Zuweisung einer Erheber-ID und einer Fahrt-ID zu jedem Datensatz, sowie der Speicherung der genauen Uhrzeit mit Datum sind vielfältige Auswertungen möglich. Durch die Erstellung von Wegeketten aus den GPS-Daten sind beispielsweise auch durchschnittliche Wegelängen berechenbar.

Bei massenafter Datensammlung mit dieser Sensorik und minimalen Erheberinformationen, wie beispielsweise Alter und Geschlecht können auch Verhaltensanalysen zum Radverkehr erstellt werden. Unter Zuhilfenahme der soziodemographischen Erheberinformationen und zugänglichen Wetterinformationen können auch zeitliche Einflussfaktoren auf den Radverkehr bemessen werden.

6 Diskussion

Die Tinkerforge-Sensoren bieten einfache Schnittstellen und sind deshalb mit deutlich geringerem Aufwand kombinierbar und programmierbar, als verschiedene Einzelsensoren von unterschiedlichen Herstellern. Allerdings vereinfachen die Bricklets oft die eigentlichen Kalibrierungs- und Sensorfunktionen der verbauten Sensoren. Die in den Sensordatenblättern beschriebene API der eigentlichen Sensoren bietet in manchen Fällen zusätzliche, nützliche Sensorfunktionen an.

Aufgrund des geringen Ausmaßes des Feldtests sind abschließende Aussagen über die Einzelfunktionalitäten noch nicht möglich. In Einzelsituationsvergleichen konnte jedoch bereits ermittelt werden, ob die Sensorauswertung brauchbare Ergebnisse liefert. Besonders im Bereich der Höhenermittlung und der Licht- beziehungsweise Farbwertermittlung könnten mit Hilfe besserer Kalibrierung und nachträglicher Glättung der Ergebnisse präzisere Rückschlüsse getroffen werden.

Erst durch die Erfassung von Massendaten werden präzise und umfassende Analysen der Fahrradstrecken möglich. Kleinere Datenmengen führen zu einer starken Verfälschung der Ergebnisse. Beispielsweise führt ein kurzer Stopp am Bankautomaten zu einer Verzeichnung großer Wartezeiten. Durch die Glättung im Rahmen von Massendaten fallen diese Fehlereinflüsse größtenteils weg. Zur Erkennung von nicht-verkehrsbedingten Standzeiten, wie zum Beispiel dem Halten zum Geld abheben, muss weitere Hardware an das Sensormodul hinzugefügt werden. Denkbar wäre beispielsweise ein Sensor im Sattel zur Erkennung ob jemand auf dem Fahrrad sitzt, oder ein einfacher Knopf zum Pausieren der Aufzeichnung.

Durch weitere Präzisierung der Messergebnisse kann das entwickelte Sensormodul eine sehr gute Datengrundlage für weiterführende Untersuchungen schaffen. Durch die Einteilung der Einzelergebnisse in Werteklassen ist eine vereinfachte Darstellung und eine vereinfachte Weiterverarbeitung möglich. Mit Hilfe von Gewichtungsfaktoren für die jeweiligen Sensorergebnisse sollte es zudem möglich sein, negative und positive Einflüsse zu verschneiden, beziehungsweise gegeneinander verrechnen zu können. Durch die individuelle Anpassung der Gewichtungsfaktoren, kann jeder Nutzer der Daten einen eigenen Bewertungsschlüssel schaffen, durch den er individuell-angepasste Fahrradstreckenanalysen bekommen kann.

Zum tatsächlichen Einsatz des Sensormoduls müsste noch ein Sensor-Gehäuse entwickelt werden, das den Sensor vor Witterungseinflüssen und sonstigen Schäden schützt, jedoch die Sensorfunktionen nicht negativ beeinträchtigt.

7 Fazit

In Einzelfalluntersuchungen haben sich die meisten Erfassungsverfahren als belastbar gezeigt. Aus den gewonnenen Daten lassen sich bereits durch geringen Interpretationsaufwand Rückschlüsse zum Komfort der gefahrenen Fahrradstrecken gewinnen. Durch weitere

Feinjustierung, Kalibrierung und „tiefere“ Interpretation können die Ergebnisse präzisiert und erweitert werden. Individuelle Fehler bei den Sensorerfassungen können über große Mengen an Daten, also über eine große (breite) Masse an Erhebern, geglättet werden.

Literaturverzeichnis

[AC09] Akar, G., & Clifton, K. J. (2009). The influence if individual perceptions and bicycle infrastructure on the decision to bike. 88th Annual Meeting of the Transportation Research Board. Washington, D.C.

[BB16] bbbike.cgi. BBBike Routensuche. Abgerufen am 27. 06 2016 von http://www.bbbike.de/cgi-bin/bbbike.cgi

[BFG84] Bosselmann, P., Flores, J., Gray, W., Priestly, T., Anderson, R., Arens, E., Kim, J.-J. (1984). Sun, Wind, and Comfort A Study of Open Spaces and Sidewalks in Four Downtown Areas. Berkeley, California: University of California, Institute if Urban and Regional Development.

[BS14] Bosch Sensortec. (2014). Data Sheet BNO 055 Intelligent 9-axis absolute orientation sensor.

[De15] Dezani, H. (2015). Urban road optimization based on safety and cyclists´ effort required by bike tracks. IEEE 18th International Conference on Intelligent Transportation Systems. Las Palmas, Spanien.

[DV14] Das Verkehrspolitische Programm des ADFC. (2014). Berlin: Allgemeiner Deutscher Fahrrad-Club e.V.

[FG10] FGSV, F. f.-u. (2010). Empfehlungen für Radverkehrsanlagen ERA. Köln: FGSV Verlag.

[GL16] Gravitystorm Limited. OpenCycleMap Documents. Abgerufen am 27. 06 2016 von http://www.opencyclemap.org/docs/

[Gö11] Göttlicher, S. (2011). "Weiche" Planungsinstrumente auf regionaler Ebene. In Angewandte Geographie. Heidelberg: Springer Verlag.

[GT11] Global Technology Inc. (2011). FGPMMOPA6H GPS Standalone Module Data Sheet. Taiwan.

[HI10] Honeywell International Inc. (2010). HIH-5030/5031 Series Low Voltage Humidity Sensors Datasheet. Golden Valley, Minnesota.

[HMM13] Hoffmann, M., Mock, M., & May, M. (2013). Road-quality classification and bump detection with bicycle-mounted smartphones. St. Augustin, GER: Frauenhofer IAIS.

[II20] infas Institut für angewandte Sozialwissenschaft GmbH, Deutsches Zentrum für Luft- und Raumfahrt e.V. Institut für Verkehrsforschung. (2010). Mobilität in Deutschland 2008. Bonn und Berlin: Bundesministerium für Verkehr, Bau und Stadtentwicklung.

[IS16] Infrastruktur, B. f. (2016). Bundesverkehrswegeplan 2030. Berlin: Bundesministerium für Verkehr und digitale Infrastruktur.

[MS12] Measurement Specialities. (2012). MS5611-01BA01 Barometric Pressure Sensor, with LCP cap Datasheet.

[MKL15] Miah, S., Kaparias, I., & Liatsis, P. (2015). Evaluation of MEMS Sensors Accuracy for Bicycle Tracking and Positioning. London, UK: IEEE.

[SC06] Sharp Corporation. (2006). GP2Y1010AU0F Compact Optical Dust Sensor.

[Sc09] Schäfer, K.-H. (2009). Öffentlichkeitsarbeit in der Verkehrsplanung. In Handbuch der kommunalen Verkehrsplanung. Berlin: VDE Verlag.

[SE14] Stadtenwticklungsforschung, I. f.-u. (2014). Transferstelle Mobilitätsmanagement. Abgerufen am 31. 05 2016 von http://www.mobilitaetsmanagement.nrw.de/cms1/index.php?option=com_content&view=article&id=201&Itemid=7

[SS07] Suhr, W., & Schlichting, H.-J. (2007). Mit Pedalkraft gegen Berge und Wind. In Physik unserer Zeit (S. 294-298). Weinheim: Wiley-VCH Verlag.

[TA12] Texas Advanced Optoelectronic Solutions Inc. (2012). TCS3472 Color Light-to-Digital Converter with IR-Filter Datasheet. Plano, Texas.

[TF16] Tinkerforge GmbH. Tinkerforge.com. Abgerufen am 29. 05 2016 von http://www.tinkerforge.com/de

[TI07] Texas Instruments. (2007). TMP102 - Low Power Digital Temperatur Sensor With SMBus™/Two-Wire Serial Interface in SOT563 Datasheet. Dallas, Texas.

[TS19] Titov W, Schlegel T: Monitoring Road Surface Conditions for Bicycles: Using Mobile Device Sensor Data from Crowd Sourcing. In: Krömker H: HCI in Mobility, Transport, and Automotive Systems: First International Conference, MobiTAS 2019, Held as Part of the 21st HCI International Conference, HCII 2019 (Orlando, FL, USA, July 26-31, 2019) Proceedings. Cham: Springer, S. 340-356.

[TJT08] Thomas, T., Jaarsma, R., & Tutert, B. (07. 11 2008). Temporal variations of bicycle demand in the Netherlands: The influence of wheater on cycling. Transportation Research Board.

[TNH15] Taniguchi, Y., Nishii, K., & Hisamatsu, H. (2015). Evaluation of a Bicycle-Mounted Ultrasonic Distance Sensor for Monitoring Road Surface Condition. 7th International Conference in Computational Intelligence, Communciation Systems and Networks. Riga, Lettland: IEEE.

[XBA15] Xiaofeng, L., Bin, L., Aimin, J., Shixin, Q., Chaosheng, X., & Ning, X. (2015). A Bicycle-borne Sensor for Monitoring Air Pollution near Roadways. Changzhou, China: IEEE

R. Reussner, A. Koziolek, R. Heinrich (Hrsg.): INFORMATIK 2020,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2021 1137

Die sensorbasierte Vermessung des Radverkehrs

Analyse des Radverkehrs mit einem SensorBike mit ubiquitären Sensoren

Jochen Eckart[1] Jule Merk[2]

Abstract: Vorgestellt wird ein Ansatz zur Erhebung des Radverkehrs mit Hilfe von ubiquitären Sensoren in Form eines SensorBikes. Ziel ist die Erfassung des Radverkehrs aus der Perspektive der Radfahrenden, um neue Ansätze zur Förderung des Radverkehrs zu gewinnen. Basierend auf ersten Erfahrungen aus Erhebungen mit dem SensorBike werden deren Anwendungsfälle sowie die Herausforderungen für die Datenerhebung und Auswertung dargestellt. Thematisiert werden die Möglichkeiten, aber auch Grenzen und Herausforderungen der Erhebung mit einem SensorBike.

Keywords: Fahrrad; Sensoren; SensorBike; Anwendungsbereiche; Abstand; Erschütterungen; Leistungsmessung

1 Der Bedarf nach einem Perspektivwechsel

Radverkehr ist ein wichtiger Verkehrsträger der Zukunft. Radfahren vereint persönliche (aktive Mobilität, körperliche Fitness) und gesellschaftliche (emissionsfrei, leistungsfähig, flächensparsam, stadtverträglich) Vorteile für eine zukunftsfähige und nachhaltige Mobilität in Stadt und auf dem Land. Die Förderung des Radverkehrs basiert bisher häufig auf der Prämisse „was denken Planer, was die Radfahrenden brauchen“. Um den Radverkehr fortzuentwickeln erscheint ein Verständnis davon „was Radfahrende wirklich brauchen“ erforderlich. Um den zunehmenden Handlungsbedarf für den Radverkehr in den Bereichen Verkehrssicherheit, Leistungsfähigkeit und Komfort zu bewältigen, ist verstärkt die radfahrerzentrierte Perspektive zu berücksichtigen. Die bestehenden Erhebungsinstrumente zum Radverkehr in Form von Nutzerbefragungen, Unfallanalysen oder die Analyse lokaler Verkehrssituationen sind dafür nicht ausreichend. Um eine radfahrerzentrierte Perspektive zu schaffen, sind digitale Tools zur Erfassung der Daten der Radfahrenden einzusetzen und die gewonnenen Informationen für Planungsprozesse nutzbar zu machen. Dabei werden die Daten der Radfahrenden während der Fahrradfahrt selbst mit Hilfe von ubiquitären Sensoren erfasst und für die Förderung des Radverkehrs nutzbar gemacht. Die sensorgestützten Studien ermöglichen die Untersuchung einer radfahrerspezifischen Perspektive, die sowohl

[1] Hochschule Karlsruhe Technik und Wirtschaft, Studiengang Verkehrssystemmanagement, Moltkestr. 30, 76133 Karlsruhe, jochen.eckart@hs-karlsruhe.de
[2] Hochschule Karlsruhe Technik und Wirtschaft, Studiengang Verkehrssystemmanagement, Moltkestr. 30, 76133 Karlsruhe, jule.merk@hs-karlsruhe.de

doi:10.18420/inf2020_106

den Radfahrer als Nutzer, sowie die genutzte Fahrradinfrastruktur in das Zentrum der Untersuchung rückt. Ziel ist neue Potenziale rund um schnelleres, sicheres, komfortableres und kraftsparendes Radfahren zu schaffen.

2 Entwicklung der sensorgestützten Forschung im Radverkehr

Die wissenschaftliche Auseinandersetzung mit dem Verkehrsteilnehmer „Radfahrer“ erfolgt bereits seit mehreren Jahrzehnten (vgl. [MM90] und [Kn95]). Dabei kamen zunächst insbesondere beobachtende Studien wie Nutzerbefragungen, Unfallanalysen oder Analysen lokaler Verkehrssituationen zum Einsatz. Die Entwicklung von sensorgestützten Erhebungen des Radverkehrs erfolgt erst in den letzten Jahren im größeren Umfang. Ansätze zur Erhebung des Radverkehrs mit Hilfe von Sensoren wurden u.a. in Kampagnen und Vorhaben in Berlin [Ta18], Köln [Be19] oder Dresden [BS19] eingesetzt.

Diese Entwicklung ist unter anderem darauf zurückzuführen, dass Sensoren zunehmend günstiger und ubiquitär verfügbar werden. Zudem hat sich die Qualität der Sensoren deutlich gesteigert, was die Daten zuverlässiger für den Einsatz in Forschung und Planungspraxis macht. Zusätzlich ist die Handhabung der Sensoren deutlich einfacher geworden, so dass auch Akteure außerhalb der Informatik und Messtechnik diese nutzen können.

Dabei kann auf verschiedene, bereits im Massenmarkt vertretene Sensoren und Anwendungen zurückgegriffen werden. So sind z. B. die aus dem Freizeit- und Sportbereich stammenden Sensoren, wie Fahrradcomputer und Fitness-Apps preisgünstig und ausgereift und können auch für weitere Einsatzzwecke zur Messung des Radverkehrs umgenutzt werden [EH19]. Viele Erhebungsprojekte konzentrieren sich auf Sensorik die in handelsüblichen Smartphones eingebaut sind (insbesondere GPS sowie Erschütterungs- und Beschleunigungssensoren) [Vi20]. Für zahlreiche relevante Fragen wie z. B. die Überholabstände und –geschwindigkeiten werden jedoch weitere Sensoren benötigt. Weiterhin werden aber auch basierend auf günstigen Elektronikbausätzen wie z. B. Arduino eigene Sensorsysteme entwickelt.

Die im Radverkehr angewandten Messmethoden lassen sich in zwei Typologien unterteilen. Zum einem klassische Labor- und Feldversuche mit Radfahrenden, in kleinen Stichprobengrößen. Zum anderen werden im Rahmen von partizipativer Forschung und mit Hilfe von Crowd-Sensing Ansätzen auch Untersuchungen mit deutlich größeren Teilnehmerzahlen durchgeführt. Dies basiert häufig auf Kampagnen zur Messung des Radverkehrs durch Initiativen in den Kommunen oder durch kommerzielle Consulting Firmen (z. B. bikecitizen [Bi20]).

Die Beobachtung dieser Messprojekte zeigt, dass die Ergebnisse der Messungen vielfach eher zurückhaltend in der Planungspraxis der Radverkehrsförderungen aufgegriffen werden. Die Daten und Informationen aus den bisherigen Messprojekten sind aufgrund ihrer Art und Struktur vielfach nicht leicht genug nutzbar. Der Mehrwert der erhobenen Daten ist für

die Akteure aus der Planungspraxis im Vergleich zum Erhebungsaufwand nicht ausreichend. Es stellt sich daher die Frage, wie die erhobenen Daten anschlussfähig an die Prozesse der Radverkehrsförderung gemacht werden können. Zudem ist zu verdeutlichen, welche neuen Möglichkeiten zur Förderung des Radverkehrs durch die Vermessung der Radfahrenden geschaffen werden.

Darüber hinaus bestehen teilweise noch praktische Herausforderungen bei der Sammlung, Aggregation und Auswertung der Messdaten. Es stellt sich die Frage, wie eine leicht verständliche und schnelle Handhabung der erhobenen Daten durch die Akteure aus der Praxis gewährleistet werden kann. Zudem stellen sich Herausforderungen bei der Skalierung dieser sensorbasierten Messungen des Radverkehrs, um eine umfassende Anwendung in der Praxis zu ermöglichen.

3 Das SensorBike der Hochschule Karlsruhe

Um den Einsatzbereich der Messung des Radverkehrs mit Sensoren zu erweitern, wurde an der Hochschule Karlsruhe ein SensorBike entwickelt. Das SensorBike ist ein Messfahrrad, welches als Untersuchungsinstrumentarium für die angewandte Forschung im Radverkehr entwickelt wurde. Es basiert auf einem handelsüblichen Trekkingrad, welches durch den Trapezrahmen und die geringe Rahmenhöhe für beide Geschlechter nutzbar ist. Durch den Anbau von diversen Sensoren (Tab. 1) kann eine Vielzahl an Daten zeitgleich erhoben und in der Analyse miteinander verknüpft werden. Die Sensoren und Anwendungsbereiche des SensorBikes umfassen die Einflussgrößen für den Kraftaufwand, die Verkehrssicherheit sowie den Fahrkomfort der Radfahrenden. Die Einflussgrößen, die sich auf den Kraftbedarf beim Radfahren auswirken, wie Längsneigung, Windgeschwindigkeit, Fahrbahnoberfläche, Reifendruck, Gewicht etc. werden erfasst und dem Energieverbrauch der Radfahrenden sowie weiteren Vitalparametern wie Puls gegenübergestellt. Einflussgrößen, die sich auf den Fahrkomfort der Radfahrer auswirken, wie Witterung/Klima, Erschütterungen, Luftqualität, Lärmbelastung, Belichtung etc., bilden wichtige Einflussgrößen für die Verkehrsmittelwahl und Routenwahl der Radfahrenden. Die Erhebung der Einflussgrößen zur Verkehrssicherheit, wie Seitenabstände und Geschwindigkeit des umliegenden Verkehrs, Bremsbeschleunigungen, Verkehrskonflikte etc., ermöglichen neue Ansätze zur Bewertung der Verkehrssicherheit des Radverkehrs.

Tab. 1: Die verbauten Sensoren des SensorBikes.

Themenfeld	**Sensor/Technologie**	**Anwendungsbereich**
Sicherheit	Markierungstaste / Feedbacktaste und Zeitlogger	Markierung von positiven / negativen Stellen durch Radfahrenden
	Beschleunigungssensor	Kritische Brems- und Beschleunigungsvorgänge

Tab. 1: Die verbauten Sensoren des SensorBikes.

Themenfeld	Sensor/Technologie	Anwendungsbereich
	Erschütterungssensor	Fahrbahnbelag, Komfortgefühl der Radfahrenden
Sicherheit	Kamera vorne und hinten	Verkehrssituationen, Verkehrskonfliktanalyse, Sichtverhältnisse, Wetter, Beleuchtung, Umfeldanalyse, Fahrbahnoberflächen
	Entfernungsmesser	Seitenabstände zum überholenden Fahrzeug und zum Seitenraum
	Bremssensoren	Detektion für Berührung und Zug an Bremsen; Identifikation kritischer Bremsvorgänge
Kraftaufwand	Leistungsmesskurbel im Tretlager	Messung Energiebedarf zum Radfahren
	Körperfunktionssensoren	Messung von Herzfrequenz, Blutdruck, Hautleitfähigkeit, Hauttemperatur, Sauerstoffsättigung
	Fahrradcomputer	Geschwindigkeit, Trittfrequenz
	GPS	Routennachverfolgung inkl. Höhenprofil
Komfort	Thermometer, Hygrometer, Lichtsensor	Körperliches Wohlbefinden mit Temperatur, Luftdruck, Luftfeuchte, Beleuchtungsstärke, Windgeschwindigkeit
	Feinstaubmessgerät	Schadstoffbelastung Luft
	Schallpegelmesser	Umgebungslärm, Stress
	Beschleunigungssensor	Fahrbahnzustand (Beschaffenheit Oberfläche, Feuchtigkeit)
	Gyroskop	Schwankungen im Fahrrad, Kurvenfahrt
	Drucksensor	Luftdruck Fahrradreifen

Das SensorBike wurde unter Nutzung verschiedener Sensoren bereits in über 16 Projekten und Abschlussarbeiten an der Hochschule Karlsruhe eingesetzt. Im Folgenden werden die Erfahrungen aus diesen Projekten zu folgenden Fragen geteilt:

- Welche Users Cases bzw. Anwendungsbereiche für die sensorbasierten Messungen des Radverkehrs bestehen in der Praxis der Radverkehrsförderung?
- Welche praktischen Herausforderungen bestehen beim Skalieren, Sammeln, Aggregieren und Auswerten der Daten für die Vermessung des Radverkehrs?

4 Anwendungsbereiche für das SensorBike

Um die praktische Anwendung der sensorbasierten Messung des Radverkehrs zu unterstützen sind zunächst die Themenfelder zu identifizieren, in welchen die Messungen auch in der Praxis genutzt werden können. Um die Ergebnisse der Messungen möglichst praxisrelevant zu gestalten, bietet sich eine Orientierung an den Aufgaben der kommunalen Radverkehrsförderung und –planung an.

4.1 Gesamtkonzepte Fahrradförderung

In der Praxis hat sich die Strategie der „Radverkehrsförderung mit System“ bewährt, die am Gesamtsystem Fahrrad mit den Elementen technische Infrastruktur, digitale Infrastruktur, Serviceleistungen für den Radverkehr sowie die Fahrradkultur ansetzt, wie am Beispiel der Stadt Karlsruhe [St13] ersichtlich. Diese ganzheitlichen Konzepte haben in zahlreichen Kommunen zur umfassenden Steigerung des Radfahranteils geführt.

Die Daten aus einer sensorbasierten Erhebung des Radverkehrs können zur Entwicklung oder Fortschreibung von Gesamtkonzepten zur Fahrradförderung z. B. im Rahmen von Verkehrsentwicklungsplänen eingesetzt werden. Die Ergebnisse zur sensorbasierten Erhebung des Radverkehrs können als Ergänzung zu etablierten Instrumenten wie dem ADFC Fahrradklimatest [Ad20] oder dem ByPad-Verfahren [By20] dienen. So können durch die Sensoren Potenziale erschlossen werden, um das Radfahren schneller, komfortabler und sicherer zu machen und damit das Fahrrad in Konkurrenz mit anderen Verkehrsmitteln zu stärken. Die sensorbasierten Daten können die Radverkehrsförderung dabei unterstützen, weg von der verinselten Betrachtung einzelner Situationen, hin zu einer ganzheitlichen Betrachtung gesamter Routen oder Gebiete zu kommen. Die Daten aus einer sensorbasierten Erhebung des Radverkehrs stehen flächendeckend zur Verfügung und erzeugen damit ein zuvor nicht vorhandenes Gesamtbild einer Kommune. Dies zeigte sich in einem Projekt in Wiesbaden (vgl. [Sc14] und [Ju14]), bei welchem die abgefahrenen Routen aller Radfahrenden übereinandergelegt wurden. Das Ergebnis ist eine vollständige Straßenkarte von Wiesbaden. Auch die in Kampagnen von bikecitizen [Bi20] erhobenen GPS-Daten des Radverkehrs (Route, Geschwindigkeit, Verzögerungen etc.) erlauben eine Gesamtbetrachtung ganzer Städte. Die sensorbasierten Messungen des Radverkehrs ermöglichen eine langfristige Erfassung. So kann eine Erhebung zur Langzeitwirkung von innovativen Maßnahmen zur Förderung des Radverkehrs durch einen Vorher-Nachher-Vergleich erfolgen. Die bisher meist nur qualitativ vorliegenden Größen zum Komfort, Kraftbedarf und Geschwindigkeit

der Radfahrenden können durch Sensoren umfassender erhoben und quantifiziert werden. Durch die Quantifizierung werden die Informationen besser handhabbar für Entscheidungs- und Planungsprozesse gemacht. Im Ergebnis können die Daten eine Priorisierung von Maßnahmen im kommunalen Planungsprozess unterstützen. Dies hilft in der Breite die Qualität der Planung zu steigern.

Für den Einsatz sensorbasierter Messungen zur Entwicklung von Gesamtkonzepten für den Radverkehr ist der passende Zeitpunkt entscheidend. Gesamtkonzepte zur Radverkehrsförderung werden in den Kommunen nur selten (alle 10-15 Jahre) aufgestellt. Damit sind die Zeitfenster begrenzt, in denen die Kommunen die beschriebenen, umfassenden sensorbasierten Daten für die Aufstellung von Gesamtkonzepten benötigen. Entsprechend sind realisierte Anwendungsfälle des Einsatzes von Sensordaten bei Gesamtkonzepten der Radverkehrsförderung bisher eher selten.

4.2 Partizipation der Nutzer

Konzepte zur Förderung des Radverkehrs werden meist durch eine umfassende Beteiligung der Radfahrenden sowie der Öffentlichkeit begleitet. Dies ermöglicht die Belange der Radfahrenden in die Planungsprozesse einzubringen und insgesamt die Qualität der Radverkehrskonzepte zu steigern. In vielen Partizipationsprozessen ist jedoch zu beobachten, dass die Beteiligung nicht repräsentativ ist und einzelne organisierte Interessengruppen ein hohes Gewicht in den Diskussionsprozessen besitzen. Eine sensorbasierte Erhebung des Radverkehrs mittels Crowd-Sensing kann helfen zu gewährleisten, dass Partizipation nicht von wenigen Aktiven bestimmt wird, sondern dass alle Radfahrenden repräsentativ eingebunden werden können.

Die sensorbasierte Partizipation der Radfahrenden kann eine wichtige Rolle für das Fahrradklima spielen. Damit wird den Radfahrenden ermöglicht, eine direkte Rückmeldung an die professionellen Planer zu geben. Die täglichen Erfahrungen der Radfahrenden werden aufgezeichnet, dokumentiert und für die Partizipationsprozesse nutzbar gemacht. Die Messung aus Sicht der Radfahrer kann helfen, die häufig sehr emotionale Diskussion zum Radverkehr zu objektivieren bzw. bisher nur qualitativ beschriebene Eigenschaften zu quantifizieren. Die durch die Sensoren empirisch erhobenen Präferenzen der teilnehmenden Radfahrenden erhalten im Vergleich zu den individuellen Einschätzungen einzelner Akteure mehr Gewicht in den Partizipationsprozessen. Neben Sensoren ermöglichen Feedbackinstrumente (Feedbacktasten, Apps für Rückmeldungen etc.), dass die Radfahrenden eigenständig vor Ort mit geringem Aufwand positive oder negative Rückmeldungen geben können, ähnlich wie im Projekt GO Karlsruhe [Hä19] für die Zielgruppe Fußgänger.

Zahlreiche bisherige Projekte zur sensorbasierten Partizipation der Radfahrenden wie Kampagnen in Berlin [Ta18] oder Köln [Be19] sind von einzelnen engagierten Akteure die ein hohes Organisationsvermögen sowie technische Kompetenz haben abhängig. Daher sind sensorbasierte Erhebungsinstrumente erforderlich, die ein einfaches Participatory

Sensing ermöglichen. Für den Einsatz bei lokalen Messkampagnen muss das System auf Gruppen mit größeren Teilnehmerzahlen ausgelegt sein. Von Vorteil beim Participatory Sensing ist, dass die Radfahrenden sich durch den „Mitmach-Charakter“ eingebunden und mitgenommen fühlt. Eine weitere Anforderung ist, dass die gewonnenen Daten mit vergleichsweise geringem Aufwand grafisch gut aufbereitet und präsentiert werden können, damit diese für die breite Bevölkerung verständlich sind. In diesem Fall ein bestehender Bedarf aus Sicht der Praxis der Radverkehrsförderung für den Einsatz von Sensoren am Fahrrad.

4.3 Fahrradrouten und Radroutennetze

Die Entwicklung der baulichen Radverkehrsinfrastruktur (Radwege, Radfahrstreifen, Schutzstreifen, Fahrradstraßen, fahrradgerechte Knotenpunkte etc.) hat sich als ein Kernelement der Radverkehrsförderung bewährt. Dabei ist wichtig, das Augenmerk auf ein durchgängiges Gesamtnetz zu richten, nicht nur auf einzelne Streckenabschnitte. Nur wenn das vorhandene Radverkehrsnetz attraktiv gestaltet ist, kann die Radverkehrsförderung die volle Wirkung entfalten. Im Vordergrund steht die Frage, wie für Radfahrende attraktive und komfortable Routen identifiziert und definiert werden können. Der Einsatz von Sensoren kann sowohl bei der Neuplanung der Radverkehrsinfrastruktur, als auch der Aufwertung bestehender Radrouten erfolgen.

Sensorbasierte Erhebungen können die Entwicklung und Verbesserung von Radroutennetzen (hierarchisches Routennetz für verschiedene Nutzergruppen, Fahrradwegweisung etc.) oder einzelnen Radrouten (Schnellradwege, Stadtteilverbindungsrouten etc.) unterstützen und verbessern. Zum einem kann dafür aufgezeichnet werden, welche Routen die Radfahrenden gegenwärtig nutzen und welche Eigenschaften die bisher von den Radfahrenden präferierten Routen besitzen. Zum anderen bietet sich ein Vergleich verschiedener Routenalternativen an. Mit Hilfe einer integrierten Leistungsmessung am Fahrrad lässt sich ermitteln, welche Routen am leichtesten zu befahren sind. Zudem ist zu analysieren, wie hoch das Geschwindigkeitsniveau der Radfahrenden ist und wo es zu Zeitverlusten durch häufige und lange Wartezeiten bzw. vermeidbare Haltevorgänge kommt. Mit Hilfe des Erschütterungssensors können Unebenheiten im Belag verortet werden. Darüber hinaus sind Sensoren zur Klimamessung (Temperatur, Luftfeuchte, Beleuchtungsstärke) eine wichtige Ergänzung, um die Messungen in den richtigen Kontext einzuordnen. Schall- und Feinstaubsensoren zeigen die Belastung des Radfahrers durch Umwelteinflüsse. Auch eine Messung des subjektiven Stresses des einzelnen Radfahrenden ist denkbar, um kritische Stellen in der Infrastruktur zu finden und von Radfahrenden präferierte Routen zu identifizieren.

Erste Projekte von Röder [Rö20] und Hauenstein [Ha20] zur Entwicklung und Verbesserung von Fahrradrouten zeigen ein großes Potenzial für sensorbasierte Erhebungen. Im Rahmen dieser Projekte wurden für die geplante Verbindung zwischen zwei Orten jeweils verschiedene Routenalternativen mit SensorBikes erfasst. Im Vordergrund standen dabei Faktoren wie Geschwindigkeit, Reisezeit, Leistungsbedarf, Erschütterungen oder Bremsvorgänge.

Es wurden Daten zum objektiven Vergleich der Routenalternativen aus Perspektive der Radfahrenden erhoben. Dabei konnten auch Verbesserungsvorschläge zur Optimierung der Routen abgeleitet und quantifiziert werden. Mit vergleichbar geringem Erhebungsaufwand konnten durch die sensorbasierte Erhebung zusätzliche Informationen zur Planung von Radrouten gewonnen werden. In diesem Anwendungsfeld besteht auch zukünftig eine Nachfrage in der Praxis der Radverkehrsförderung.

4.4 Verkehrssicherheitsarbeit

Die Verkehrssicherheit ist ein wichtiger Schlüssel für die Förderung des Radverkehrs. Insbesondere Gelegenheitsradfahrer sowie potentielle neue Radfahrer nennen die Verkehrssicherheit als ein wesentliches Kriterium zur Nutzung bzw. Nichtnutzung des Fahrrades. Nur wenn sich diese auf dem Fahrrad sicher fühlen, werden Sie dies auch als Verkehrsmittel nutzen. Die Verkehrssicherheitsarbeit ist damit wesentlicher Bestandteil zahlreicher Konzepte zur Radverkehrsförderung sowie Kriterium bei der Planung der Radverkehrsinfrastruktur.

Sensorbasierte Erhebungen auf dem Fahrrad können die Verkehrssicherheitsarbeit durch zusätzliche Informationen unterstützen. Die Erhebungen eignen sich besonders für die vorsorgende Verkehrssicherheitsanalyse, bei der Bereiche mit negativen Einflüssen identifiziert werden, bevor es zu Unfällen kommt. Für die Analyse der Verkehrssicherheit bieten sich verschiedene Sensoren an. So können Sensoren die Bremsbeschleunigung sowie die Betätigung der Bremsen erfassen. Situationen mit vielen plötzlichen und starken Bremsvorgängen oder Ausweichmanövern von Radfahrenden deuten auf sicherheitsrelevante Konflikte hin. Zudem bietet sich die Messung des Abstandes zu überholenden oder parkenden Fahrzeugen an, um Konflikte zu erkennen. Darüber hinaus können Radfahrende durch eine Feedbacktaste melden, wann und wo sie sich gefährdet fühlen, wodurch sich diese in Ihrer Einschätzung ernst genommen fühlen. Weiterhin bieten sich Videoaufzeichnungen vorne und hinten an, um mögliche Verkehrskonflikte (Beinahe-Unfälle) zu erfassen. Durch die Erhebung von Konflikten im Radverkehr wird die Verkehrssicherheitsarbeit um Informationen erweitert, die über die Betrachtung der eher schlecht dokumentierten Unfälle mit Radfahrerbeteiligung hinausgehen. Die Messungen des Stresses von Radfahrenden (über Hautleitfähigkeit und Hauttemperatur) ermöglicht zudem das subjektive Sicherheitsgefühl zu erfassen. Die Ergebnisse der sensorbasierten Messung können helfen, Handlungsschwerpunkte für die Verbesserung der Verkehrssicherheit zu definieren bzw. bei einzelnen Standorten die Ursachen der Verkehrskonflikte aus Sicht der Radfahrenden zu verstehen.

Die SensorBikes wurden bereits in verschieden Projekten erfolgreich für einzelne Fragen der Verkehrssicherheitsarbeit eingesetzt. Die Untersuchungen ermöglichten aktuell diskutierte Fragen, wie objektive und subjektive Verkehrssicherheit von Schutzstreifen, oder die Mindestüberholabstände von Radfahrenden mit objektiv erhobenen Daten zu analysieren. Eine Studie zur subjektiven und objektiven Verkehrssicherheit konnte zeigen, dass auf einer Infrastruktur auf der Fahrbahn häufiger Stressmomente auftreten, als dies im Seitenraum der Fall ist, die subjektive Sicherheit jedoch auch vom Radfahrtyp abhängt [Me19]. Eine Studie

zum Bremsverhalten konnte helfen, kritische Standorte mit zahlreichen plötzlichen und kräftigen Bremsvorgängen, verursacht durch andere Verkehrsteilnehmer, zu identifizieren [Rö20]. Eine Studie zu Überholabständen auf innerörtlichen Hauptverkehrsstraßen zeigte, dass der Abstand zwar bei 50 % der Überholvorgänge unter 1,5 m, jedoch nur bei 10 % unter 1 m liegen [We20] und damit die Diskussion versachlichen. Die Ergebnisse der Projekte zu aktuellen Fragestellungen weisen eindrücklich nach, dass der Einsatz von Sensoren in diesem Themengebiet gewinnbringend ist und auch für weitere Projekte der Bedarf besteht.

4.5 Messung Umweltqualität

Die Umweltqualität beim Radfahren, wie die Luftqualität und die Lärmbelastung, tragen mit zum Komfort der Radfahrenden bei und sind damit ein weiteres Kriterium bei der Verkehrsmittelwahl. Zudem wird der Beitrag des Radverkehrs zur Förderung einer umweltfreundlichen Mobilität berücksichtigt.

Sensorbasierte Messungen werden genutzt, um zu analysieren, welchen Belastungen Radfahrende ausgesetzt sind. Zudem können Fahrräder als mobile Messstationen für die Erfassung der Umweltqualität dienen. Die Fahrräder können helfen, lineare oder auch flächenhafte Umweltbelastungen zu messen. Insbesondere bei Messungen, welche nicht durch den Einfluss eines Kfz beeinträchtigt werden sollen, bieten sich Fahrräder als „Sensorträger“ an. Denkbare Anwendungen für das Sensorbike sind die Erfassung des Stadtklimas, Erhebungen für den Naturschutz oder die Messung der Luftschadstoffbelastung.

Eine Fragestellung aus der Praxis der Radverkehrsförderung ist die Belastung von Babys durch Luftschadstoffe bei der Mitnahme auf dem Fahrrad bzw. im Fahrradanhänger. In der Studie wurde die Luftschadstoffbelastung von Babys bei der Mitnahme im Auto sowie dem Fahrrad verglichen und Maßnahmen zur Reduzierung der Belastungen aufgezeigt [SPK20]. Damit können Unsicherheiten junger Eltern reduziert werden. Die Messung der Umweltqualität durch Radfahrende ist als Fachthema durchaus gefragt und bietet vielfältige Anwendungsfelder für sensorbasierte Messungen. Allerdings ist die Einbindung in die kommunale Praxis der Radverkehrsförderung nur eingeschränkt möglich.

4.6 Zustandserfassung und Bewertung von Radverkehrsanlagen

Für den Komfort und den benötigten Kraftaufwand der Radfahrenden ist der Zustand der Fahrbahnoberfläche wichtig. Ungenügend unterhaltene Fahrbahnoberflächen zählen zu den häufigsten Meldungen von Radfahrenden an die Kommunen. Zudem haben die Straßenbaulastträger Interesse die Radverkehrsinfrastruktur systematisch zu erheben und zu erhalten.

Sensorbasierte Erhebungen durch das Fahrrad können die Zustandserfassung und Bewertung von Radverkehrsanlagen unterstützen. So findet die Zustandserfassung und Bewertung der

Radverkehrsanlagen bisher meist durch die Befahrung mit Kraftfahrzeugen statt, welche mit Kameras sowie Ebenheitsmessern ausgestattet sind. Jedoch sind nicht alle Wege mit den Testfahrzeugen befahrbar. Zudem besteht die Befürchtung von Radfahrenden, dass Ihre Perspektive durch die Messung mittels eines Kfz nicht angemessen berücksichtigt wird. Durch die sensorbasierten Erhebungen beim Radfahren kann der Zustand der Fahrbahnoberfläche mittels eines Erschütterungssensors und Kameraaufnahmen der Fahrbahn erfasst werden. Die Ergebnisse können damit auch als Ergänzung zum offiziellen Verfahren dienen.

Projekte mit dem SensorBike zur Messung der Erschütterungen durch die Fahrbahn haben gezeigt, dass eine zuverlässige Erkennung verschiedener Belagsarten sowie Fahrbahnzustände möglich ist [Fu19]. In diesem Rahmen konnten auch die Befürchtungen junger Eltern zu den Erschütterungen von Babys bei der Mitnahme mit dem Fahrrad eingeordnet werden [AHV20]. Erste Modellprojekte für eine sensorbasierte Zustandserfassung und Bewertung mit dem Fahrrad wurden in Brandenburg bereits durchgeführt [La16]. Jedoch gibt es noch offene Fragen und Bedenken bei der Zustandserfassung durch Fahrräder. Beispielsweise stellen sich die Fragen, welche Erschütterungswerte für die Zustandsbewertung relevant sind oder wie sich das Fahrverhalten und die Fahrtlinie der Radfahrenden auf die Ergebnisse auswirken. Für den Einsatz von sensorbasierten Messungen zur Erfassung des Fahrbahnzustandes müssen daher Standards entwickelt werden, um die hohen Qualitätsanforderungen der Straßenbaulastträger zu erfüllen und eine Vergleichbarkeit der Ergebnisse zu gewährleisten. Ein Bedarf aus der Praxis der Radverkehrsförderung für den Einsatz der sensorbasierten Zustandserfassung und Bewertung mit Hilfe von Fahrrädern besteht.

5 Herausforderungen für die sensorbasierte Messung des Radverkehrs und dessen Skalierung

Eine ausreichende Teilnehmeranzahl ist für belastbare Ergebnisse von Studien von äußerster Wichtigkeit. Einige der vorgestellten Studien wurden als Vertiefungsstudien durchgeführt, um den gewählten Ansatz in der Messmethodik und dessen Umsetzbarkeit zu prüfen. In einem nächsten Schritt wird nun die Skalierung hin zu einem Crowd-Sensing Ansatz und die damit verbundenen Herausforderungen beim Sammeln, Aggregieren und Auswerten der Daten diskutiert.

Eine Herausforderung für die Umsetzung von Crowd-Sensing-Ansätzen ist die Gewinnung ausreichender Teilnehmerzahlen. Solche Messkampagnen sind nur in seltenen Fällen Selbstläufer mit hohen Teilnehmerzahlen. Die Möglichkeit zu einer guten Datengrundlage beizutragen ist selten eine ausreichende Motivation für die Teilnahme zahlreicher Nutzerinnen und Nutzer. In den meisten Fällen ist vielmehr erforderlich, die Crowd-Sensing-Ansätze in umfangreiche Kampagnen der Öffentlichkeitsarbeit einzubinden, um ausreichende Nutzerzahlen zu gewinnen. Zudem bietet sich der Einsatz von Multiplikatoren an. Der Aufwand für das Akquirieren der Teilnehmerinnen und Teilnehmer ist bei der Planung solcher Messungen zu berücksichtigen.

Das primäre Ziel der Skalierung ist die Einbindung möglichst vieler Teilnehmer und die zeitgleiche Einbeziehung mehrerer Standorte in eine Studie. Dabei sind auch eine hohe Datenqualität und die Zuverlässigkeit der Ergebnisse zu gewährleisten. Durch die Erhöhung der Teilnehmerzahlen zeigt sich bei Crowd-Sensing-Studien zudem die Herausforderung des Umgangs mit großen Datenmengen. So liegen eine Vielzahl von Datenpunkten vor, die mit Hilfe von vorliegenden Orts- und Zeitangaben zu Aussagen für bestimmte Streckenabschnitte bzw. Zeiträume aggregiert werden müssen. Die Daten der verschiedenen Nutzer sind zusammenzuführen und zu verschneiden. Die aktuell noch sehr anspruchsvolle und zeitaufwendige Datenaufbereitung sollte in Zukunft mit interdisziplinärer Hilfe aus dem Bereich Informatik vereinfacht werden. Für das Datenmanagement erscheint daher die Entwicklung passgenauer Software sehr erstrebenswert. Diese sollte Funktionen zur automatischen Verschneidung der Daten sowie zur Fehlersuche aber auch zur Ausgabe von kumulierten Ergebnissen beinhalten. Dabei müssen die Instrumente für die Auswertung großer Datenmengen auch für technisch weniger versierte Nutzer verwendbar sein. Aus diesem Grund sollten die eingesetzten Modelle einfach, verständlich und transparent sein. So kann auch die Akzeptanz der Praxisakteure gegenüber den Ergebnissen deutlich erhöht werden und zu einer verbesserten Radverkehrsförderung beitragen.

Bei der Planung ist sehr genau abzuwägen, welche Daten und welcher Umfang von Daten für die Studie erforderlich sind. Dabei sollten die Daten bereits eine grundlegende Filterung erfahren, um die Datenqualität zu erhöhen. Zudem ist die Datenrate auf das, für den Messzweck erforderliche Minimum zu reduzieren, um die zu verarbeitende Datenmenge zu begrenzen.

Als eine Alternative zu Massendatenerhebungen mit hohen Teilnehmerzahlen bietet sich bei einigen Fragestellungen die Durchführung von Messkampagnen mit kleinen Teilnehmerzahlen und einem durchdachten Messdesign an. So konnte ein Teil der bisherigen Projekte aufgrund der aufwendigen Ausstattung und des hohen Betreuungsaufwandes nur in Vertiefungsstudien mit einer geringeren Anzahl von Probanden durchgeführt werden. Bei einigen Fragestellungen kann durch die Entwicklung von einfachen und duplizierbaren Sensorsystemen die Anzahl der Probanden erhöht werden. Andere Fragestellungen werden jedoch aufgrund der aufwendigen Ausstattung oder der Notwendigkeit der Kontrollierbarkeit der Rahmenbedingungen auch in Zukunft nur mit geringen Teilnehmerzahlen durchführen lassen.

Gerade auch Datenschutz und Datensicherheit spielen für das Crowd-Sensing beim Radverkehr eine wichtige Rolle. Vorgehensweisen und Erkenntnisse aus dem Datenschutz sind für die spezifischen Rahmenbedingungen weiterzuentwickeln. So sind technische Schutzmechanismen von Verschlüsselung bis Anonymisierung bzw. Pseudonymisierung vorzusehen.

6 Fazit

Die Aufzählung der möglichen Einsatzbereiche von sensorbasierten Messungen des Radverkehrs verdeutlicht, dass viele Ansätze bereits in Einzelstudien erprobt wurden. Die bisher erfolgversprechendsten Ansätze für die Unterstützung der Praxis der Radverkehrsförderung wurden dabei in klar umrissenen Anwendungsfeldern, wie der Zustandserfassung und Bewertung von Radverkehrsanlagen, der Verkehrssicherheitsarbeit oder der Planung von Radrouten erzielt. In diesen Fällen können die sensorbasierten Messungen helfen, bestehende Fragen der Radverkehrsförderung zu adressieren. Der Einsatz sensorbasierter Systeme bei generellen Fragestellungen wie der Entwicklung übergeordneter Radverkehrskonzepte steht bisher noch vor größeren Herausforderungen. Der Dialog zwischen den Akteuren der Radverkehrsförderungen sowie den Akteuren der Messkampagnen muss gestärkt werden. Die sensorbasierten Messungen des Radverkehrs sollten sich verstärkt an dem Informationsbedarf der Akteure der Radverkehrsförderung orientieren.

Wenn die genannten Herausforderungen gemeinsam mit den Akteuren aus dem Bereich der Informatik / Messtechnik sowie der Radverkehrsförderung gemeistert werden, bietet die sensorbasierte Messung des Radverkehrs große, bisher ungenutzte Potentiale zur Förderung des Radverkehrs.

Literaturverzeichnis

[Ad20] ADFC Fahrradklima-Test, `https://fahrradklima-test.adfc.de/`, Stand: 14.08.2020.

[AHV20] Andriof, Benjamin; Haffelder, Silvio und Vonnieda, Leon: SensorBike: Radfahren mit Baby – Die Belastung von Babys durch Erschütterungen und Luftschadstoffen – Erschütterungen am Fahrrad und Auto. Hochschule Karlsruhe – Technik und Wirtschaft, Projekt von Studierenden 2020.

[Be19] Beutelspacher, Klaus: Abstand Messen Köln. www.abstand-messen-koeln.jimdofree.com/, Stand: 14.08.2020.

[Bi20] bikecitizens, www.bikecitizens.net/de/neues-datenanalysetool-fuer-die-radverkehrsplanung/, Stand: 14.08.2020.

[BS19] Becker,Udo; Schlag, Bernhard et al.: RadVerS - Mit Smartphones generierte Verhaltensdaten im Verkehr - Differenzierug des Nutzerverhaltens unterschiedlicher RadfahrerInnengruppen. TU Dresden, 2019.

[By20] BYPAD – Bicycle Policy Audit, www.bypad.org/, Stand: 14.08.2020.

[EH19] Eichner, Rick und Hauenstein, Jan: Sensorbike – Erhebungsmethodik. Hochschule Karlsruhe – Technik und Wirtschaft, Projekt von Studierenden 2019.

[Fu19] Furtado, Alexandre Thomaz: SensorBike: Evaluation of Bicycle Ride Comfort – According to ISO 2631-1. Hochschule Karlsruhe – Technik und Wirtschaft, Projekt von Studierenden 2019.

[Ha20] Hauenstein, Jan: SensorBike: Radroutenplanung aus Sicht der Radfahrer – Am Beispiel Karlsruhe/Ettlingen. Hochschule Karlsruhe – Technik und Wirtschaft, Bachelorthesis 2020.

[Hä19] Häußler, Elke et al.: Den Fußverkehr in Städten fördern: Partizipative Forschung im Reallabor GO Karlsruhe. GAIA – Ecological Perspectives for Science and Society, Band 28, Nummer 4, S. 396-397, 2019.

[Ju14] Jung, Hendrik: Rad ohne Wege – Weiter Weg zur Fahrradstadt Wiesbaden. Sensor Wiesbaden, 01.05.2014, www.sensor-wiesbaden.de/rad-ohne-wege-weiter-weg-zur-fahrradstadt-wiesbaden/, Stand:15.08.2020.

[Kn95] Knoflacher, Hermann: Fußgeher- und Fahrradverkehr: Planungsprinzipien. Böhlau Verlag, Wien, 1995.

[La16] Land Brandenburg, Verkehrsministerium: Auftakt der Zustandserfassung von Radwegen. Pressemitteilumg vom 29.08.2016, www.mil.brandenburg.de/cms/detail.php/bb1.c.456982.de, Stand: 15.08.2020.

[Me19] Merk, Jule: Vergleich der objektiven Verkehrssicherheit und des subjektiven Verkehrsstresses bei Schutzstreifen und Radfahrstreifen im Vergleich zu eigenständigen Radwegen. Hochschule Karlsruhe – Technik und Wirtschaft, Masterthesis 2019.

[MM90] Monheim, Heiner; Monheim-Dandorfer, Rita: Straßen für alle-Analysen und Konzepte zum Stadtverkehr der Zukunft. Hamburg, 1990.

[Rö20] Röder, Annika: Optimierung und Vergleich regionaler Radrouten unter Einsatz des SensorBikes – am Streckenbeispiel Karlsruhe – Weingarten (Baden). Hochschule Karlsruhe – Technik und Wirtschaft, Bachelorthesis 2020.

[Sc14] Scholz&Volkmer GmbH: Radwende – Der Radweg ist das Ziel. Projekt 2014, www.s-v.de/de/produkte/radwende/, Stand:15.08.2020.

[SPK20] Sommer, Anabelle; Prinzing, Benedikt und König, Björn: Abschlussbericht - Messung der Feinstaubbelastungen von Babys beim Radfahren und Autofahren. Hochschule Karlsruhe – Technik und Wirtschaft, Projekt von Studierenden 2020.

[St13] Stadt Karlsruhe, Stadtplanungsamt: Radverkehr – 20-Punkte-Programm. Zwischenstand und Fortschreibung des 20-Punkte-Programms zur Förderung des Radverkehrs in Karlsruhe. 2013.

[Ta18] Tagesspiegel Online: Radmesser: Die genaue Methode – Wie wurden die Überholabstände gemessen? www.tagesspiegel.de/gesellschaft/medien/radmesser-die-genaue-methode-wie-wurden-die-ueberholabstaende-gemessen/23710682.html, Stand: 14.08.2020.

[Vi20] Vialytics GmbH, www.vialytics.de/, Stand: 14.08.2020.

[We20] Welz, Christoph: Erhebung und Analyse des Überholabstands vom motorisierten Individualverkehr zu Radverkehr auf Stadtstraßen – am Beispiel der Stadt Karlsruhe. Hochschule Karlsruhe – Technik und Wirtschaft, Bachelorthesis 2020.

R. Reussner, A. Koziolek, R. Heinrich (Hrsg.): INFORMATIK 2020,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2021 1151

Verbindungsaufbau zu mobilen Public Displays

Anforderungen und Datenfluss

Christian Rickert[1], Thomas Schlegel[2]

Abstract: Public Displays werden bereits in vielen Städten eingesetzt, zum Beispiel als elektronische Fahrplanauskünfte, Infoboards in Innenstädten oder Selbst-Check-In Stehlen an Flughäfen. Wenige dieser Public Displays bieten einen Verbindungsaufbau zum eigenen Smartphone an. Im Fokus steht die Übertragung der eigenen Strecke auf das Public Display. Zusätzlich wird auf die Reisebegleitung und den Kommunikationsfluss vom Start bis zum Ziel eingegangen. Es wird erläutert, was mit den Daten während einer Reise passiert und welche Rolle die Plattform zur Routenberechnung einnimmt. Anhand von Anforderungen, die durch Personas und Szenarien ermittelt werden, wird ein erster Prototyp in Form einer Smartphone-Anwendung entwickelt, der den Verbindungsaufbau simulieren soll. Dadurch können erste Aussagen darüber getroffen werden, welche Ansprüche an solch ein System gestellt werden müssen.

Keywords: Public Displays; Bluetooth; QR; W-LAN; Android; Smartphone; SmartWindow; öffentlicher Verkehr; Straßenbahn

1 Einleitung

Reisebegleitungen gibt es in vielen verschiedenen Formen, vom Navigationssystem im Fahrzeug bis hin zu verschiedenen Apps auf dem Smartphone. Diese Anwendungen wollen dem Nutzer die Reise vom Start zum Ziel erleichtern. Oftmals wird dabei zu Beginn der Reise ein Ziel angegeben und während der Reise wenig mit dem System interagiert. Gerade bei der Nutzung des öffentlichen Verkehrs begrenzt sich die Nutzung von Reisebegleitungen auf die Suche nach der nächstbesten Verbindung zum Zielort. Eine aktive und mobile Reisebegleitung zielt darauf ab, den Nutzer auf der gesamten Strecke zu unterstützen, indem sie ihm im Falle einer Störung oder Ausfalls gezielte Informationen zukommen lässt, die die Reise vereinfachen. Diese Unterstützung erhält er durch sein Smartphone und ein neues, innovatives System. Im Rahmen des Projektes „SmartMMI“ (Modell – und Kontextbasiert Mobilitätsinformation auf smart Public Displays und Mobilgeräten im öffentlichen Verkehr) werden sogenannten „SmartWindows“ in den Bahnen des öffentlichen Verkehrs eingebaut. Das intelligente Fenster soll dem Nutzer während der Fahrt in der Straßenbahn zur Seite

[1] Karlsruhe University of Applied Sciences, Moltkestrasse 30, 76133 Karlsruhe, Germany, chrisrickert27@gmail.com | https://www.hs-karlsruhe.de/
[2] Institute of Ubiquitous Mobility Systems (IUMS), Karlsruhe University of Applied Sciences, Moltkestrasse 30, 76133 Karlsruhe, Germany, iums@hs-karlsruhe.de | http://iums.eu

doi:10.18420/inf2020_107

stehen und mit nützlichen und interessanten Informationen versorgen. In diesem Paper werden Anforderungen vorgestellt, wie dieses System möglichst nutzerfreundlich und verlässlich funktioniert. Am Ende wird diskutiert, wie die Anforderungen erfüllt werden können und was notwendig zur optimalen Nutzung des Systems ist.

2 Bestehende Systeme

Zu Beginn folgt ein kurzer Überblick über Public Displays und digitale Fahrgastinformationssysteme.

Dynamische Fahrgastinformationssysteme gibt es in verschiedenen Formen. Viergutz [VI15] beschreibt die Anforderungen an unterschiedliche dynamische Fahrgastinformationssysteme. Zu diesen Systemen zählen Anzeigen an Haltestellen, Abfahrtsmonitore an öffentlichen Plätzen und mobile Anwendungen. Anzeigen an Haltestellen sowie Abfahrtsmonitore sind an die Fahrgäste gerichtet und zeigen allgemeine Informationen an. Die Anforderungen an eine Anzeige an einer Haltestelle sind unter anderem aktuelle, zeitgenaue Informationen über die Abfahrtszeitpunkte. Die Angabe über die noch verbleibende Zeit bis zur Abfahrt eignet sich hier. Abfahrtsmonitore haben ähnliche Anforderungen, haben jedoch den Vorteil, dass sie auch in Geschäften oder Unternehmen platziert werden. Abfahrtsmonitore werden heutzutage auch als Public Displays bezeichnet und bieten eine grafische Oberfläche. Eine Nutzerinteraktion ist nicht zwingend notwendig, kann aber in Betracht auf den Standort des Displays sinnvoll sein. Im Gegensatz dazu sind mobile Anwendungen individueller. Die Abfahrtszeiten werden nicht als Countdown, sondern meist als Uhrzeit angegeben.

GeoSignage [GE19] entwickeln Managementsysteme für Infotainment. Die Systeme bestehen aus einem Interface als Web-Client und einer Anwendung für unterwegs. Sie erhält standortbezogene Daten und kann gezielte Werbung oder Anzeigen auf den Bildschirmen abspielen lassen. Zusätzlich gibt es einen W-LAN Zugang, den die Nutzer mit ihrem Smartphone nutzen können, um sich die Werbung selbst auszusuchen.

CHK [IN19] bauen interaktive Public Displays an Haltestellen ein. Nutzer können über einen Touchscreen die nächsten Abfahrtszeiten abrufen und individuell anzeigen lassen. Weiterhin können Informationen über alle Haltestellen in der Umgebung angezeigt werden und Nutzer können ihre eigene Route an den Bildschirmen planen.

3 Anforderungsanalyse

3.1 Interaktionszyklus

Der Interaktionszyklus beschreibt in drei Phasen, wie die Nutzung des SmartWindows abläuft. Phase 1 beschreibt den Verbindungsaufbau, sobald der Nutzer neben einem Fenster sitzt. Phase 2 handelt von der Übertragung der Route und allen Vorgängen, während der Nutzer daneben sitzt. Die Vorgänge während der Trennung der Verbindung werden in Phase 3 betrachtet (Abb. 1).

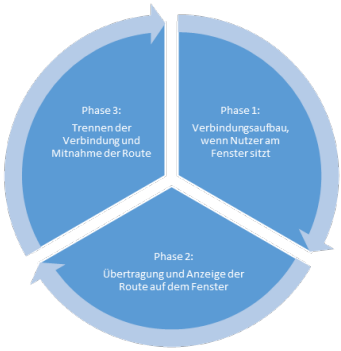


Abb. 1: Interaktionszyklus

Den jeweiligen Phasen werden funktionale und nichtfunktionale Anforderungen zugeordnet. Nach der späteren Vorstellung des Prototyps wird über die erfüllten Anforderungen diskutiert. Die Anforderungen werden aus Szenarios gewonnen, in den unterschiedliche Personas eingebunden werden. Durch die unterschiedlichen Situationen und die erwarteten Funktionen des Systems können eine Vielzahl an Anforderungen gewonnen werden.

3.2 Funktionale und nichtfunktionale Anforderungen

Funktionale Anforderungen beschreiben eine konkrete und erwartete Funktion eines Systems, zum Beispiel ein automatischer Verbindungsaufbau. Sie sind Funktionen, die das System erfüllen soll. Nichtfunktionale Anforderungen sind auch „Qualitätsanforderungen“[PR15]. Sie beschreiben Anforderungen, die nicht von funktionalen Anforderungen abgedeckt werden, zum Beispiel wie zuverlässig oder bedienbar ein System sein soll. Anforderungen werden an Phase 1 gestellt (Tab. 1):

Tab. 1: Anforderungen der Phase 1

Typ	**Phase 1: Verbindungsaufbau**
Funktional 1.1	Das Smartphone muss erkennen, dass sich ein SmartWindow in der Nähe befindet
Funktional 1.2	Das Smartphone sollte um Bestätigung fragen, bevor eine Verbindung eingegangen wird
Funktional 1.3	Das SmartWindow sollte pro Fensterseite nur eine Verbindung gleichzeitig herstellen.
Funktional 1.4	Das SmartWindow muss eine Möglichkeit zur Verfügung stellen, eine mitgebrachte Route entgegenzunehmen
Nichtfunktional 1.1	Das System sollte mehrere Sprachen zur Auswahl bieten

Die Anforderungen wurden durch Personas und Szenarien bestimmt. Die Personas wurden aus der Anforderungsspezifikation des Projekts „SmartMMI“ [SC18] übernommen. Phase 2 besitzt folgende Anforderungen (Tab. 2):

Tab. 2: Anforderungen der Phase 2

Typ	**Phase 2: Übertragung und Anzeige**
Funktional 2.1	Das System muss die Daten des SmartWindows mit dem Smartphone synchronisieren
Funktional 2.2	Das System muss den Nutzer über Unfälle o.Ä. auf SmartWindow und Smartphone benachrichtigen
Funktional 2.3	Das System sollte den Nutzer informieren, wenn die Verbindung vorzeitig getrennt wurde
Funktional 2.4	Das System sollte bei einem ungewollten Verbindungsabbruch erneute einen Verbindungsaufbau durchführen
Nichtfunktional 2.1	Das System muss eine stabile Verbindung während der Fahrt aufrechthalten.
Nichtfunktional 2.2	Das System sollte einfach bedienbar sein

Die Phase 2 beschreibt die Nutzung des SmartWindows und welche Daten dem Nutzer angezeigt bzw. gesendet werden sollen. Verbindungsabbrüche müssen selbstständig behoben werden. Auch Echtzeitdaten, die am Bildschirm angezeigt werden, müssen mit dem Smartphone synchronisiert werden, damit keine Diskrepanz entsteht und der Nutzer nicht unnötig verwirrt wird. In Phase 3 spielen folgende Anforderungen eine Rolle (Tab. 3):

Tab. 3: Anforderungen an Phase 3

Typ	**Phase 3: Trennung und Verlassen der Bahn**
Funktional 3.1	Das System muss dem Nutzer die Möglichkeit bieten, seine Route auf das Smartphone zu übertragen
Funktional 3.2	Das System sollte erkennen, dass der Nutzer den Sitzplatz verlassen hat
Funktional 3.3	Das SmartWindow muss bei Verbindungstrennung vorhandene Daten löschen
Nichtfunktional 3.1	Das System sollte bei Verbindungstrennung keine Daten verlieren

Die Anforderungen der Phase 3 werden an ein problemloses Verlassen des Sitzplatzes gerichtet. Hierbei kann es passieren, dass der Nutzer seinen Platz plötzlich verlässt. In diesem Fall sollten keine Daten verloren gehen und die Route auf sein Smartphone übertragen werden, sodass er die Navigation nutzen kann, sobald er an der Haltestelle steht.

4 Aufbau des Systems

4.1 Aktuelle Situation

Die Anforderungsspezifikation des Projekts „SmartMMI“ gewähren einen Einblick in die Bestandteile des aktuellen Systems (Abb. 2).

Zunächst werden die Verbindungen zwischen Plattform und Smartphone sowie Plattform und Fahrzeug(-antenne) genauer betrachtet. Die Plattform ist zuständig für die Reisebegleitung bzw. die Routenberechnung, als auch die Mitteilung der Echtzeitdaten wie Störungen oder Fahrtausfälle. Der Nutzer kann die Reisebegleitung unabhängig vom SmartWindow nutzen. Dafür benötigt er lediglich eine Smartphoneanwendung. Aktuell ist der Nutzer in der Lage, durch das Einscannen eines QR-Codes am SmartWindow eine Strecke, die auf dem Fenster läuft, mitzunehmen. Mit diesem Beitrag soll untersucht werden, wie das „Mitbringen“ eines Weges ermöglicht werden kann, damit die Reisebegleitung vom Start zum Ziel durchgängig ist. Damit ist gemeint, wie die eigene, mitgebrachte Route auf verlässliche Weise auf das SmartWindow übertragen werden kann.

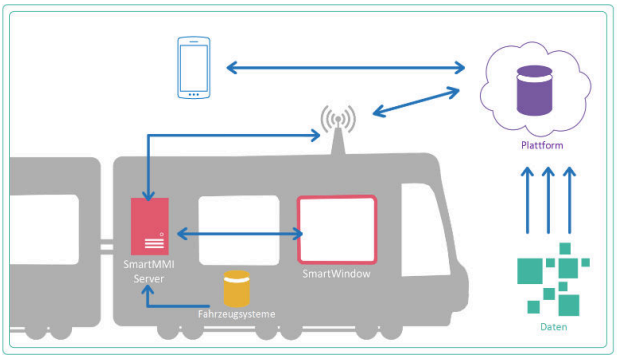


Abb. 2: Aufbau des Systems

4.2 Neue Situation

Das Einbauen der Möglichkeit, eine Strecke auf das SmartWindow zu übertragen, kann über viele Wege funktionieren. Dazu zählen Bluetooth, QR-Code, W-LAN und mobile Daten, da diese Grundfunktionen von Smartphones sind. NFC wird nicht in Betracht gezogen, da zur Bearbeitungszeit keine Möglichkeit besteht, in Bahn oder Fenster einen NFC-Chip einzubauen. Die geringe Reichweite und Datenübertragungsrate stellen ein potenzielles Problem dar. Obwohl eine aktiv-aktiv Verbindung über NFC möglich ist, ist NFC nur brauchbar, wenn die Verbindung nicht unterbrochen wird. Falls in Zukunft die Möglichkeit besteht, NFC-Chips einzubauen, könnten die Armlehnen in Erwägung gezogen werden, da dort das Smartphone dauerhaft platziert werden kann.

- Bluetooth

  Die Herausforderungen bei der Nutzung von Bluetooth sind die Reichweite und die Anzahl der verbundenen bzw. gekoppelten Geräte. Ohne Restriktionen spielt es keine Rolle, ob der Nutzer neben dem Fenster oder am anderen Ende des Abteils steht. Er ist in der Lage eine Verbindung mit dem Fenster herzustellen, auch wenn er nicht direkt am Fenster sitzt. Die Nutzung soll aber nur für Nutzer neben dem Fenster möglich sein. Deshalb muss bei der Verwendung von Bluetooth darauf geachtet werden, dass nur ein Nutzer gleichzeitig eine Verbindung je Fensterhälfte eingehen kann. Dies ist möglich, wenn ihm bei der Anmeldung an seiner Fensterseite das SmartWindow als verfügbares Gerät angezeigt wird. Beispielsweise kann das Fenster erst gefunden werden, sobald der Nutzer aktiv einen Button drückt. Bei der Trennung muss auch erkannt werden, ob der Nutzer tatsächlich die Bahn verlassen hat, oder ob die Verbindung vorzeitig durch ein technisches Problem unterbrochen wurde. Der Vorteil von Bluetooth liegt bei der bidirektionalen Kommunikation zwischen den Geräten.

- QR-Code

  Da ein QR-Code bereits beim Mitnehmen der Strecke eingesetzt wird, bietet es sich an, diesen auch beim Übertragen der Strecke auf das SmartWindow zu verwenden. Ein QR-Code kann eine URI oder ein Link zu einer Homepage sein. Im Falle einer URI können darin die notwendigen Daten zur Nutzung der Routenberechnung kodiert werden. Jedoch besteht bei der Nutzung des QR-Codes keine direkte Verbindung zwischen Smartphone und SmartWindow, es werden lediglich Daten übertragen, wenn der Code gelesen wird. Zusätzlich muss er vor dem Verlassen des Sitzplatzes frühzeitig daran denken, seinen Weg mitzunehmen, da das Fenster kein Wissen darüber hat, ob er seinen Sitzplatz verlassen hat. Zwar könnten die Daten über das mobile Netz synchronisiert werden, aber dafür muss das Fenster erkennen, welcher Nutzer gerade am Fenster saß.

- W-LAN

  Die Nutzung von W-LAN hat ähnliche Voraussetzung wie Bluetooth. Es muss sichergestellt werden, dass der richtige Nutzer am Fenster angemeldet wird und das diese Verbindung nicht von einem anderen Nutzer im Netzwerk unterbrochen werden kann. Auch hier spielen Reichweite und Anmeldung wieder eine Rolle. Dem Nutzer könnte ein dynamisch generiertes Passwort angezeigt werden, mit dem er sich im W-LAN Netzwerk anmelden kann. Das Verlassen des Fensters ist weniger problematisch, da bei einer zu großen Entfernung die Verbindung automatisch abbricht. Nach Möglichkeit sollten keine Daten bei einer automatischen Trennung verloren gehen, sondern vorher bereits mit dem Smartphone synchronisiert werden.

- Mobile Daten

  Anders als bei den bereits genannten Technologien wird bei der Nutzung der mobilen Daten keine direkte Verbindung zum Fenster eingegangen, sondern der Plattform

mitgeteilt, dass der Nutzer sich jetzt am Fenster befindet. Die Authentizität kann durch einen Code geprüft werden, der am Fenster generiert wird und den der Nutzer in der Smartphoneanwendung eingeben muss. Das Verlassen kann durch den Abgleich des Standortes geschehen. Eine mobile Datenverbindung geht mit einer stabilen Datenverbindung einher.

5 Konzept

In diesem Kapitel wird das Konzept und die Designentscheidungen vorgestellt sowie ein potenzieller Kommunikations- und Datenfluss für eine mobile Reisebegleitung. Für die Kommunikation in beide Richtungen, bzw. die Übertragung von Echtzeitdaten sind W-LAN und Bluetooth von Vorteil, da sie nicht abhängig vom Mobilfunknetz und der Verbindungsqualität sind. Zusätzlich ist Bluetooth in jedem Smartphone verbaut [BL19]. QR-Code und mobile Daten sind dagegen simpler bei der Gestaltung und Anwendung, verlangen jedoch mehr Einwirken durch den Nutzer.

Das Konzept handelt vom Verbindungsaufbau via QR-Code. Für den Nutzer sieht es nach einer Verbindung aus, jedoch ist es lediglich ein Datenaustausch. Sobald der Nutzer die Bahn betritt und sich an die jeweilige Fensterseite setzt, wird ihm die Möglichkeit geboten, sich am Fenster anzumelden. Was der Nutzer nicht weiß, ist, er stellt keine Verbindung mit dem Fenster her, sondern der Plattform wird mitgeteilt, dass sich in dieser Bahn an diesem Sitzplatz ein Nutzer befindet, der die Funktionen des SmartWindow nutzen möchte. Das Anmelden funktioniert über einen QR-Code, den der Nutzer mit seinem Smartphone und vorzugsweise in der Smartphoneanwendung einliest. Dieser QR-Code enthält Daten über das Fenster und die Bahn, wie zum Beispiel Richtung und Linie der Bahn, eine eindeutige Nummer des SmartWindows und auf welcher Seite des Fensters der Nutzer sitzt. Hat der Nutzer diesen QR-Code eingelesen, folgt der nächste Schritt. Der Plattform müssen alle nötigen Informationen mitgeteilt werden. Über die Smartphoneanwendung schickt der Nutzer diese Daten nun an die Plattform. Zusätzlich gibt er noch Informationen über sich selbst mit: Seinen Nutzernamen bzw. ID und seinen Start- und Zielort. Sobald diese Daten bei der Plattform ankommen, werden die Daten verarbeitet. Die Plattform ordnet die Strecke dem Fenster zu damit sie dort angezeigt werden kann. Weiterhin wird während der Fahrt in Echtzeit Ausfälle und Störung einbezogen. Für den Fall, dass auf der Strecke des Nutzers eine Störung erwartet wird, kann das System dies bei der Berechnung berücksichtigen und dem Nutzer eine alternative Route vorschlagen oder ihn zum Umstieg raten. Die Mitnahme des Weges funktioniert dann als Absicherung. Unter der Annahme, dass das System in der Bahn eine bessere Verbindung zum Server hat, kann durch das manuelle Mitnehmen der Strecke dafür gesorgt werden, dass keine Daten verloren gehen.

5.1 Kommunikationsfluss

Im folgenden Kapitel wird diskutiert, wie die einzelnen Komponenten miteinander kommunizieren und der Kommunikationsfluss entsteht.

- Datenaustausch zwischen Smartphone und Plattform

  Die Reisebegleitung beginnt, sobald der Nutzer sich entscheidet, eine Reise anzutreten. Auf dem Smartphone hat er die zugehörige Anwendung mit folgenden, persönlichen Daten:

  - Nutzerdaten (Name, E-Mail, Identifikation)
  - Standort
  - Zielort
  - Abgespeicherte Orte
  - Favorisierte Ziele
  - Favorisierte Abfahrtszeiten
  - Favorisierte Linien

  Wenn der Nutzer sein Ziel auswählt, beginnt der erste Schritt des Datenaustauschs. An die Plattform werden sein aktueller Standort, sein Ziel und seine bevorzugten Linien gesendet. Die Plattform berechnet die Strecke und ungefähre Dauer der Reise ein erstes Mal und sendet die Daten an das Smartphone des Nutzers. Gleichzeitig wird eine aktuelle Session erzeugt, die eine genaue Identifikation erhält. Diese Identifikation wird in den nächsten Schritten gebraucht. Der Nutzer erhält zunächst die voraussichtliche Route und beginnt seine Reise. Im Hintergrund berechnet die Plattform die Route periodisch immer wieder neu, um im Falle von Änderungen des Nutzerverhaltens, der Verkehrssituation, o.Ä. weiterhin eine korrekte Route liefern zu können.

- Datenaustausch zwischen Smartphone und SmartWindow

  Sobald der Nutzer eine Straßenbahn betritt und sich neben ein SmartWindow setzt, wird eine Verbindung hergestellt. Folgende Daten werden vom Smartphone auf das Fenster übertragen:

  - Sitzplatz (Rechte oder linke Seite des Fensters)
  - Nutzerdaten
  - Sessionnummer

Der Sitzplatz spielt eine Rolle, da das SmartWindow gleichzeitig von zwei Personen bedient werden kann, die am Fenster sitzen. Das SmartWindow befindet sich immer an einer Sitzgruppe, bestehend aus vier Sitzen, in der sich die Personen gegenübersitzen.

Das SmartWindow benötigt noch weitere Daten, bevor die Reisebegleitung am Fenster angezeigt wird. Vom Fahrzeugsystem benötigt das SmartWindow folgende Daten:

- Linie der Bahn
- Richtung der Bahn
- Aktueller Standort auf der Linie
- Standort des Fensters in der Bahn

- Datenaustausch zwischen SmartWindow und Plattform

  Damit während der Fahrt die Reisebegleitung läuft, müssen nun alle Daten an die Plattform gesendet werden. Die Plattform berechnet periodisch die Reise und aktualisiert diese auf dem SmartWindow. Folgende Daten werden vom SmartWindow an die Plattform gesendet:

  - Sessionnummer
  - Nutzerdaten
  - Linie der Bahn
  - Richtung der Bahn
  - Aktueller Standort auf der Linie
  - Zielort

  Die Sessionnummer wird mit dem Smartphone abgeglichen und die Route wird auf dem SmartWindow angezeigt. Der Nutzer kann mit dem SmartWindow interagieren und nach Belieben die Route ändern. Entscheidet er sich, ein neues Ziel auszuwählen, wird erneute eine Anfrage mit den aktualisierten Daten an die Plattform gesendet und die Strecke berechnet.

- Individuelle und allgemeine Informationen

  Da die Informationen auf dem SmartWindow von allen Passagieren einsehbar sind, sollten sensible Daten, wie Ziel und Start nicht direkt angezeigt werden. Das SmartWindow soll als Ergänzung des Smartphones dienen und nicht als redundante Anzeige von Daten, die der Nutzer auch auf dem Smartphone sieht. Die am SmartWindow angezeigte Strecke kann auf den Teil der Reise beschränkt werden, der in der Bahn abläuft. Falls der Nutzer umsteigen muss, werden ihm am Fenster eine Auswahl an

möglichen Umsteigegelegenheiten angezeigt. Das SmartWindow kennt zwar die Ziele der verbundenen Geräte, zeigt sie aber nicht an. Stattdessen kann diese Information genutzt werden, um Störungsinformationen am Bildschirm anzuzeigen und zu frühzeitigen alternativen Routenvorschlägen führen.

6 Prototyping

Im Folgenden wird ein Prototyp vorgestellt, der das Konzept testen soll. Hierfür wird ein der Datenaustausch über einen QR-Code erzeugt. Die Möglichkeit, eine eigene Strecke mitzubringen und somit eine Lücke in der durchgängigen Reisebegleitung zu füllen, soll hier simuliert werden.

Der Nutzer setzt sich neben das Fenster und liest den QR-Code ein und überträgt die bereits erwähnten Daten auf das Smartphone. Dort werden weitere Daten hinzugefügt und an einen Server gesendet, der diese verarbeiten soll. Der Prototyp besteht aus zwei Komponenten: Eine Anwendung auf dem Smartphone und dem Web-Client.

6.1 Web-Client

Der QR-Code zum Testen der Verbindung enthält folgende Daten:

- Identifikation des Fensters
- Richtung der Bahn
- Linie der Bahn
- Seite des Fensters

Als Identifikation wird ein Platzhalter verwendet, Richtung und Linie der Bahn sind als statische Angaben vorhanden und werden bei der Generierung ermittelt. Der QR-Code wird dann nach Drücken des neuen Buttons (Abb. 3) angezeigt, mit einer Aufforderung, die SmartMMI-App zum Einlesen des Codes zu verwenden.

6.2 Smartphoneanwendung

Die Smartphoneanwendung muss zwei grundsätzliche Aufgaben erfüllen: Das Einlesen des QR-Codes und das Übertragen der Daten. Da es sich hier um einen Prototyp handelt und kein Zugang zur Trias-API und somit zur Routenberechnung besteht, wird sehr vereinfacht vorgegangen. Das Smartphone liest den QR-Code ein und filtert anschließend die Daten und

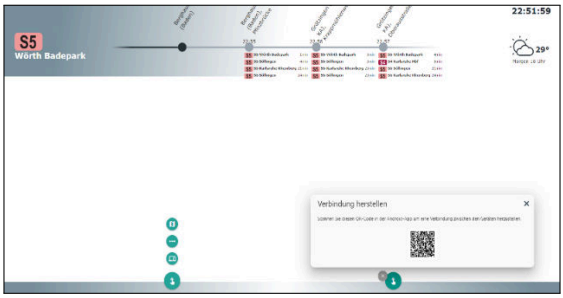


Abb. 3: QR-Code mit Aufforderung

die URL heraus und zeigt sie an. Den Daten des Fensters werden weitere Informationen über die Strecke angehängt und in ein JSON-Format umgewandelt. Die JSON-Datei wird an die URL gesendet und somit der Plattform mitgeteilt, dass sich an diesem Fenster jener Nutzer mit einer Strecke befindet, die er angezeigt bekommen möchte.

7 Fazit

Da der Prototyp simpel gehalten wurde, bleibt viel Raum für Verbesserungen und Erweiterungen offen, damit die Anforderungen in vollem Umfang erfüllt werden können. In der jetzigen Form können wenige Anforderungen durch den Prototyp erfüllt werden.

Zu den erfüllten Anforderungen gehören das Mitbringen und Mitnehmen einer Route (Funktional 1.4 und 3.1) sowie die Verbindung zu nur einem Nutzer pro Fensterseite (Funktional 1.3). Die Synchronisierung (Funktional 2.1) wird dadurch erfüllt, dass die Daten des Smartphones übertragen, und der Weg zum Anzeigen berechnet wird. Das Umsetzen einer stabilen, tatsächlichen Verbindung, die einen Datenaustausch der beiden Geräte ermöglicht, sprengt den Rahmen dieses Papers. Die große Menge an nicht erfüllten Anforderungen zeigt, dass der Anspruch an ein nutzerfreundliches, innovatives System sehr hoch ist.

Die mobile und aktive Reisebegleitung benötigt eine vorsichtige Planung der Kommunikation. Es muss sichergestellt werden, dass der Nutzer die korrekten Informationen zum korrekten Zeitpunkt erhält. Er soll nicht mit unnötigen Funktionen zur Interkation gezwungen, sondern durch geringe und gezielte Benachrichtigungen unterstützt werden. Das Ziel der Reisebegleitung ist die Erleichterung der Reise.

In einem weiteren Schritt könnte eine aufwendige Erweiterung der Anwendung durchgeführt werden, damit mehr Anforderungen erfüllt werden können. Sobald der Prototyp einem gewissen Standard entspricht, eignen sich Nutzertests, um die neuen Funktionen testen und evaluieren zu lassen. Diese Art der Evaluierung bietet viel Aufschluss darüber, welche der Anforderungen eine höhere Priorität haben als andere. Wird ein großer Fokus auf Komfort

gelegt, sollten so wenig Nutzerinteraktionen wie möglich erforderlich sein. Wird der Fokus auf einen großen Funktionsumfang gelegt, muss das System trotzdem weiterhin einfach bedienbar und intuitiv bleiben. Die Stabilität des Systems muss dabei immer in Betracht gezogen werden. Im Falle einer Verbindungstrennung muss dem Nutzer genug Transparenz geboten werden, damit er weiß, dass gerade etwas nicht funktioniert. Eine Antwort auf die Frage, welchen Nutzen dieses System hat, bietet dieses Paper nicht. Die ermittelten Anforderungen und der entworfene Prototyp zeigen dennoch, dass genug Potential besteht, eine durchgängige Reisebegleitung zu realisieren.

8 Danksagung

Dieses Projekt wurde im Rahmen des Forschungsprojekts „SmartMMI – modell- und kontextbasierte Mobilitätsinformationen auf Smart Public Displays und Mobilgeräten im öffentlichen Verkehr“ und wurde gefördert durch das Bundesministerium für Verkehr und digitale Infrastruktur als Teil der Forschungsinitiative mFund (Funding ID:19F2042A).

Literaturverzeichnis

[BL19] Bluetooth Market Update. Bluetooth Technology Website. Abgerufen am 22.06.19 von https://www.bluetooth.com/bluetooth-resources/2019-bluetooth-market-update/

[GE19] GeoSignage Sverige. Railway Technology. Abgerufen am 22.06.19 von https://www.railway-technology.com/contractors/operation/geosignage-sverige/

[IN19] Interactive Kiosk. Abgerufen am 22.06.19 von https://www.connectpointdigital.com/interactive-kiosk/

[PR15] Pohl, K. & Rupp, C (2015): Basiswissen Requirements Engineering: Aus- und Weiterbildung zum “Certified Professional for Requirements Engineering”: Foundation Level nach IREB-Standard (4., überarbeitete Auflage ed.). dpunkt.verlag, Heidelberg.

[SC18] Schlegel T. (2018): SmartMMI - Anforderungsspezifikationen, Karlsruhe.

[VI15] Viergutz K. (2015): Echtzeitdaten-Fahrgastinformation der RNV Rhein-Neckar-Verkehr GmbH, Berlin.

R. Reussner, A. Koziolek, R. Heinrich (Hrsg.): INFORMATIK 2020,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2021 1163

Analysis and Comparison of the Gaze Behavior of E-Scooter Drivers and Cyclists

Depending on Road Surface Quality in a Real Test Environment

Mathias Trefzger[1] Waldemar Titov[1] Karol Sgodzaj[1] Thomas Schlegel[1]

Abstract: In this paper, we contribute an eye tracking study to evaluate the gaze behavior of e-scooter drivers and cyclists on high and low quality road surfaces. We recorded the surface quality with sensors and put the different surfaces in relation to the gaze behavior. We recorded eye movements of the participants and performed an Area of Interest (AOI) sequence analysis to identify gaze patterns. Found sequences show that on the high quality surface participants focused most commonly the distant road section and then shifted to nearer sections. Individual advantageous gaze sequences are omitted if the surface is poor. We found a significant difference in the attention distribution of the two means of transport. In addition, we can confirm previous results showing that low quality road surfaces cause the gaze to shift forward. However, the participants did not adapt their speed to the worse surface.

Keywords: Bicycle; E-Scooter; Eye tracking; Sensors; Traffic safety; Visual analytics

1 Introduction

Since June 15 2019, e-scooters are permitted to drive on german roads. Shortly afterwards there were many reports of accidents. These have contributed that many people perceive the e-scooter as a dangerous means of transport. We want to check this first impression with the help of proven eye tracking methods. Starting with the question on how the road surface quality influences the gaze behavior of e-scooter drivers, we conducted a first study to break down the gaze behavior. Our approach is based on two studies by Vansteenkiste et al. [Va14, Va17]. To make the traffic behavior of e-scooter drivers comparable, we conducted a study with e-scooter drivers and cyclists. We have chosen this comparison because both means of transport move with similar speeds and use the same infrastructure like cycle paths and roads. In addition, both have an almost unrestricted visual field, are subject to environmental conditions like weather and have to balance while driving respectively riding.

In this paper, we investigate the gaze behavior on high quality and low quality road surfaces and compare the differences between e-scooters and bicycles. In order to make the difference between the selected route sections visible, we have recorded the vibrations caused by the

[1] Institute of Ubiquitous Mobility Systems (http://iums.eu), Karlsruhe University of Applied Sciences, Moltkestr. 30, 76133 Karlsruhe, Germany, iums@hs-karlsruhe.de

doi:10.18420/inf2020_108

roads in a preliminary study using a mobile application [Ti19]. Monitoring and classifying road conditions using bike-mounted smartphones has been a rising topic in the last few years. The reason for the rise is the availability and user acceptance of mobile devices, and general market penetration. Modern smartphones are equipped with many highly sensitive sensors and have a high computing capability. Therefore, mobile devices can be utilized for monitoring conditions in context of safety and traveling comfort of non-motorized transport types. Many authors among others [Mo08, Mo13] have stated the suitability of smartphones being used for road condition determination.

A study conducted by [La11] examines the suitability of smartphone sensor data for road condition determination. Two android devices with different sensors and computing capacities are evaluated in terms of data quality and data density by capturing the acceleration data of a smartphone. The results indicate a relatively high deviation caused by different utilized suspensions. Tackling this challenge, we use a bicycle and an e-scooter without any suspension in our study.

Because of the higher task complexity on low quality road surfaces, we expect a higher percentage of gaze in the functional space [La91] and less irrelevant fixations [Va13]. Due to the significantly poorer suspension of the e-scooter, we expect that this effect will be more pronounced with e-scooters.

2 Methodology used for Selecting a Suitable Real Test Environment

To evaluate whether the road surface quality has an impact on the attention distribution of cyclists and e-scooters, we set up a study concept consisting of two studies. Prior to the eye tracking study, we evaluated the road surface condition in a pre-study. Aiming to select suitable and meaningful representative high and low quality road sections, as our real test environment, we used our prior developed road surface evaluation tool. In [Ti19] we introduced the crowd sensing monitoring tool for road surface condition called GyroTracker. Furthermore, we made sure that the test routes had the same conditions in terms of lane width, type of road and traffic. The following subchapters introduce the method that we utilized in the pre-study to determine the road surface conditions.

2.1 Pre-Study for Accessing the Road Surface Quality

The assessment of the road surface quality uses a two-stage condition recording and evaluation procedure. In [Ti19] we previously described the two-stage procedure of data collection and classification in more detail.

In the first stage, data is collected via mounting the smartphone on the handlebar of a bicycle respectively of an e-scooter using a bicycle navigation mount as shown in figure 1. Once the app is started, the location service of the device will run in the background.

Upon completion, user's own location will be displayed on the map. Due to the high spatial relevance of the measurement data, the data recording can only begin after the device's location is available. Once located, users can start recording of both location and sensor data. During the bicycle or e-scooter ride, GyroTracker reads the four sensors: location, gyroscope, acceleration, and linear acceleration and stores the data into the smartphones database. After the completion of a recording, the measurement-id is incremented in the background, ensuring the identification of each recording.

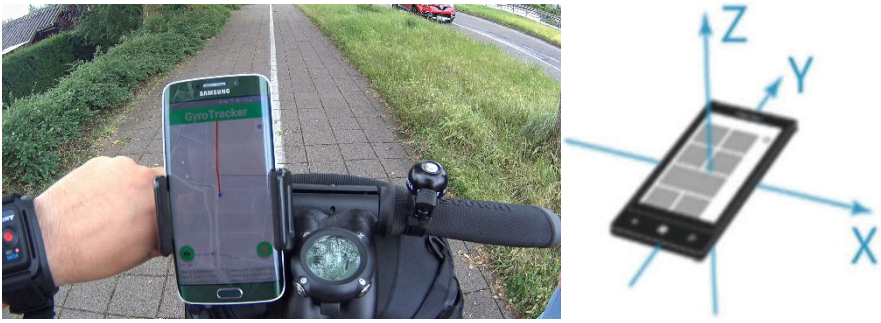


Fig. 1: Left: Collecting road data with the gyrotracker app. Right: Vertical Acceleration values

In the second stage, the recorded data is cleaned to be further processed and analyzed. To ensure the assignability of the three streaming sensors (acceleration, gyroscope and linear acceleration) with the location data, the location data must be interpolated. For that purpose, we used the simultaneously recorded timestamps of the individual sensors. After the data preparation, we analyzed the recorded data based on the concepts of the international roughness index [Du14 and Sa95]. Therefore, we used the vertical component of the linear acceleration values to calculate the road surface roughness. Dependent upon our way of mounting the smartphone the vertical acceleration components, shown in figure 1, can be extracted. Thereby the vertical component of linear acceleration is calculated as the sum of modulus of y-values and modulus of z-values of the linear acceleration data as shown in formula 1.

$$\sum |y - values| \; + \; |z - values|(1) \tag{1}$$

Based on the described approach the road surface conditions of our test environment were measured and evaluated. The results of the monitoring are compared to each other for a founded statement of the road surface conditions that our eye tracking study then is based on.

2.2 Results of the Road Surface Quality Measurements

The analyzed acceleration data collected by the bicycle and the e-scooter is shown in figure 2. The blue graphs display respectively the data of the high quality road surface with the orange graph being the mean value.

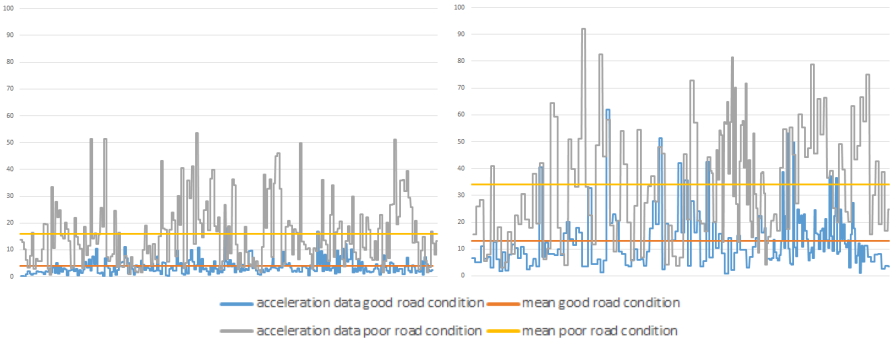


Fig. 2: Recorded road surface acceleration data with a bicycle (left) and an e-scooter (right) on low and high quality road surface

The gray graphs display respectively the data of the low quality road section with the yellow graph being the mean value. A significant deviation between the good road surface section and the poor road surface section can be seen.

The mean value of the acceleration data collected by a bicycle on a road section with high quality road surface is 4^2. Which is exactly four time less than the mean value of the acceleration data collected by a bicycle on a road section with poor road surface quality amounting to be 16 m/s^2.

In comparison to the data collected by the e-scooter it is overall higher for both road sections. The mean value of the acceleration data collected by the e-scooter on a high quality road surface is 13 m/s^2. Which is close to the mean of the acceleration data collected by a bicycle on the low quality road. The mean value of the acceleration data collected by the e-scooter on the low quality road sums up to be 34 m/s^2.

Comparing both used modes of collecting the acceleration data the bicycle (as expected) seems to be a much smoother transportation type. The ratio of the data collected by the bicycle amount to be exactly 4, the ratio of the e-scooter data amounts to be 2.6.

Additional to acceleration data we analyzed the rotation speed data collected by the gyroscope sensor. A significant deviation between the road surfaces labeled high and low quality surface is shown. Furthermore, figure 3 indicates a significant difference in perceived road surface quality between the two types of transport.

During the pre-study road surface evaluation, all participants were equipped with Samsung Galaxy S6 Edge with the android version 6.0.1 and 4 x 2.1 GHz processor devices. Deviating

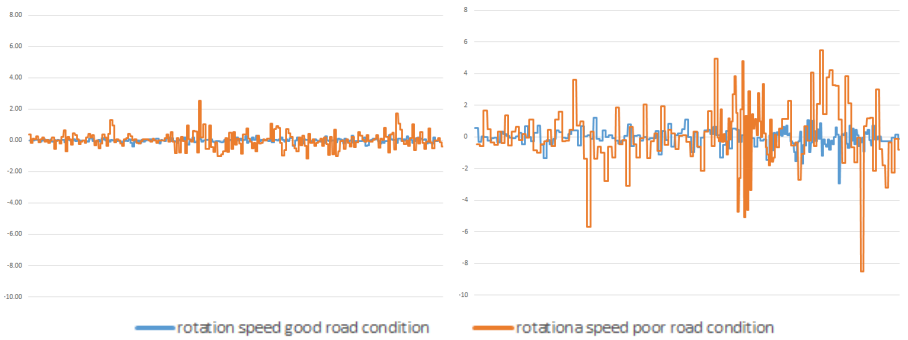


Fig. 3: Rotation speed on high and low quality roads with a bicycle (left) and e-scooter (right)

from [La11], our evaluation focus is on the method used for classification of collected data; therefore, the collection of measurement data by further devices was omitted. In the short period of our pre-study, eight measurements of different road surface qualities were carried out. The result was a data set with over 35.000 individual measuring points consisting of 945.000 single sensor values. After analyzing the collected sensor data, the road best suiting our requirements for the eye tracking evaluation was chosen. The requirements and the chosen road section is presented in the next chapter.

3 Study design

3.1 Participants

25 volunteers participated in the study. Of these, 24 eye tracking recordings were usable. 14 male and 10 female volunteers aged from 18 to 62 years participated (7 from 18-29; 12: 30-49, 4: 50-62, 1: > 63). Among them were trainees, students and employees. They showed different use of the bicycle and e-scooter in everyday life and the experiences associated with it. The average participant owned a bicycle and used it daily during spare time for errands and on the way to work. Only five participants had used an e-scooter before.

3.2 Used Devices and Transportation Modes

We used head mounted Tobii Glasses 2 to record eye movements and gaze location. The eye tracking recordings were started and stopped using a laptop with the Tobii Pro Glasses Controller Software.

The participants were provided with a bicycle and an e-scooter. Thus, all test persons used the same means of transport. The bicycle used was a conventional ladies' bicycle without

suspension. The used e-scooter was a Trekstor E.Gear (EG 3178) scooter also without suspension.

3.3 Protocol

Participants arrived individually to the dedicated location at a scheduled time. First of all a list of traffic rules in public road space was used as a reference. Afterwards the course of the study was explained to the test persons. Next, a declaration of consent had to be signed. The lenses of the eye tracking glasses were adjusted to the visual acuity of the test persons if required. Then the test persons put on the glasses and the helmet. Afterwards the eye tracking glasses were calibrated using the One Point Calibration Method. Next, the test persons were taught how to operate the e-scooter and then took a test drive to get used to it. Following, the course was explained. Then the actual ride with the e-scooter and the bicycle started. At the end of the study, the test persons had to fill in a questionnaire.

3.4 Testing Environment

The test track used is an approximately 200-meter long road section in Mannheim, Germany. The route was chosen because one-half of the road had a smooth surface (this will be referred as high quality track: HQ) and the other half contained bumps and potholes (referred as low quality track: LQ). Otherwise, the conditions were almost identical. Another advantage of the chosen road section was the low traffic volume so that the distraction caused by other road users was minimal.

Both routes had a junction where the test persons had to pay attention to the oncoming traffic. During the study, the participants had to drive back and forth along the road first using the bicycle and then the e-scooter.

3.5 Data analysis and statistics

We evaluated the eye tracking recordings using Tobii Pro Lab and Blickshift Analytics. For the evaluation of the gaze behavior, we used two similarly constructed reference images for the HQ and LQ side (see figure 4). In contrast to Vansteenkiste et al. [Va14, Va17], we divided the AOI "road" into three parts: near (“Road 1”: approx. 0-8 m), middle (“Road 2”: 8-16 m) and far (“Road 3”: $\geq$16 m). With this classification, we wanted to determine how the attention of the participants shifts with different road qualities. The length of the individual AOI road sections was chosen based on the prior experimentally evaluated breaking distance of an e-scooter from 20 to 0 km/h. The breaking distance of the bicycle was found shorter, thus being ignored.

Thereby we calculated the speed of each participant and individual type of transport in retrospect by dividing the length of the field test route by the needed time. Differences between HQ and LQ speeds were statistically tested using a t-test. Significance level was set at $P \geq 0.05$.

For the evaluation, the gaze points had to be assigned to a reference picture frame by frame by hand in the analysis software Tobii Pro Lab. Therefore, distribution to the AOIs may contain small inaccuracies. Nevertheless, differences in gaze behavior become visible.

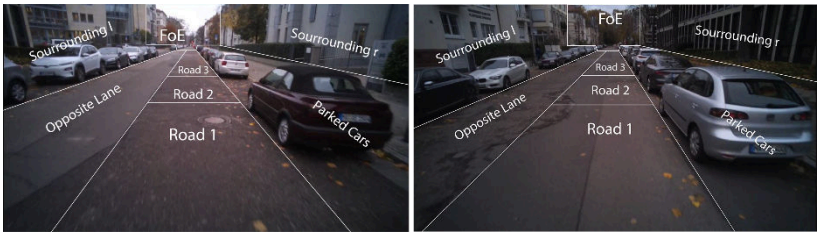


Fig. 4: Reference picture of the low (left) and high quality (right) road surface with AOI overlay.

After the evaluation with Tobii Pro Lab we used Blickshift Analytics to take a look into the gaze patterns of the participants using the Tool "Sequence Analysis" as well as visual analytic methods like Scarf Plots.

4 Results

4.1 Speed

There is no significant difference in speed for e-scooters (t=1.470 p=0.155) and cyclists (t=-1,620 p=0.119) while riding respectively driving on a good surface or bad surface. As shown in table 1 the e-scooter drivers drove approx. two km/h slower on both the LQ (t= -4.705 $p=9.715^{-05}$) and HQ (t=-3.650 p=0.001) route than the cyclists.

Tab. 1: Cycling speed (in km/h) of e-scooter drivers and cyclists on the high and low quality road surface tracks

E-Scooter Speed		**Bicycle Speed**	
High quality	Low quality	High quality	Low quality
15.93 ± 2.80	15.34 ± 2.56	17.84 ± 3.04	18.48 ± 2.97

4.2 Fixation Metrics

Additionally to the speed, we analyzed standard eye tracking metrics (Total Visit Duration, Average Visit Duration, Visit Count) to highlight the differences in the gaze behavior by the

different vehicles and road surface qualities. A visit is defined as the period of time when a participant first focuses on a region until the person looks away from that region. That means a visit can consist of at least one or multiple fixations.

Total Visit Duration

The Total Visit Duration for all AOIs was calculated to find how much time the participants spent on the AOIs (Table 2). Comparing the HQ and LQ surface a significant shift of the gaze from more distant road sections to nearer sections are visible (see fig. 5). The views on the three road AOIs increase a little bit (E-Scooter: 17%, Bicycle: 12%). Overall, e-scooter drivers spent more time looking on the road AOIs then the cyclists. On the LQ surface, the e-scooter drivers look 52% longer on the nearest road section "Road 1".

It is also noticeable that with the better road surface, the views on the parking cars are significantly longer. This is useful for protecting against doors suddenly opening in traffic.

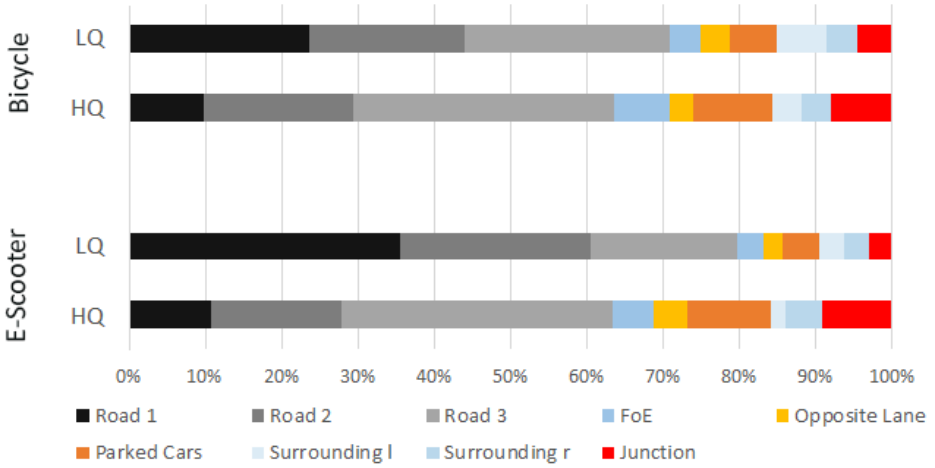


Fig. 5: Total Visit Duration percentages towards the 9 AOIs per vehicle and road type.

Tab. 2: Total Visit Duration (incl 0) in seconds

AOI	Means of Transport	Road Surface	
		High quality	Low quality
Road 1	E-Scooter	3.80 ± 3.66	13.97 ± 10.53
	Bicycle	3.02 ± 3.29	7.23 ± 6.44
Road 2	E-Scooter	6.00 ± 3.20	9.71 ± 4.27
	Bicycle	5.89 ± 3.72	6.22 ± 2.91
Road 3	E-Scooter	12.49 ± 8.13	7.53 ± 6.44
	Bicycle	10.31 ± 6.11	8,22 ± 6.87

FoE	E-Scooter	1.85 ± 2.08	1.33 ± 1.41
	Bicycle	2.18 ± 2.70	1.22 ± 2.10
Opposite Lane	E-Scooter	1.55 ± 1.35	0.96 ± 1.08
	Bicycle	0.96 ± 0.74	1.19 ± 1.18
Parked Cars	E-Scooter	3.85 ± 2.09	1.91 ± 2.42
	Bicycle	3.13 ± 2.26	1.87 ± 1.44
Surrounding l	E-Scooter	0.73 ± 0.71	1.27 ± 1.41
	Bicycle	1.15 ± 1.24	2.00 ± 1.65
Surrounding r	E-Scooter	1.64 ± 0.98	1.23 ± 1.12
	Bicycle	1.17 ± 0.82	1.22 ± 1.02
Junction	E-Scooter	3.20 ± 3.04	1.18 ± 0.96
	Bicycle	2.40 ± 2.03	1.36 ± 0.87

Average Visit Duration

Looking at the Average Visit Durations in table 3, it shows that the average time spent on the nearest road section “Road 1” increases significantly when comparing the HQ and LQ surface. This is the case for both the e-scooter and the bicycle. The difference on “Road 2” is minor and on “Road 3” the Visit Durations become a little bit shorter.

Tab. 3: Average Visit Duration in seconds

AOI	Means of Transport	Road Surface	
		High quality	Low quality
Road 1	E-Scooter	0.36 ± 0.14	0.70 ± 0.37
	Bicycle	0.37 ± 0.17	0.56 ± 0.33
Road 2	E-Scooter	0.42 ± 0.23	0.49 ± 0.21
	Bicycle	0.47 ± 0.21	0.38 ± 0.10
Road 3	E-Scooter	0.63 ± 0.31	0.50 ± 0.27
	Bicycle	0.65 ± 0.52	0.53 ± 0.39
FoE	E-Scooter	0.48 ± 0.22	0.38 ± 0.19
	Bicycle	0.55 ± 0.49	0.48 ± 0.38
Opposite Lane	E-Scooter	0.36 ± 0.15	0.34 ± 0.23
	Bicycle	0.30 ± 0.14	0.29 ± 0.19
Parked Cars	E-Scooter	0.41 ± 0.18	0.35 ± 0.24
	Bicycle	0.42 ± 0.24	0.34 ± 0.15
Surrounding l	E-Scooter	0.23 ± 0.10	0.35 ± 0.16

	Bicycle	0.33 ± 0.30	0.39 ± 0.24
Surrounding r	E-Scooter	0.35 ± 0.20	0.32 ± 0.18
	Bicycle	0.32 ± 0.17	0.39 ± 0.37
Junction	E-Scooter	0.59 ± 0.36	0.55 ± 0.36
	Bicycle	0.56 ± 0.35	0.48 ± 0.21

Visit Count

In this study, the Visit Count indicates the visual effort participants spent on the different AOIs. Table 4 shows the average visit count on the nine AOIs. As indicated in the Total Visit Duration, the number of visits to "Road 1" and "Road 2" increases from HQ to LQ. For "Road 3" the number decreases. Cyclists perform significantly fewer visits in the two nearby street AOIs than e-scooter drivers do. The number of visits by e-scooters on parked cars is almost halved. With the cyclists, the effect is not quite as pronounced.

Tab. 4: Visit Count (incl 0)

AOI	Means of Transport	Road Surface	
		High quality	Low quality
Road 1	E-Scooter	9.63 ± 7.78	18.58 ± 9.46
	Bicycle	7.38 ± 6.63	13.00 ± 8.48
Road 2	E-Scooter	14.83 ± 6.85	20.04 ± 6.60
	Bicycle	13.33 ± 6.87	16.42 ± 6.60
Road 3	E-Scooter	19.42 ± 6.44	14.38 ± 6.62
	Bicycle	17.21 ± 7.60	15.13 ± 5.89
FoE	E-Scooter	3.67 ± 3.23	3.42 ± 3.31
	Bicycle	3.92 ± 3.88	2.46 ± 2.30
Opposite Lane	E-Scooter	4.13 ± 2.85	2.75 ± 2.47
	Bicycle	3.33 ± 2.46	4.00 ± 3.30
Parked Cars	E-Scooter	9.58 ± 4.58	5.25 ± 4.11
	Bicycle	7.75 ± 3.00	5.21 ± 3.12
Surrounding l	E-Scooter	2.88 ± 2.29	3.33 ± 2.78
	Bicycle	3.58 ± 2.90	5.13 ± 2.97
Surrounding r	E-Scooter	5.25 ± 3.47	3.58 ± 2.32
	Bicycle	3.58 ± 1.47	3.67 ± 2.30
Junction	E-Scooter	5.13 ± 2.82	2.29 ± 1.43
	Bicycle	4.25 ± 1.94	2.83 ± 1.40

4.3 Visual Analysis

We use scarf plots [Ri05] for our AOI sequence analysis, which enables a closer assessment of the participants gaze behavior compared to heat maps or gaze plots. Therefore, scarf plots better illustrate the visual distribution of gaze patterns on the identified AOIs.

The following sections describe observations of the AOI sequence analysis, by first going through the main patterns observed across the e-scooter drivers and cyclists. Then we highlight noticeable patterns depending on the road surface quality, which we will later discuss.

Common Gaze Patterns

In our dataset, we searched for the most common gaze patterns. They are shown in Table 5. Generally, it can be seen that with the HQ surface the participants fixate most commonly a distant road AOI and then shift their gaze to nearer parts of the road. With the LQ surface, the gaze sequence is reversed: The nearest road AOI is fixated first then shifts to next road AOI.

Tab. 5: Most common gaze sequences found in the dataset. Legend: Road 1 (R1); Road 2 (R2); Road 3 (R3); Parked Cars (PC); Surrounding right (Sr)

Subsequence in % of input sequences	E-Scooter		Bicycle	
	HQ	LQ	HQ	LQ
100%	R3-R2	R1-R2	R3-PC	R1-R2
				R2-R3
90%	R3-PC	R3-R1-R2	R2-R3	R1-R2-R3
	PC-R3		R2-R3	
	R2-R3			
	R1-R3			
80%	R3-R1	R1-R2-R1	R2-PC	R2-R1-R2
	R3-Sr	R1-R2-R3		
	R2-PC	R2-R1-R2		
70%	R3-R2-R3	R1-R2-R1-R2	R2-R3-R2	R2-R1-R3
	R3-R1-R3		R3-R2-R3	R2-R3-R2
	R2-R3-R2			R3-R1-R2

On the HQ surface 100% of the cyclists performed, the sequence “R3-PC” as well as 90% the e-scooter drivers performed the sequence “R3-PC” and “PC-R3”. Therefore, those AOIs seem closely related – after looking at the end of the road the view to the parked cars follows. This short gaze sequence serves for the traffic safety. Looking on the sequences on the LQ surface less than 60% of cyclists and e-scooter drivers perform that gaze pattern.

Group specific patterns

We could not find consistent patterns unique for each vehicle. Therefore, we discuss more general patterns found or shared between some of them. We found that some participants tend to look at the different AOIs longer and therefore jump less back and forth between the AOIs. In contrast to this group, there are also test persons who look at the individual AOIs very briefly and constantly jump back and forth. This can be seen in figures 6 and 7.

5 Discussion

We designed our study in a way to allow us to determine how different road surface qualities have an effect on the gaze behavior on e-scooter drivers and cyclists and if there are differences between the two vehicles.

As Vansteenkiste et al. [Va14, Va17] we found no effect of the road surface quality on the cycling and driving speed. Neither cyclists nor e-scooter drivers did adapt their speed to the worse road surface. As can be seen in the graphs above (figure 2), the deflections due to unevenness of the LQ road surface are clearly visible compared to the HQ road surface condition. The driving comfort is thus significantly reduced. The vibrations themselves seem not to be the main cause for different gaze behavior: The vibrations for the e-scooter drivers on the HQ surface was almost the as on the LQ for the cyclists.

The e-scooter drivers drove approximately 2 km/h slower than the cyclists did on the different surfaces. There was therefore a significant effect on the participants in the study. This could be due to insecurity with the vehicle, caused by the lack of experience with the newer type of vehicle or due to the stronger vibrations caused by the lack of proper suspension. To test this hypothesis a further study with experienced e-scooter drivers is necessary.

In comparison with the study of Vansteenkiste et al. [Va14], there is, contrary to our expectations, only a comparatively small difference in the distribution of cyclists’ attention between poor and good road surfaces (54% and 47% in our study compared to 63% and 25% from Vansteenkiste et al.). Although part of the difference is due to the different test tracks, the difference between the two studies is high. Another reason for the difference may be the different width of the roadway. It is conceivable that, with a considerably narrower roadway of 1.3 m (LQ) to 2.0 m (HQ), more attention must be paid to ensuring that the driver does not drive over the lane boundary For a detailed comparison of the studies see table 6.

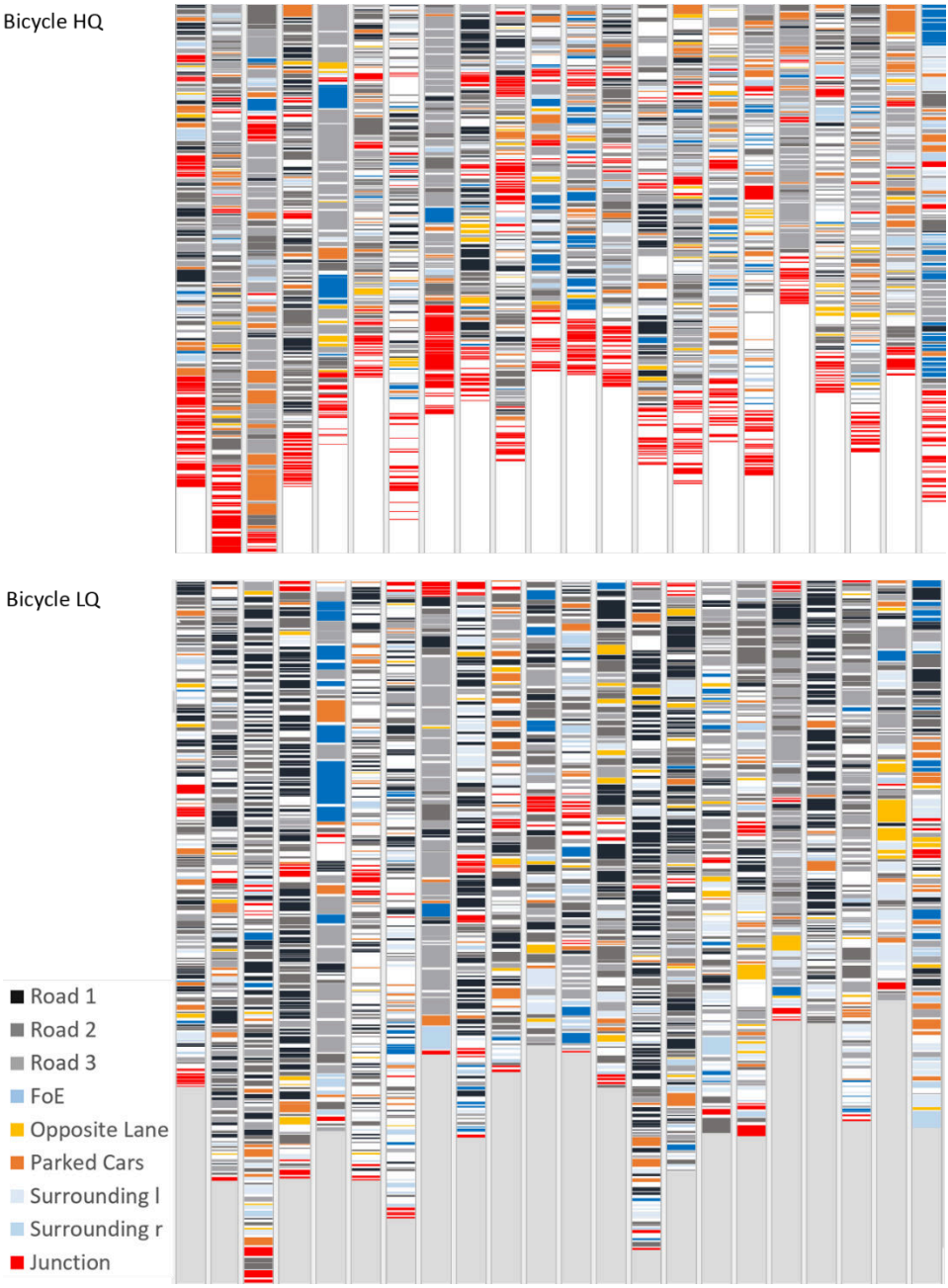


Fig. 6: Scarf Plot of all participants riding a bicycle. All visualizations have been normalized to the same height.

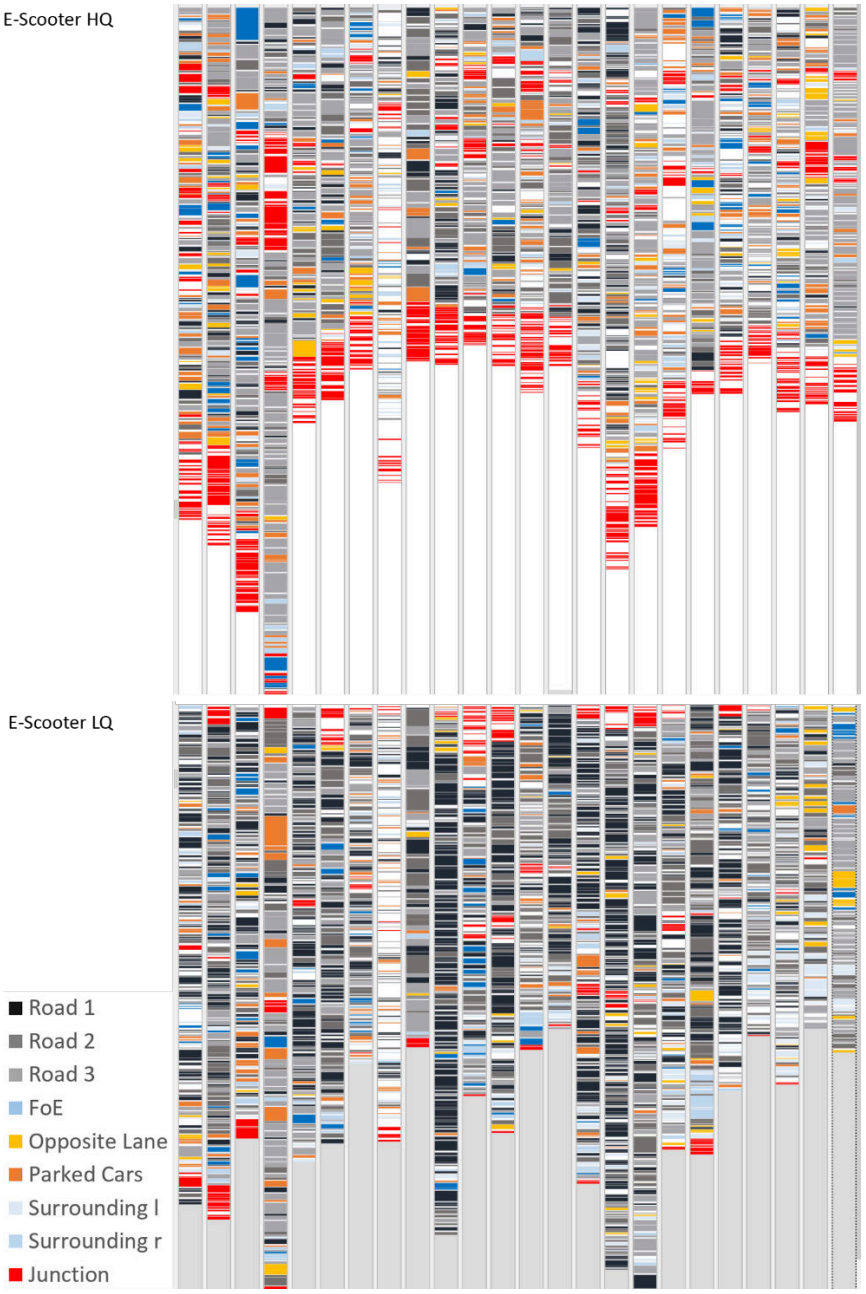


Fig. 7: Scarf Plot of all participants driving an e-scooter. All visualizations have been normalized to the same height.

Tab. 6: Key data from the two studies of Vansteenkiste et al. compared to our study

	Va14	**Va17**	**Our Work**
Participants			
Number	5	15 adults, 12 children	24
Gazepoint density	all over 80%	Participants were included when share of NoData was less than 50%	E-Scooter: 86,8 ± 6,57 Bicyle: 83,13 ± 7,11
age	22-24	adults 25.93 ± 2.71, children 9.08 8 ± 2.07	18-62
gender	1 male, 4 female	adults: 8 female, 7 male children: 8 female, 4 male	14 male, 10 female
User frequency	daily	-	daily – never
Purpose	Main means of transport	-	free time, errands, work, main means of transport, no use
Study route			
Location	Ghent, NL	Ghent, NL	Mannheim, GER
Evaluated Route Distance	120 and 136m	120 and 136m	200 m each
Lane width	HQ 2m, LQ 1,3m	HQ 2m, LQ 1,3m	HQ 3m, LQ 3m
Other			
device	IviewX Head Mounted Eye	SMI Eye Tracking Glasses 2.0	Tobii Pro Glasses 2
weather	overcast	clear	Overcast
Means of transport	bicycle	bicycle	bicycle and e-scooter

Also surprising is the relatively small difference in the distribution of attention of e-scooter drivers compared to cyclists. Due to the much stronger vibrations while driving the e-scooter (shown in figure 2), the participants focused significantly longer on the nearest road AOI in front of them compared to the cyclists. However, it can be clearly seen that the views of more distant road sections shift to nearer parts of the road. While the high quality road hardly shows any difference between e-scooter drivers and cyclists, the low quality road

shows a different distribution of the road AOIs. The e-scooter drivers focus largely on the near road AOI (29% to 19%) and the second nearest road AOI (21% to 17%). The share on the most distant road AOI thus falls shorter to 16%. Leading to a less foresighted style of traveling. With the good road surface, a full 30% of the e-scooter drivers looked at the most distant road AOI. This shows that the gaze distribution is drawn into closer areas due to poorer surface quality. In addition, a low quality surface has a greater impact on the vision of e-scooter drivers than on cyclists. If the road surface is good, there is little difference in gaze behavior between the two means of transport. The increasing Average Visit Durations from HQ to LQ strongly indicate a higher perceived risk by the participants.

Chapman and Underwood describe that dangerous situations are characterized by a narrowing of visual search, shown by an increase in fixation durations and a decrease in saccade angular and a reduction in the variance of fixation durations [Ch98]. These signs were also observed in our study. From the results, we conclude that poor road surface quality increases the risk that e-scooter drivers and cyclists are more likely to overlook other road users. The risk is greater for e-scooter drivers than for cyclists. If one considers that tourists, who usually are not familiar with the roads, often drive e-scooters we assume that the risk increases further.

6 Conclusion and Future Work

In this paper, we reported on the findings of an eye tracking study conducted with adult participants who rode the provided bike respectively drove the e-scooter on a high and low quality road. Our main goal was to analyze participants gaze behavior and patterns. Low quality road surfaces cause a shift of the visual attention at nearer road sections. We measured with our "Gyrotracker" app that the vibrations caused by the road surface were significantly stronger with the e-scooters. Therefore, we conclude that in addition to the quality of a road surface, the suspension of a vehicle also has a strong influence on the distribution of vision.

The evaluation of the gaze patterns showed that on the high quality surface most participants focused most commonly first on the most distant road section and then shifted to nearer parts of the road. With the low quality surface, this pattern is reversed.

On the high quality surface almost all e-scooter drivers and cyclists performed the sequence of first focusing the most distant road section and then the parking cars. On the low quality surface this pattern is significantly used less, which likely increases the risk of accidents through opening car doors.

With our analysis, we could not find unique patterns for the different vehicles. However, seemingly the participants can be grouped into persons who look on the AOIs shorter and jump frequently between them, as well as persons who focus longer at the AOIs but therefore jump less frequently between them.

In future work we want to evaluate the gaze sequences in even more detail. We therefore want to compare the gaze behavior of different groups of people, for example regular cyclists and people who hardly ride a bike at all. Here we want to find out if certain parameters or properties influence the gaze behavior and patterns. In addition, we want to extend the eye tracking with additional metrics, which will be recorded by additional sensors on the vehicles. This will allow us to include the external conditions in evaluating the traffic behavior.

7 Limitations

Most participants never drove an e-scooter before. Therefore, insecurity could have had an influence on the gaze behavior. This issue could be clarified in future research when participants are experienced driving e-scooters.

Bibliography

[Ch98] Chapman, R.; Underwood, G.: Visual search of driving situations: Danger and experience. In: Perception 27, p. 951-964, 1998. DOI: 10.1068/p270951

[Du14] Du, Y.; Liu, C.; Wu, D.; Jiang, S.: Measurement of international roughness index by using-axis accelerometers and GPS. Mathematical Problems in Engineering, 2014, 2014. Jg. DOI: https://doi.org/10.1155/2014/928980.

[Ho13] Hoffmann, M.; Mock, M.; May, M.: Road-quality classification and bump detection with bicycle-mounted smartphones. In: Proceedings of the 3rd International Conference on Ubiquitous Data Mining- Volume 1088. CEUR-WS.org, p. 39-43, 2013.

[La91] Laurent, M., & Thomson, J. A.: Anticipation and control in visually- guided locomotion. International Journal of Sport Psychology, 22(3-4), p. 251–270, 1991

[La11] Lauer, J.; Jochem, A.; Zipf, A.: Straßenzustandsermittlung durch Klassifikation mobiler Sensordaten von Smartphones. Doktorarbeit. Master's thesis, Abteilung Geoinformatik Geographisches Institut, Universität Heidelberg.

[Mo08] Mohan, P.; Padmanbhan, V.; Ramjee, R.: Nericell: rich monitoring of road and traffic conditions using mobile smartphones. In: Proceedings of the 6th ACM conference on Embedded network sensor systems, p. 323-336, 2008. DOI: https://doi.org/10.1145/1460412.1460444.

[Ri05] Richardson, D.; Dale, R.: Looking To Understand: The Coupling Between Speakers' and Listeners' Eye Movements and Its Relationship to Discourse Comprehension. In: Cognitive Science 29 (6), p. 1045 – 1060, 2005.

[Sa95] Sayers, M.: On the calculation of international roughness index from longitudinal road profile. In: Transportation Research Record, Nr. 1501, 1995.

[Sc11] Schepers, J.P.; Den Brinker, B.P.L.M.: What do cyclists need to see to avoid single-bicycle crashes? Ergonomics 54 (4), p. 315–327., 2011. DOI: https://doi.org/10.1080/00140139.2011.558633.

[Ti19] Titov, W.; Schlegel, T.: Monitoring Road Surface Conditions for Bicycles – Using Mobile Device Sensor Data from Crowd Sourcing. In HCII 2019: HCI in Mobility, Transport, and Automotive Systems. Springer, p. 340-356, 2019. DOI: 10.1007/978-3-030-22666-4_25.

[Va13] Vansteenkiste, P..; Cardon, G.; D'Hondt, E.; Philippaerts, R.; Lenoir, M..: The visual control of bicycle steering: The effects of speed and path width. Accident Analysis & Prevention Volume 51/2013, p. 222-227, 2013. DOI: https://doi.org/10.1016/j.aap.2012.11.025.

[Va14] Vansteenkiste, P.; Zeuwts, L.; Cardon, G.; Philippaerts, R.; Lenoir, M..: The implications of low quality bicycle paths on gaze behavior of cyclists: A field test. Transportation Research Part F: Traffic Psychology and Behaviour Volume 23/2014, p. 81-87, 2014. DOI: https://doi.org/10.1016/j.trf.2013.12.019.

Workshop on Tools and Concepts for
Communication and Networked Systems

Tools and Concepts for Communication and Networked Systems – Or: How to build resilient IoT Systems?

Mesut Güneş[1], Sebastian Zug[2], Matthias König[3]

The first workshop on Tools and Concepts for Communication and Networked Systems (TCoNS) will be on Friday 2. October 2020 at Kongresszentrum Karlsruhe, co-located with the 50. GI-Jahrestagung INFORMATIK2020.

The goal of the workshop is to provide a platform for discussions about the evolution of the Internet of Things (IoT) from a practical point of view that is best described in the subtitle of the workshop, namely "How to build resilient IoT Systems?"

The Internet of Things (IoT) and related future networking concepts promise the ubiquitous availability of data. Applications can aggregate and evaluate relevant sets of data, and they can provide highly flexible, context-aware services, which can interact with each other and form a new type of emergent behavior.

The first part of the story has already become reality. Sensors aggregating current values as well as data storages providing historical information emerge wildly around us in terms of computational capacity and data quantity. But did we already achieve the goals from an application and coordination point of view? Embedded, low performance IoT devices transmit their data periodically or event-driven to a database that serves the applications' requests. By separating data producers from recipients, the traditional approach limits the flexibility and efficiency of the system. In contrast, accessibility and controllability of the IoT nodes immediately by (multiple) applications ensure a finely tuned configuration of individual embedded systems as well as the whole network. It will no longer be valuable to restrict IoT systems to static behavior, thus new methods, models, and algorithms are required to ensure functionality, resiliency, and security.

This workshop addresses current research related to the implementation and realization of future IoT applications and systems. One focus is put on the practical side of challenges, i.e., how to build an IoT system at least in a prototypical way. The focus is mainly on concepts, tools, and the toolchain, required for this endeavor. Furthermore, flexible right management, capability and performance profiles, and request evaluation for dynamically composed IoT

[1] Otto-von-Guericke Universität Magdeburg, Fakultät für Informatik, Universitätsplatz 2, 39106 Magdeburg, Germany, mesut.guenes@ovgu.de
[2] Technische Universität Bergakademie Freiberg, Informatik, Germany, sebastian.zug@informatik.tu-freiberg.de
[3] Bielefeld University of Applied Sciences, Campus Minden, Artilleriestraße 9, 32427 Minden, Germany, matthias.koenig@fh-bielefeld.de

doi:10.18420/inf2020_109

settings will be targeted. We want to discuss how to realize abstract representations of these aspects in order to automatically react to adapted requests, changed network configurations, or system states.

All submissions to the workshop have gone a rigid review by at least three reviewers, who pointed out ways to improve the papers. We thank all anonymous reviewers for their contribution to the workshop. Luckily enough, we could accept all submitted eight papers to the workshop, which represent diverse perspectives of the ongoing endeavor of how to build resilient IoT systems.

Organizers:

- Mesut Güneş, Otto-von-Guericke Universität Magdeburg
- Sebastian Zug, Technische Universität Bergakademie Freiberg
- Matthias König, Fachhochschule Bielefeld

Workshop TPC Chairs: Frank Engelhardt, Marian Buschsieweke, Ali Nikoukar

Publicity Chair: Katja Nothnagel

Technical Program Committee:

- Andreas Reinhardt, Technische Universität Clausthal
- Anna Förster, Universität Bremen
- Bettina Schnor, Universität Potsdam
- Björn Scheuermann, Humboldt Universität zu Berlin
- Christian Bettstetter, Universität Klagenfurt
- Christian Renner, Universität zu Lübeck
- Claudia Linnhoff-Popien, Ludwig-Maximilians-Universität München
- Jochen Schiller, Freie Universität Berlin
- Joerg Nolte, Brandenburgische Technische Universität Cottbus-Senftenberg
- Karin Anna Hummel, Johannes Kepler Universität Linz
- Kay Roemer, Technische Universität Graz
- Kurt Tutschku, Blekinge Institute of Technology
- Lars Wolf, Technische Universität Braunschweig
- Oliver Hahm, RIOT OS / Zühlke
- Reinhard German, Freie Universität Berlin
- Stefan Fischer, Universität zu Lübeck
- Thomas Schmidt, HAW Hamburg

R. Reussner, A. Koziolek, R. Heinrich (Hrsg.): INFORMATIK 2020,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2021 1185

Evaluation of Interoperability Between Various Implementations of the Thread Protocol Stack

Sebastian Miethe[1], Silvia Krug [2]

Abstract: The increasing popularity of Internet of Things (IoT) applications leads to a continuous expansion into further application areas. Since each application poses its own challenges and requirements, many different solutions have been developed to build parts of the IoT. However, this diversity results in new challenges as well. Especially if a cooperation between heterogeneous components or a later addition of components is required, interoperability becomes challenging. Precision agriculture is one example application area that requires both conditions. In this paper, we assess the interoperability of three different implementations of the Thread protocol stack. To this end, we propose an empirical analysis method that can similarly be applied to any other interoperability evaluation. Our results show that the versions under test exhibit various interoperability problems and only certified devices work without issues.

Keywords: Internet of Things; Protocol Interoperability; Thread; OpenThread

1 Introduction

Smart or precision agriculture opens up new possibilities for farmers to manage their assets efficiently. Therefore, farmers have become early adopters for digital technologies such as the Internet of Things (IoT) [KJL19]. However, the fast development of technologies results in many design options on the market that farmers cannot overview. Besides that, limited budgets can lead to successive installations. This leads to the requirement of future-proofed and extendable technologies, which is made possible through interoperability between heterogeneous system components. If two systems can exchange data and work together according to a protocol specification, they are considered interoperable [Zh08]. To what extend this can be guaranteed is however an open question.

There are several reasons for limited interoperability, ranging from constrained hardware resources through faulty or incomplete protocol implementations and even security reasons [KB17]. Interoperability analyses as such are nothing new. In [Ay18] for example, the authors show that no implementations of 6LoWPAN covers all features. The reasons for this are the complexity of the protocol and hardware restrictions. Missing implemented features caused package drops and so limits interoperability. In [Ko11], the authors analyze Routing Protocol for Low Power and Lossy Networks (RPL), an IoT routing protocol.

[1] IMMS Institut für Mikroelektronik- und Mechatronik-Systeme gemeinnützige GmbH (IMMS GmbH), Ehrenbergstraße 27, 98693 Ilmenau, Germany — sebastian.miethe@imms.de
[2] IMMS GmbH — silvia.krug@imms.de, https://orcid.org/0000-0003-0282-5471

doi:10.18420/inf2020_110

They show that certain versions can work together, but performance metrics might drop. Another analysis, targeting Constrained Application Protocol (CoAP), has shown that the choice of programming language may lead to faster responses to client requests while some implementations were not able to work with others at all due to different protocol versions [IOU17]. Reasons for limited interoperability and its effects are divers for different protocols and entire IoT-systems. Limited interoperability results in problems when *joining a network* or during *network operation* in terms of stability and transfer-related *performance parameters*. So further investigations on interoperability in IoT are needed.

In this paper, our focus was to experimentally evaluate the interoperability of three implementations of the Thread protocol stack and their impact on real deployments. We developed an adaptive model allowing various experiments to verify the extend of cooperation between the chosen implementations. Other more formal approaches to verify the interoperability were therefore out of the scope of this work. In addition, we evaluate typical performance metrics to observe variations caused by lacking interoperability. As explained later, Thread addresses interoperability directly however issues have to be expected nevertheless.

2 Fundamentals

2.1 The Thread Protocol stack

Thread is a protocol stack for wireless mesh networks first released in 2015. The main goals of Thread are in low power consumption, secure networking, being economical in purchase and operation, simple network installation and operation, and scalability. It is based on well-known standards and covers the physical through transport layers (see Figure 1). The application layer is intentionally unspecified to allow various IP-based applications. [Th17b]

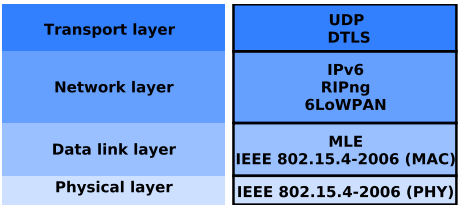


Fig. 1: Thread protocol stack [Th15b]

2.2 Joining a Thread Network

Since we aim to test the ability to join an network, it is necessary to understand how this process works. A so-called *joiner* needs to complete the three steps *discovery*, *commissioning*, and *attaching* in order for it to successfully join a network [Th15b].

In the *discovery process*, a joiner actively searches for an existing network via Mesh Link Establishment (MLE). It sends out a *Discovery Request* and waits for a *Discovery Response*. Afterwards, the *commissioning process* according to the Mesh Commissioning Protocol (MeshCoP) starts [Th15a]. Every network needs exactly one router assuming the role of *commissioner*. The joiner needs to authenticate at the commissioner with a key and its Medium Access Control (MAC) address. Then, the network credentials can be passed to the joiner via Datagram Transport Layer Security (DTLS). After the commissioning step, the joiner needs to find a parent to which it can *attach* itself. The joiner requests this with a multicast message for potential parent [Th15a]. After *attaching*, it becomes a full member of the network. If the number of routing devices in the network is below a certain threshold, the joiner request the leader to become a routing device, via CoAP [Th15b].

2.3 Thread Product Certification

Several companies develop own implementations of Thread. *The Thread Group, Inc.* offers a certification program which is voluntary and shall ensure interoperability between Thread devices [Th17a]. If a SoC and protocol implementation passes all tests, it can use the label *Thread Certified Component*. This is relevant because we test certified and uncertified products allowing us to draw conclusions on the significance of the certification.

3 Measurement Setup and Implementation

The interoperability of the three implementations Openthread, Kinetis Thread Stack and Mbed Thread Stack will be evaluated. OpenThread and the Kinetis Stack implement version 1.1.1, while Mbed implements version 1.1.0 of the Thread-specification. These versions should work together well, because only errata were corrected in version 1.1.1.

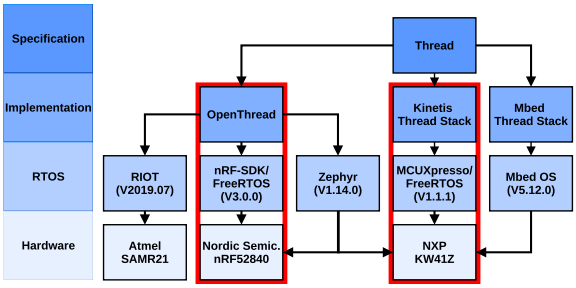


Fig. 2: Utilized soft- and hardware

Figure 2 gives an overview of the implementations, cross-referenced with the respective Real-Time Operating Systems (RTOSs) and its version as well as the used hardware platform. The two combinations outlined in red are officially *Thread Certified Components*. Using

these implementations, we will perform experiments in selected scenarios: *joining a network* and the *network operation with* and *without application data traffic*.

3.1 Modeling

We developed a general model as common approach for testing the interaction between implementations. It is applicable to other protocols. Figure 3 shows the model for testing the implementations in the process of joining a network. The SUT (System Under Test) contains two or more IUTs (Implementation Under Test) and observes their interactions. To control the SUT and its IUTs, inputs are needed. Most of the used RTOSs offer a Command Line Interface (CLI) which allows the user to manage things like joining a network manually via commands. We use the CLIs as an input. The CLI can also be used as output because it allows the user to request information about the network status and settings. E.g. a *neighbor table* can be requested to check if nodes unexpectedly lose connection to the network. Since Thread uses *IEEE802.15.4*, an *IEEE802.15.4*-compliant sniffer with *Wireshark* is used to capture the network traffic adn compare it according to the analyses in Section 2.2.

3.2 Joining a Network

To verify if every implementation is able to join a network of nodes with a different implementation, we set up a node (IUT A) which advertises a network. Network settings like channel, network name, and Pre Shared Key for the Device (PSKd) have to be adjusted in the source code or via the CLI. Then, the alternative implementations (IUT B) join the network one by one. We captured the packet exchange during this process in order to compare it to the expected exchange according to the specification. Figure 3 shows an incomplete finite state machine for the joining process used as reference for evaluation.

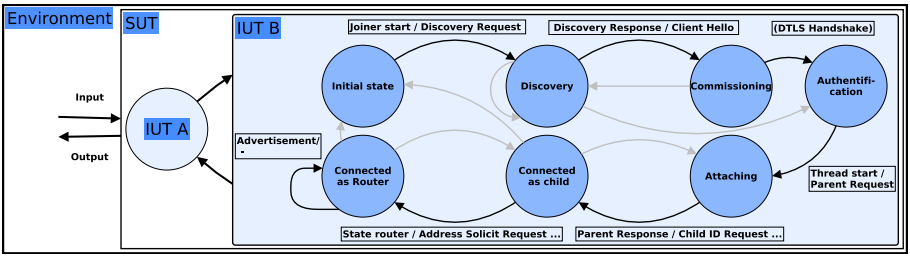


Fig. 3: Packet exchange for joining a network

3.3 Network Operation Without Application Data Traffic

When a node has joined a network, no disconnections among any other nodes should happen. After topology changes the network is kept alive via so-called *advertisements*.

Those are periodically sent by every node, as determined by the *Trickle Algorithm* [Th17b]. To examine mutual influences between pairs of implementations, we set up individual networks consisting of two nodes and observed them for one hour each. We recorded the packet exchange and neighbor table of each setup, to check that nodes remain continuously connected. In addition, we set up a network with all combinations of implementations and hardware platforms and observed them for five days to evaluate any long-term effects.

3.4 Network Operation With Application Data Traffic

Finally, we verify if parameters such as latency, UDP- or MAC-packet error rate (PER) are influenced by the implementation. We use UDP to generate user data traffic since it is used at the highest layer of the Thread stack (see Figure 1).

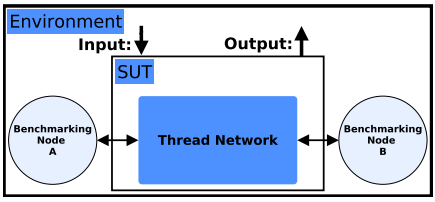


Fig. 4: Model for measuring latency and PER

Figure 4 shows two nodes with an application for benchmarking UDP-traffic in a Thread-network. This network consists of two nodes that were varied in their implementation. After *Benchmarking Node A* has sent a UDP-packet it waits for a UDP-acknowledgment from *Benchmarking Node B*. The *latency* [*ms*] is measured as the Round-Trip-Time (RTT) from sending the UDP-packet until obtaining the UDP-acknowledgment. Furthermore, the *UDP-PER* [%] is recorded as the percentage of retransmissions needed due to missing UDP-acknowledgments. Similarly, the *MAC PER* [%] is determined.

We performed the measurement in two different configurations. In configuration I 1.000 UDP packets were sent with a payload of 20 Byte each and an acknowledgment timeout of 120 ms. In configuration II the payload and timeout are increased to 100 Byte and 216 ms respectively. The increase in size forces the 6LoWPAN-layer to perform fragmentation and show a variation in delay. We chose the *ack timeout* based on empirical tests. It is quite generous, because we are interested in packet loss due to interoperability issues and not due to high latency. The transmission rate is not fixed, the sender sends the next packet after getting the acknowledge or after the time exceeds. To avoid package loss due to low signal strength caused by obstacles and path loss, we kept the nodes in close proximity and a maximum distance of one meter. We set up the wanted network topology manually with the help of MAC filtering via the CLIs configuring MAC-blacklists. In addition, we ensured that all nodes rated incoming packets with *Link Quality* = *3*, which corresponds to an RSSI of more than 20 dB, which would be rather high in a real world scenario. The network

was varied by permuting the three tested implementations as table 1 shows. We chose the *nRF-SDK* on the *nRF52840* as Openthread variant since it worked without errors in previous tests. For Kinetis, the *MCUXpresso-SDK*, and for the Mbed Thread Stack *MbedOS* on the *NXP KW41Z* was used (see Figure 1). We tested every permutation five times with the two configurations and calculated the average values from these measurements.

	homogeneous			heterogeneous		
IUT A	Kinetis	Mbed	Openthread	Kinetis	Mbed	Openthread
IUT B	Kinetis	Mbed	Openthread	Mbed	Openthread	Kinetis

Tab. 1: Combinations of IUTs whose latencies and PER were to be measured

4 Results and Discussion

4.1 Joining a Network

Now, we examined the experiments. Beginning with the network joining process, all combinations where observed three times to validate that occurring errors are reproducible. Table 2 summarizes the results. It shows all tested combinations of a given *commissioner* and a *joiner* node.

			Joiner					
			nRF52840		KW41Z			SAMR21
	Hardware	Software	nRF-SDK	Zephyr	Zephyr	Mbed	Kinetis	RIOT
Commissioner	nRF-52840	nRF-SDK						1*
		Zephyr					2*	
	KW41Z	Zephyr						
		Mbed						3*
		Kinetis						4*
	Atmel	RIOT						

Tab. 2: Overview of the results for the network joining process

The Mbed Thread Stack does not implement the joiner and commissioner. Since we focused on testing the initial state of each implementation, we were not able to test the Mbed Thread Stack at this point. Combinations marked with green exhibited the expected packet exchange as described in section 2. No issues occurred so interoperability appeared to be given. Combinations marked with yellow were also able to join, but some deviations from the expected packet exchange occurred. This shows that all combinations with *Zephyr OS* on the *KW41Z* System-on-a-Chip (SoC) were affected. We observed that duplicates of packets appeared frequently during the DTLS-handshake. At a closer look, the duplications happened because *Zephyr* exceeds the *THR_DTLS_INIT_RETRANSMIT_TIMEOUT* = 8 s during the DTLS-handshake. This leads to retransmissions as stated in the specification

[Th17b]. In some combinations, it also happened that retransmitted packets were interpreted by the receiver as another DTLS handshake request. This caused a temporary processing of two DTLS handshakes in parallel between the same two nodes. Thus, in a worst case scenario, 46 % more packets had to be sent for the joining process, which is especially detrimental for energy-constrained devices. Furthermore, a long runtime on batteries is mandatory for IOT-systems to be practicable for farmers. All combinations with *RIOT* are highlighted with red. When *RIOT* is the joiner, the process triggers a *RIOT kernel panic* on the CLI and it stops working. As commissioner, it does not reply to the joiner during the DTLS handshake. This behaviour prevents any joining and could happen due to package drops caused by missing features in the implementation.

4.2 Network Operation Without User Data Traffic

4.2.1 Individual Observation

In the next step, we observed all implementations separately in networks of two nodes. The network credentials where predefined in source code to skip the commissioning. Nearly all combinations behaved as expected, except for the ones marked with 1* through 4* in Table 2. In combinations 1* and 2* *RIOT* respectively *Zephyr*, stopped sending the periodical *advertisements* at some point. These are necessary to keep connections up between the participants of the network and thus keeps the network as a whole alive. As a result, the remaining node discarded the routing entry for the other node. This way, *RIOT* had not been part of the network for about 4:50 min and *Zephyr OS* even for 8:30 min.

In combinations 3* and 4*, *RIOT* did not reply to the requests from the Mbed Thread Stack and Kinetis Thread Stack to become routing devices. Therefore, they remained as a child. It was interesting for us that *RIOT* did reply to all other combinations, which are based on Openthread, namely *Zephyr*, *nRF-SDK* and *RIOT* itself. Conversely, *Zephyr* and *nRF-SDK* replied to *Mbed* and Kinetis Thread Stack. This indicates that there is an issue with *RIOT* in combination with Openthread, but not with Openthread itself. The reason might be that *RIOT* drops the CoAP request due to implementation incompatibilities.

4.2.2 Collective Observation

After observing all implementations separately, the operation of all nodes at once was monitored for five days. We did this three times to verify that observed effects where reproducible. The duration in Table 3 show how long the nodes participated in the network during the total duration of 120 h.

nRF52840		KW41Z			SAMR21
nRF-SDK	Zephyr	Zephyr	Mbed	Kinetis	RIOT
120 h	36.4 h	70.7 h	120 h	120 h	120 h

Tab. 3: Overview of the results for the network operation for five days

The *RIOT* node was part of the network all the time but disrupted the rest of the network. That is why we marked it with yellow. *RIOT* exceeded the *MAX_NEIGHBOR_AGE* = 100 s period for sending *advertisements* [Th17b]. We found that this caused the other nodes to deleted its routing entry and created it again when it started sending again. This happened periodically about every 72 min and is problematic in two ways. At first, nodes consider their *RIOT* peer gone and thus are not able to send packets to it. After the *RIOT* node starts sending again, all nodes want to disseminate the change in the network. This results in a significant increase in the number of messages exchanged and thus again in a higher energy consumption. The combinations of cells marked with red completely disconnected from the network. As we noticed this happened because they stopped sending *advertisements* after a certain time. As the microcontroller did not respond any more via the CLI, *Zephyr OS* probably crashed.

4.3 Network Operation With User Data Traffic

To evaluate differences in performance between the selected implementations, we performed the experiment as detailed in Section 3.4 with one participant of each of the three implementations. Figure 5 shows the measured values for the homogeneous networks. The latency shows that Openthread is the fastest implementation, which could be due to differences in the used hardware platforms. Interesting is the UDP-PER with the Mbed Thread Stack, which is extremely high compared to the others. We found out that this was caused by missing acknowledgments on the MAC layer that were also responsible for the higher MAC-PER. Those missing acknowledgments by the Mbed Thread Stack also caused connection losses and reattachments with the other participants. Those disappearing paths to neighboring nodes reduces the redundancy in a network and therefore reduce its resilience.

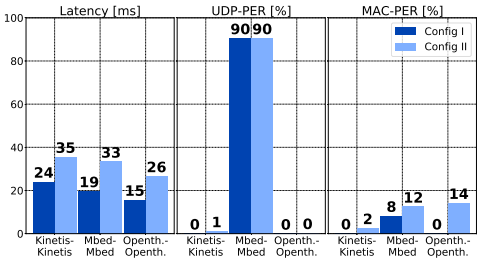


Fig. 5: Homogeneous networks

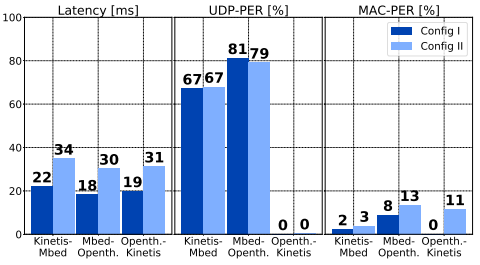


Fig. 6: Heterogeneous networks

The results with the heterogeneous networks are rather predictable. The combinations with a Openthread participant e.g. have the lowest latency because Openthread already was the fastest in the homogeneous networks. The same is true for the UDP-PER. We found that combinations with the Mbed Thread Stack show bad results since Mbed Thread Stack was again responsible for missing acknowledgments. Interestingly, the Kinetis Thread Stack-Mbed Thread Stack combination showed around 14 % less PER than Mbed Thread Stack-Openthread. Still, a PER of 67 % is not good compared to the 0 % at Openthread-Kinetis Thread Stack.

The overall results show that interoperability between various Thread stacks cannot be taken for granted. The fact that errors occurred with OpenThread used in combination with *RIOT* and *Zephyr* but not with the *nRF-SDK*, even thought they all use the same version of Thread, shows that not just the implementation but the used RTOS can be the cause of problems. Only the Kinetis Thread Stack and OpenThread on the *nRF-SDK* exhibited no issues, full interoperability and stable network operation. All other combinations caused problems that restricted the interoperation with other nodes and made operation of the network unreliable and therefore less resilient.

The two certified stacks thus are to be preferred for application development because this lowers the risk of having to deal with limitations and issues during development and later operation. For the considered domain of agricultural applications in particular, robust network operation and future-proofing interoperability are prime concerns. Especially, topology changes are crucial for agriculture applications. There, it is likely that nodes get disconnected e.g. to save power or due to changed vegetation properties altering the signal strength at the receiver. Dealing with such harsh conditions, requires a stable and robust network operation as base in order to cope with different dynamic effects. This aspect will be covered in future studies.

5 Conclusion and Future Work

In this paper, we evaluated the interoperability of the Openthread, Kinetis- and Mbed thread stack, targeting the application in agriculture scenarios. We achieved this evaluation by modeling typical scenarios and performing empirical tests. The investigations confirmed that interoperability cannot be taken for granted and should be evaluated. The results confirm that officially Thread-certified products work together flawlessly with other certified stacks. This makes networks more stable as no unexpected behaviour could be observed. All other implementations experienced issues.

Continuing this work, further Thread-certified components should be analyzed to confirm that they effectively prevent interoberability issues. In addition, more scenarios besides the network joining process and operation should be considered. One crucial aspect we will focus on is the behavior in case of topology changes due to harsh environmental conditions as typically experienced in outdoor agriculture applications.

Acknowledgements

Gefördert durch:

aufgrund eines Beschlusses
des Deutschen Bundestages

The EXPRESS project is supported by funds of the Federal Ministry of Food and Agriculture (BMEL) based on a decision of the Parliament of the Federal Republic of Germany. The Federal Office for Agriculture and Food (BLE) provides coordinating support for digitisation in agriculture as funding organisation, grant number FKZ 28DE102C18.

References

[Ay18] Ayers, H.; Thomas Crews, P.; Hua Kian Teo, H.; McAvity, C.; Levy, A.; Levis, P.: Design Considerations for Low Power Internet Protocols. 2018.

[IOU17] Iglesias-Urkia, M.; Orive, A.; Urbieta, A.: Analysis of CoAP Implementations for Industrial Internet of Things: A Survey. Procedia Computer Science 109/, pp. 188–195, 2017.

[KB17] Konduru, V. R.; Bharamagoudra, M. R.: Challenges and solutions of interoperability on IoT: How far have we come in resolving the IoT interoperability issues. In: 2017 International Conference On Smart Technologies For Smart Nation (SmartTechCon). Pp. 572–576, 2017.

[KJL19] Klerkx, L.; Jakku, E.; Labarthe, P.: A review of social science on digital agriculture, smart farming and agriculture 4.0: New contributions and a future research agenda. NJAS - Wageningen Journal of Life Sciences 90-91/, 2019.

[Ko11] Ko, J.; Eriksson, J.; Tsiftes, N.; Dawson-Haggerty, S.; Terzis, A.; Dunkels, A.; Culler, D.: ContikiRPL and TinyRPL: Happy Together, 2011.

[Th15a] The Thread Group, Inc.: Thread Commissioning, 2015, URL: www.threadgroup.org/Portals/0/documents/support/CommissioningWhitePaper_658_2.pdf, visited on: 07/10/2019.

[Th15b] The Thread Group, Inc.: Thread Stack Fundamentals, 2015, URL: www.threadgroup.org/Portals/0/documents/support/ThreadOverview_633_2.pdf, visited on: 06/07/2019.

[Th17a] The Thread Group, Inc.: Thread Certification Program Opens: Top Reasons To Get On Board Now!, 2017, URL: www.threadgroup.org/news-events/blog/ID/146/Thread-Certification-Program-Opens-Top-Reasons-to-Get-on-Board-Now#.XRCFY69R09l, visited on: 06/16/2019.

[Th17b] The Thread Group, Inc.: Thread Specification, version 1.1.1, 2017.

[Zh08] Zhong, N.; He, Z.; Kuang, J.; Zhuo, Z.: Optimal Protocol Interoperability Test Generation via Heuristic Algorithm. In: 2008 International Conference on Internet Computing in Science and Engineering. Pp. 278–281, 2008.

R. Reussner, A. Koziolek, R. Heinrich (Hrsg.): INFORMATIK 2020,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2021 1195

The PASTA threat model implementation in the IoT development life cycle

Andreas Wolf[1] Dimitrios Simopoulos[2] Luca D'Avino [3] Patrick Schwaiger[4]

Abstract: Recently, IoT usage has grown rapidly. Security risks are rising analogously, though. Our paper introduces an approach to identify and address security threats by applying the PASTA (Process for Attack Simulation and Threat Analysis) threat model to the IoT domain. By adapting PASTA, we optimize the threat analysis based on domain knowledge and specific needs of IoT. With integration of the PASTA results into the development process and the IoT software development life cycle, we reduce security risks. A prototype demonstrates the feasibility of the concept for security vulnerability reduction via an integrated DevSecOps toolchain.

Keywords: Internet of Things; Threat Model; DevSecOps; Cyber-Security; Security Testing

1 Introduction

The Internet of Things (IoT) is becoming an integral part of our lives. With all the negative news about data leaks, privacy violations and device compromises, trust in technology is the key acceptance and success factor for IoT. Even though positive effects of IoT are proven, like resource and cost saving (weather based irrigation control), security (intrusion detection) and even life saving (smoke detectors), it will only be accepted, if we can trust it.

IoT differs from standard applications and systems due to the inherent complexity and the number of involved components. Standard applications usually consist of a few up to 10th of components (back-end servers, back-end networks, Internet, (browser)-client). IoT however starts on a different scale. Even in the smallest Cloud-based home automation scenario we have dozens of sensors on doors and windows, IP-Cameras, sun-blinds with light and wind sensors and motors, an IoT-Hub as gateway between e.g. a ZigBee IoT network, Wi-Fi and the Internet. Industrial scenarios scale that up to multiple on premise IoT networks with hundreds of sensors and actuators, multiple (hierarchical) hubs and concentrators, cloud links and services and interfaces to multiple back-end systems. Therefore, IoT always requires a system perspective.

In the context of this paper, an IoT system covers at least the local IoT network as well as services provided by the Internet. Services are vital for the system, as they provide the

[1] AKKA DSO GmbH, Taunusstraße 36, 80807 Munich, Germany, andreas.wolf@akka.eu
[2] AKKA DSO GmbH, Taunusstraße 36, 80807 Munich, Germany, dimitrios.simopoulos@akka.eu
[3] AKKA DSO GmbH, Taunusstraße 36, 80807 Munich, Germany, luca.davino@akka.eu
[4] AKKA DSO GmbH, Taunusstraße 36, 80807 Munich, Germany, patrick.schwaiger@akka.eu

doi:10.18420/inf2020_111

respective functionalities. Potential existence of threats leads to vulnerabilities, which are weaknesses of a system that make an exploit possible and can be found on network, host or application levels, including operational practices [UM15]. Successful risk handling plays a significant role in the system functionality. Although there are techniques facilitating the general software life cycle, special treatment is necessary for IoT systems, applications, and operating systems.

Due to the variety of components in an IoT system, we have different types of security measures to apply. It starts with the hardware, a Microcontroller unit (MCU) -based board or a System-on-Chip (SOC) that can contain different types of design flaws. It continues with communication links with authentication, authorization, encryption and validation protocols, which due to their computational complexity are often trades for security. IoT systems also include software in the form of firmware, operating systems, services and applications opening them for the standard threats like buffer overflows, especially because of the limited memory resources [MM19] or Random Number Generator (RNG) attacks [Ke98]. Finally, secure boot and component registration absence can lead to undesirable system functionality and malicious software boot.

This paper discusses methods and practices for dealing with potential security risks and complex threats during the IoT system life cycle. We adapt a specific threat model for deriving design security requirements, architecture- and implementation-countermeasures and verification for the Continuous Integration tool chain. A methodology is defined to predict and manage the potential threats that arise in the process of developing and using any such IoT system.

2 Literature Review

Threat models are integral to risk identification. A plethora of such models were created over time. The most mature threat model was proposed by Loren Kohnfelder and Praerit Garg in 1999 under the name STRIDE and was widely used for several years by Microsoft [KG99]. The specific model evaluates the overall design of the system, either through data flow charts (DFD's) or through a repository of known threats. Threat identification is based on its name acronym: Spoofing, Tampering, Repudiation, Information disclosure, Denial of service, Elevation of privilege. Thus, security experts can break down and analyse in depth potential risks, processes and flows. Although it provides the user with checklists and methodologies, it cannot represent any security architectural decisions and it is time consuming.

Another threat model is LINDDUN, focusing on software architecture and data security [WJ15]. LINDDUN stands for Linkability, Identifiability, Non-repudiation, Detectability, Disclosure of information, Unawareness, Non-compliance. The model uses data flow diagrams to represent the system's overall state. Constant analysis of the system threats helps to identify its weaknesses. LINDDUN achieves that through workflow iterations. A

powerful identification tool is the rich knowledge-base and its documentation [WSJ14]. The drawbacks are the high workload and time consumption that the specific model needs.

The National Institute of Standards and Technology provides an additional threat model, the Common Vulnerability Scoring System (CVSS) [SM09]. A final numerical score depicts the severity of specific threats based on the assessment of vulnerability characteristics. An online calculator is available for the results [Ib11]. Three metric groups - base, temporal and environmental, play an important role during the recognition and analysis phase, determining the output score. Due to the calculation method, analysts doubt its accuracy. This is a major reason the final score values may vary, which results in combination with other existing threat models for better results [Ha13].

One of the oldest and most widely used threat models is the Attack Tree. Bruce Schneier presented that threat model in 1999, with the help of extensive tree diagrams [Sc99]. The root represents the purpose of the attack, while the leaves depict different ways. Every major target of an attack is broken down into more discrete roots. So, with the specific threat mode, the user iterates through the problem and breaks it down into more attack trees. This threat model focuses on experienced analysts, assuming they have cybersecurity experience and thus does not provide them with detailed documentation or guidelines.

A recently proposed threat model is the Persona non Grata (PnG) [Cl14]. Its goal is to profile the potential attacker and their skills. This helps to outline vulnerabilities and threats early and understand the attacker's possible motives and skills. While this approach fits into the agile methodology and it is easy to use, the threat identification is limited to a specific set of threat types [Me18].

Quantitative threat modelling is a hybrid method, combining features from attack trees, STRIDE, and CVSS [PMK16]. Analysts build attack trees based on the five STRIDE categories and afterwards they apply the CVSS method to compute the respective scores for the tree components. This method focuses more on cyber-physical systems and risks with complex dependencies among their components.

Trike is an additional threat model, launched as a security audit framework in 2005 [Me18]. It aims more in the defensive side of a system. Analysts need an overview of the system's environment and create a requirement model. For the risk assessment, Trike normalizes the likelihood of the respective risk occurrence. The downsides, however, are lack of documentation and doubt of its calculation accuracy.

In 2003, the CERT division presented its own method, OCTAVE (Operationally Critical Threat, Asset, and Vulnerability Evaluation) [Al03]. Strategic evaluation of potential organizational risks is its primary focus. As this model fits larger organizations, it assesses operational risks and activities, not processes. Using this model, however, is a long process and its documentation is lengthy.

3 Introduction of the PASTA Threat Model

Threat models create an abstraction of the system to get a better understanding of the complexity of a system in order to find security flaws and potential attack vectors [Sh18]. Provided that, system architects and developers can respond to design failures and improve the system security [MLY05].

PASTA is a risk-centric threat modelling framework, proposed by Tony UcedaVélez in 2012 [UM15]. It includes different layers of abstraction and considers high-level layers, like the business logic, as well as the lower layers, like the concrete attack vectors. Especially in the IoT sector, a good understanding of abstract system architecture and business objectives is needed. The specific model fills the gap between technical and business risk analysis related to cyber threats. It provides analysts with an overview of manageable threats and risks to the addressable application environment. Thus, PASTA is the proposed threat model for IoT projects. The PASTA threat model consists of seven (7) stages, as shown in Figure 1, and can be applied to already existing projects or during the definition phase of a new project. We define the mapping of these seven stages of the PASTA Model to the IoT environment in the following chapter.

Fig. 1: The seven (7) stages of PASTA [UM15].

4 Applying the PASTA model to the IoT-ecosystem

Defining goals is the first stage in PASTA. Understanding business and its pain points is the basis for understanding and valuing risk. The system's main goals are set, and key services and target markets are specified, resulting in security and compliance requirements. Stride's security compliance can be a starting point to identify and address threats and risks. Business impact analysis to assess the importance of specific services, applications or devices can be based on ISO 22301 methodology. By defining maximum interruption, recovery time and priority, risk can be classified and prioritized from a business perspective.

The definition of the technical scope is the second stage providing the initial attack surface. Component and network diagrams reveal network design flaws like single failure points. For the respective devices, a detailed definition of accessibility, hardware, firmware, operating systems, and executed tasks, helps to identify device specific security requirements and vulnerabilities. Special attention is paid to unused communication and network interfaces, to identify side channels for an attack. Also, it is important to track hardware extensions, such as additional external modules. Technology enumeration, divided into architectural layers, outlines better the technology used. Specifically, a sub-categorization of the operating systems, programming languages, software libraries and communication protocols, ensures comprehensive technological enumeration.

Application decomposition is the next stage, defining system applications (units of deployment) and services (functional elements).The versatility and complexity of IoT networks carries the risk of not covering all applications and services. To reduce this risk, a detailed use-case diagram is essential to keep the overview of its actors, roles and features. The use case diagram shows unused features that can be eliminated or deactivated to reduce the attack surface. It also shows the intended use of the services to identify patterns to distinguish between intended and unintended use. If a feature has poor security performance the use cases indicate if we must redesign or secure it or if we can just eliminate it. Since correct and trustworthy data is the key to an IoT system, a data flow diagram (DFD) gives an important alternative view of the system and its vulnerability. The DFD in Figure 2, includes all data entry and exit points from sensors and actors via manual data entry screens to interfaces. Colours are used to denote the trustworthiness of an item (green: trustworthy) (orange: less trustworthy) (red: not trustworthy) (black: unknown) [AWT07]. The red-dotted circle wraps the trusted boundaries. This annotated DFD analyses the possible impact of incorrect or manipulated data and gives a data driven indicator for the need to secure a component or validate their data before using to increase trust.

In the fourth stage the threats of the respective system are identified. This threat analysis the cornerstone of the model and leads to a customised threat library [UM15]. Such a library is a collection of threats that may harm the overall system functionality. The first step is to be aware of all attack targets, the attack surface. Targets are the technical and functional components identified in stage 2 and 3. Attack targets are classified by likelihood of an attack at that target which is influenced e.g. by the popularity of the component, the outside

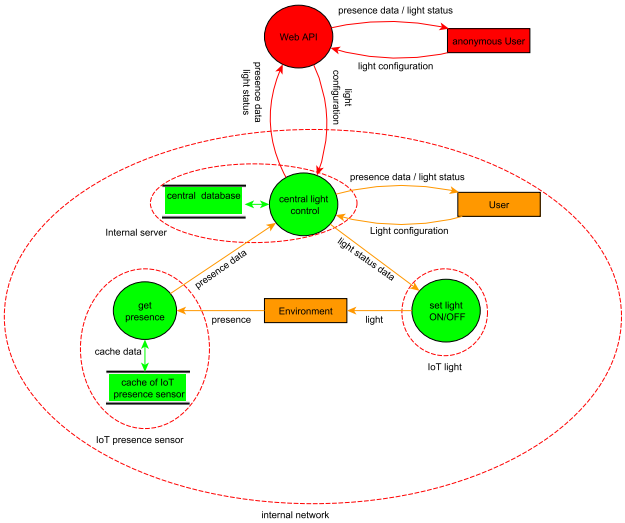


Fig. 2: Representation of an IoT light control system with different trust boundaries.

visibility or a history of known attacks. A second classification is based on business impact analysis. To provide a well-structured threat library, threats are categorized e.g. starting with the categories of the STRIDE model [KG99]. We introduce extortion and physical attacks as important threat categories. In addition to commonly known threats the field of IoT adds new threats like sensor spoofing and battery drain attacks. Typically IoT devices have constrained computational power, less memory and a limited battery capacity. Challenging such limitations is a typical strategy for attackers to find new threats [HFH15]. The list of categories is not static. It must be extended if new IoT system concepts get introduced, that lead to totally new types of threats.

Vulnerability and weakness analysis are the fifth stage. The main task is to enumerate the system vulnerabilities on a technical level using Common Weakness Enumeration (CWE). A comprehensive enumeration requires a vulnerability analysis of any used technology. Public vulnerability enumerations facilitate that task. Existing technologies are often adapted. Thus, security experience with standard Internet technology can often be transferred to the IoT environment. For instance, the Constrained Application Protocol (CoAP) and the Hypertext Transfer Protocol (HTTP) share similar concepts and vulnerabilities [Sh14]. Already created network and data flow diagrams offer insightful information to assess vulnerabilities in system architecture. The abuse-case diagram maps a threat to a component of the system. That extends the normal UML use-case diagram creation with abuse-case shapes, showing how a system is used by an attacker. Abuse-case diagrams create the link between a threat and an applied vulnerability to the system. To adapt it to IoT, IoT devices are considered actors in the Figure 3.

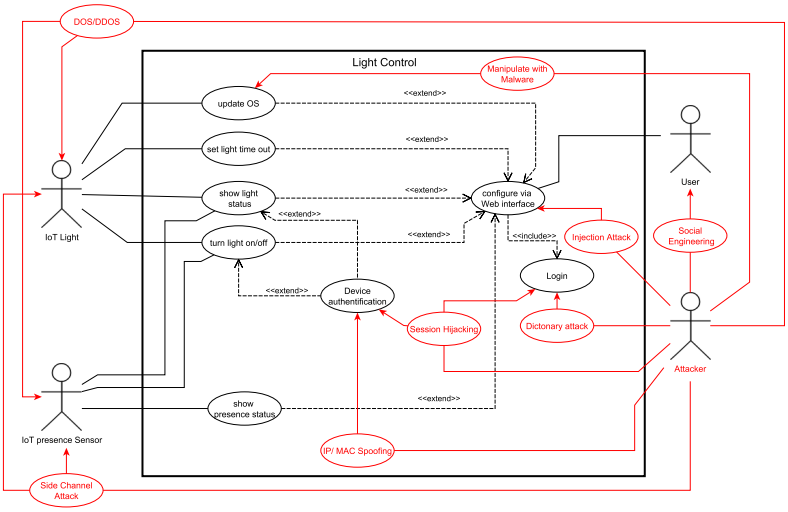


Fig. 3: An abuse-case diagram of a light control system.

The sixth stage, attack modelling, applies vulnerabilities to the system. Attack trees help to visualize discrete vulnerabilities into tree forms [Sc99]. To define the main targets of attack trees, the threats from the threat library are applied to the attack targets. Additionally, the abuse diagrams provide a good overview of which attacks can be executed on the system components. At the end of this task, the attack trees show how different vulnerabilities are used to pose a threat to the system.

The final stage is residual risk analysis. The system's most likely threats and vulnerabilities are evaluated to find countermeasures, mitigate, and detect threats. The use of CVSS (Common Vulnerability Scoring System) or CWSS (Common Weakness Scoring System) helps to assess the relevance and impact of the vulnerabilities in the enumeration. CVSS/CWSS is not optimized for the IoT environment. However, there are promising approaches of IoT-optimized risk assessments like an IoT risk assessment of CyVaR [Ra18]. Based on the threat, appropriate countermeasures must be taken Attack detection, mitigation or prevention. Mitigation is used to reduce attack damage and can be performed automatically by the system after attack detection or manually by isolating or deactivating an attacked or compromised target. Prevention methods often involve a performance trade, but risk analysis provides information needed to decide on it. Specifically, for high-risk threats performance-expensive prevention is acceptable. Countermeasures should be assigned to corresponding threats in the threat library.

DevSecOps is the methodology to improve the software development life cycle by combining development, security and operations. It extends DevOps to integrate security into the agile development process [MC17]. Security analysis in combination with the PASTA threat

model ensures high threat identification accuracy. That gives us the ability to verify the correct handling of those risks through testing. Regression testing has is a key component of the DevOps tool chain. Conventional regression tests only ensure the functionality of the business logic and seldom non-functional requirements like response time or throughput. A DevSecOps tool chain should be able to perform evidence-based and reproducible security regression tests, to continuously guarantee robustness and security of the IoT system in an agile development process. Unlike normal unit tests, security tests do not isolate individual system functions. Full system simulation is needed to observe possible side effects and simulate a realistic scenario. Security tests are derived from the security and compliance requirements defined in the PASTA threat model. A security test failure indicates a possible vulnerability and an uncovered non-functional requirement. Like functional tests, the implemented threats of the threat library can be classified according to criticality do detect the frequency in which they are performed.

To prove our concepts, we created a prototype IoT system, based on the RIOT operating system. RIOT, which is the basis for DoRIoT, is well suited because of the native ability to run as a virtual instance. Virtualisation allows us to do industry standard two-stage validation based on Software in the Loop (SIL) and Hardware in the Loop (HIL). The system contains two applications for sensor and actuator. The latter allows the internal LED to be switched on or off, when the sensor triggers it via a message across the network. Networking is based on CoAP over UDP IPv6. We first implemented the prototype, then applied code- and memory-analysis tools to identify code-level weaknesses and then applied the threat model to the prototype. All checks are done in an extended continuous integration pipeline that automatically creates a reference system without attacks and test systems, that include active attackers. Thus, we can compare normal system behaviour with attacking system behaviour. As expected, the threat model revealed system threats and weaknesses. Vulnerabilities were detected in the network and service layers, such as Vulnerability to DOS/DDOS attacks and ping flooding that could bring down the full RIOT OS. Tests could be repeated automatically to verify results and to show the effects of changes in the prototype.

5 Conclusion

IoT opens a wide range of possibilities but also poses new threats. The more IoT devices are integrated into our lives, the greater is the demand for security. In this paper we analysed and evaluated different threat models to determine which one suits IoT best. In our investigations, the PASTA threat model seemed the best approach. We improved the PASTA threat model by applying existing research and industry standards like the ISO 22301 for the Business Impact Analysis. We gained a hardware perspective by adding a network diagram and considering IoT specific properties like accessibility and hardware limitations. We modified the Data Flow Diagram by adapting the colour coding to highlight the trustworthiness of a data source based on location and reliability. To adopt the threat library, we evaluate new threat categories. The technology-enumeration and the network diagram of stage two create the IoT context to the threat library.

The adapted threat model made it possible to identify threats and vulnerabilities of an IoT system, and to assess the evidence-based risks. This methodology improves the security of the development life cycle and of existing IoT systems.

Based on our research, we created a concept for security tests based on threat model theory and data. This concept fulfilled all the conditions for the fast and agile development life cycle of DevOps, contributing to DevSecOps. To prove the concept, we created an IoT prototype system and the implementation of the respective security tests. A future objective is to gain experience in implementing the concept to broaden and improve this methodology.

Test results proved that system vulnerabilities exist and must be detected and fixed. The integration into the continuous integration pipeline proved that security tests can be implemented as automated regression tests. We could show that a threat model, based on a DevSecOps model, improves the security of an IoT system and does not restrict the agility and efficiency of an IoT project.

The presented results are part of the DoRIoT research project [20].

References

[20] Dynamic runtime for organically (dis-)aggregating IoT-processes, Aug. 2020, URL: http://doriot.ovgu.de/.

[Al03] Alberts, C.; Dorofee, A.; Stevens, J.; Woody, C.: Introduction to the octave approach. Pittsburgh, 2003.

[AWT07] Abi-Antoun, M.; Wang, D.; Torr, P.: Checking threat modeling data flow diagrams for implementation conformance and security. In: Proceedings of the twenty-second IEEE/ACM international conference on Automated software engineering. Pp. 393–396, 2007.

[Cl14] Cleland-Huang, J.: How well do you know your personae non gratae? IEEE software 31/4, pp. 28–31, 2014.

[Ha13] Hanford, S.: Common vulnerability scoring system, v3 development update. In: Technical report, Forum of Incident Response and Security Teams (FIRST). 2013.

[HFH15] Hossain, M. M.; Fotouhi, M.; Hasan, R.: Towards an analysis of security issues, challenges, and open problems in the internet of things. In: 2015 IEEE World Congress on Services. IEEE, pp. 21–28, 2015.

[Ib11] Ibidapo, A. O.; Zavarsky, P.; Lindskog, D.; Ruhl, R.: An analysis of CVSS v2 environmental scoring. In: 2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing. IEEE, pp. 1125–1130, 2011.

[Ke98] Kelsey, J.; Schneier, B.; Wagner, D.; Hall, C.: Cryptanalytic attacks on pseudorandom number generators. In: International Workshop on Fast Software Encryption. Springer, pp. 168–188, 1998.

[KG99] Kohnfelder, L.; Garg, P.: The threats to our products. Microsoft Interface, Microsoft Corporation 33/, 1999.

[MC17] Myrbakken, H.; Colomo-Palacios, R.: DevSecOps: a multivocal literature review. In: International Conference on Software Process Improvement and Capability Determination. Springer, pp. 17–29, 2017.

[Me18] Mead, N. R.; Shull, F.; Vemuru, K.; Villadsen, O.: A Hybrid Threat Modeling Method. Carnegie MellonUniversity-Software Engineering Institute-Technical Report-CMU/SEI-2018-TN-002/, 2018.

[MLY05] Myagmar, S.; Lee, A. J.; Yurcik, W.: Threat modeling as a basis for security requirements. In: Symposium on requirements engineering for information security (SREIS). Vol. 2005, Citeseer, pp. 1–8, 2005.

[MM19] Mullen, G.; Meany, L.: Assessment of Buffer Overflow Based Attacks On an IoT Operating System. In: 2019 Global IoT Summit (GIoTS). Pp. 1–6, 2019.

[PMK16] Potteiger, B.; Martins, G.; Koutsoukos, X.: Software and attack centric integrated threat modeling for quantitative risk assessment. In: Proceedings of the Symposium and Bootcamp on the Science of Security. Pp. 99–108, 2016.

[Ra18] Radanliev, P.; De Roure, D. C.; Nicolescu, R.; Huth, M.; Montalvo, R. M.; Cannady, S.; Burnap, P.: Future developments in cyber risk assessment for the internet of things. Computers in Industry 102/, pp. 14–22, 2018.

[Sc99] Schneier, B.: Attack trees. Dr. Dobb's journal 24/12, pp. 21–29, 1999.

[Sh14] Shelby, Z.; Hartke, K.; Bormann, C.; Frank, B.: RFC 7252: The constrained application protocol (CoAP). Internet Engineering Task Force/, 2014.

[Sh18] Shevchenko, N.; Chick, T. A.; O'riordan, P.; Scanlon, T. P.; Woody, C.: Threat Modeling: a Summary of Available Methods. no. July/, 2018.

[SM09] Scarfone, K.; Mell, P.: An analysis of CVSS version 2 vulnerability scoring. In: 2009 3rd International Symposium on Empirical Software Engineering and Measurement. IEEE, pp. 516–525, 2009.

[UM15] UcedaVélez, T.; Morana, M. M.: Risk centric threat modeling. Wiley Online Library, 2015.

[WJ15] Wuyts, K.; Joosen, W.: LINDDUN privacy threat modeling: a tutorial. CW Reports/, 2015.

[WSJ14] Wuyts, K.; Scandariato, R.; Joosen, W.: LIND (D) UN privacy threat tree catalog. CW Reports/, 2014.

R. Reussner, A. Koziolek, R. Heinrich (Hrsg.): INFORMATIK 2020,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2021 1205

Enhancing Resilience in IoT Networks using Organic Computing: Challenges and Requirements

Dominik Weikert[1] Christoph Steup[2] Sanaz Mostaghim[3]

Abstract: In this paper we present and analyse the requirements and challenges of the Task Allocation Problem for Internet-of-Things (IoT) Networks, especially Wireless Sensor Networks (WSNs). As IoT is comprised of a variety of heterogeneous devices and network configuration may change regularly due to low-power nodes failing or communication disruptions, a static allocation of tasks to individual nodes cannot be assumed. Therefore, task allocation has to be carried out and managed for every dynamic change in the network along the lifetime of the network. In dynamic task allocation, a NP-hard problem, the calculation of a new optimal allocation could quickly become a bottleneck for network performance, giving rise to the need of organic computing solutions to provide self-organised task allocation solutions for such networks.

Keywords: IoT; Organic Computing; Dynamic Task Allocation

1 Introduction

The Internet-of-Things (IoT) and especially Wireless Sensor Networks (WSNs) are comprised of multiple potentially low-power nodes distributed throughout the environment to monitor, process and manipulate. Nowadays this technology is already used in many application fields like collaborative monitoring [Yu17], traffic control [Ka14], agricultural irrigation [Ou14], intelligent transportation management [Ph10] and environmental monitoring [ZLH13]. The dynamic nature of these applications requires flexible, but reliable distribution of tasks to nodes within the network. This requires reliable reallocation on failures and dynamic optimization of performance on environmental changes. The methodologies in the context of Organic Computing (OC) [MSSU11] offer means to enable controlled self-organisation for IoT applications. OC provides the capability to deal with a large number of entities such as sensors that can work in a self-organised way but at the same time can be controlled and adapted to the environmental changes. One major aspect in OC is the dynamic optimization which involves observation and monitoring the current state of the system and adaptation to possible changes.

This paper aims to provide theoretical foundations for the integration of OC with IoT as the system under observation and control. In contrast to classical OC applications, WSN and

[1] Otto-von-Guericke University Magdeburg, Faculty of Computer Science, Germany, dominik.weikert@ovgu.de
[2] Otto-von-Guericke University Magdeburg, Faculty of Computer Science, Germany, steup@ovgu.de
[3] Otto-von-Guericke University Magdeburg, Faculty of Computer Science, Germany, sanaz.mostaghim@ovgu.de

doi:10.18420/inf2020_112

mobile IoT systems are limited in energy [PAG09]. This induces a new major challenge in the maximisation of the lifetime of the overall system. Therefore, incorporating energy-awareness to the OC components is crucial, especially in use-cases where recharging or replacing the individual nodes or their batteries is infeasible. Consequently, energy efficiency is the key factor in nearly all applications of WSN technology if these want to be adopted by real world users. Typically, energy is drained by four components of each node, the processor, sensors, actuators and the communication system. Among these, the communication is generally the largest energy consumer [Ra02]. Multiple solutions may provide partial remedies for the energy consumption of the communication. The communication can be made more efficient by minimising waiting times and limiting the active time of the receiver. Approaches to optimizing the communication itself have already been intensively studied and may depend on specific hardware requirements. However, a more global optimisation approach could provide additional benefits. Thus, another solution is the optimisation of the allocation of tasks to nodes of IoT Network, especially in multi-hop networks. This allows to minimise communication in the whole network by putting processing in line between data sources and data sinks. Additionally, it allows to balance the energy consumption of the nodes to prevent early battery drainage of single nodes, which may result in a partition of the IoT network into multiple disconnected sub-networks. Unfortunately, this problem is NP-Hard and needs advanced algorithms to be efficiently performed within the network. Classical task allocation algorithms may provide good solutions, but at costs, which already drain the nodes before the task even started.

In this paper we provide an in-depth analysis of the task allocation problem in IoT networks. Our goal is to analyse existing state-of-art solution regarding the challenges of IoT as well as OC and establish a formal model for evaluation and analysis of task allocation methods. Additionally, we review and classify the various methods that have been used to tackle this problem in the literature. The paper is structured as follows: In Section 2, we give a complete analysis of the challenges involved in the task allocation problem. Afterwards, Section 3 provides an overview of the state-of-the art algorithms. We formally describe the model for task allocation problems within IoT networks in Section 4 and conclude the paper in Section 5.

2 Challenges

The task assignment problem as an instance of the generalised assignment problem, which is known to be NP-hard [Yu14]. In its most basic form, we consider a network of homogeneous nodes to which a set of independent tasks are assigned. Given N_N nodes and a set of N_T tasks, the total number of possible assignments is $N_T^{N_N}$. Without exploiting additional knowledge on the system, a direct calculation of the optimum allocation is prohibitively expensive for all but even the smallest networks. Additionally, this simplistic model is not applicable for real-world networks. Therefore, the following section analyses additional challenges for IoT networks:

Task Requirements The basic problem stated above does not consider any requirements imposed by the tasks themselves. However, in reality multiple requirements may be imposed by certain tasks. The first issue is that not all tasks are independent from each other: A sensing task may have to be performed and the measurement transmitted before that information can be processed somewhere else in the network. These dependencies need to be incorporated into an optimal task allocation, ensuring sufficient and timely information flow to nodes which have to carry out tasks depending on this information. In addition, tasks can have specific constraints, such as a deadline by which a set of tasks has to be finished, which will have to be ensured during task allocation. Tasks will have to be scheduled in a way that the deadlines of all tasks are kept. Critical task sets would have to be allocated to the most reliable nodes in the network, imposing further constraints on task allocation. Additionally, tasks may have a specific sensor or actuator requirement at a specific location, requiring availability of a node capable of satisfying such task constraints.

Heterogeneity of Nodes Typical networks consist of nodes with varying processing powers, battery reserves, communication capabilities. Furthermore, each node may have only specialised sensors, such as temperature or humidity sensors. As such, no one-fits-all solution can be implemented and the task allocation algorithm must incorporate the knowledge about differing node capabilities. Furthermore, node capabilities may deteriorate over time as energy levels begin to decrease or environmental conditions change. In theory, bandwidth and topology of the network can be easily determined, but the actual values in practice may greatly differ from this and even differ at different times during network operation. As such, the task allocation algorithms need to incorporate the current state, capabilities of each node, and their link quality to their neighbours, in order to provide optimal solutions.

Dynamic Network Conditions Dynamics in networks typically result from failures or mobility of nodes. **Failures** can lead to node or link loss changing the network structure. With the limited power capacity of nodes, it can be assumed that nodes fail during sufficiently long operation times. Additionally, new nodes could be added to extend the capacity of the network during run-time, giving new options for more efficient task allocations. Furthermore, nodes may not be positioned at the same coordinates for the entire duration and may move either due to environmental or other mobility factors. In addition, there may be temporary node outages or disruptions in communication inducing transitional failures. Overall, any single node failure can have extensive repercussions for network structure due to the multi-hop communication model. A Node failure may not only impact the tasks allocated to the failing node, but also other tasks with communication paths incorporating the failed node. As a result, some communication paths may have to be rerouted whenever a node fails, resulting in increasing and shifting task loads to other nodes in the network. **Mobility** of nodes is another aspect causing change in the network structure and possibly invalidating the task allocation. The mobility of nodes is typically intrinsic either because the IoT devices are mobile or due to external forces such as environmental influences. Moving nodes may influence the network communication by disconnecting some links due to increased range and facilitating new links with nodes close to their new positions. Additionally, spacial

constraints of tasks involving sensing or acting may be violated because of the movement. Given these dynamic conditions, even an optimal task allocation may only stay optimal for a very short duration of the lifetime of the network and re-allocation is required to be carried out multiple times creating time-varying task allocations. Consequently, efficient and reliable task allocation algorithms that go beyond classical approaches are needed. Network dynamics due to node failures is a well-studied problem, but only few works consider the challenge imposed by node mobility. Jin et al. [Ji13] placed particular focus on network dynamics including node mobility, ensuring an up-to-date solution could always be used when changes of the network structure occur.

Load Balancing While achieving high network lifetime for a set of tasks is essential, an aspect that should not be ignored is the balancing of task load among all nodes in the network. As task requirements may change over time or new sets of tasks may be introduced, it is very beneficial to keep a variety of nodes available. In this way, the load can be distributed which can ensure a balanced energy level throughout the network. Load balancing on processing and energy consumption are among the well-studied goals for task allocation and the focus of several works in the literature e.g., [KOT18, ETX12, Gu15].

Quality of Service (QoS) Given the above challenges and especially the node failures, it cannot be assumed that communication always is perfect. With wavering link quality between nodes, transmissions might fail or be incomplete and data might have to be resend, increasing the communication costs by factors of two or more as messages might have to be retransmitted multiple times. Therefore, the link quality between pairs of nodes cannot be ignored, as it can have a significant impact on the network lifetime. An additional factor is the quality of the provided sensing, processing and actuating tasks. By assigning the task to multiple nodes, service quality can be improved at the cost of increased energy requirements. This can be measured by the percentage of lost packages, accuracy of sensed information, speed of processing and the ratio between successful and failed actuating tasks. Zhang et al. [Zh19] present an approach especially focused on ensuring the reliability of the task allocation.

Task Allocation Efficiency Any allocation of tasks will drain the resources of the network as tasks are moved from one node to another. This requires communication and processing for all involved nodes and has to be incorporated into the network model. As such, an efficient mechanism to redistribute tasks needs to be developed. Otherwise, any performance gained through optimization may be unmade by the cost of applying the new allocation to the network. OC provides means to determine these allocations and decide on the reallocation of tasks. Efficient algorithms nee to minimise the overall reallocation cost by either reallocating seldom or by only reallocating parts of the tasks.

Network Observation Quality One of the major strengths of OC is to monitor the system and control whenever necessary. Therefore, network observation plays an important role which enables us to keep track of all the above aspects for the task allocation. Usually, task allocation algorithms depend on having as much and as accurate information as possible to make intelligent decisions about the best options within the network. However, collecting

Tab. 1: Overview of modelling completeness of state-of-the-art approaches

Reference	Type	Dynamic	Balance	Hetero-geneity	QoS	Obser-vation	Task model	MOO
[Ji13]	GA	++		+		++	++	
[HX11]	GA	+			+		+	
[KOT18]	GA	+	++	+			+	+
[Gu15]	PSO	+	++		++		+	
[Ya14]	PSO	+	++	+			+	
[Yi17]	Consensus	+	+	++		++	+	
[ETX12]	Auction	+	++			+	+	
[CPA14]	Consensus	+	++	++	+	+	+	
[XZ20]	ACO	+		+	+		+	
[SLT18]	PSO	+	+			+	+	++
[Zh19]	PSO	+			++		+	

information about each node with high frequency leads to highly increased network load, as every node would need to relay its current status to a dedicated observer node or broadcast it in the network for distributed consensus. This keeps the network information completely up-to-date and reliable, but drains all nodes of energy. This runs contrary to the goal of increasing network lifetime by performing optimisation of the task allocation. With that in mind, perfect information cannot be assumed and an appropriate algorithm needs to incorporate uncertainty of network observations in the decisions making of the OC mechanisms.

Multi-Objective Optimisation (MOO) When combining the previous challenges of the dynamic network structure, load balancing, observation quality and allocation cost, it quickly becomes clear that trade-offs are necessary between some of the goals of the optimisation. As such, the task allocation problem needs to be formulated as a Dynamic Multi-objective Optimisation Problem (DMOP): Network lifetime, energy level or consumption balancing, reliability and quality of information may be maximised while minimising the cost to find an optimal allocation and execute it on the network.

3 Classification of the state-of-the-art approaches

Considering the above challenges, we provide an overview and a classification of the state-of-the-art approaches from the literature. Table 1 provides a general overview of the recent works in the literature and gives a summary of the incorporated aspects, which map directly to the challenges described in Section 2. A complete handling of a challenge in

Tab. 2: Overview of metrics coverage in state-of-the-art approaches

Metric	Evaluated in
Latency L	[Ji13, Gu15, Ya14, Yi17, ETX12, SLT18]
Load Balance C	[KOT18, Ya14, Yi17, ETX12, SLT18]
Energy Consumption E	[HX11, KOT18, Gu15, Ya14, Yi17, ETX12, SLT18, Zh19]
Reliability R	[Gu15, Zh19]
QoI I	[HX11, CPA14, SLT18]
Network Lifetime NL	[Ji13, KOT18]

the model used by the respective authors is denoted by (++) while a partial incorporation or evaluation is denoted by (+). A missing entry means that the particular aspect was disregarded. We additionally specify the generic type of optimisation algorithm being used. As visible, Genetic Algorithms (GA), Particle Swarm Optimisation (PSO) and Ant Colony Optimisation (ACO) are the most frequently used meta-heuristics. Additionally, the table shows that Consensus-based methods can help to deal with the observation challenge, while Auction-based methods help to increase the load balance in the network. Nevertheless, it becomes clear that most of the existing approaches only study certain aspects of the task allocation problem in detail while neglecting others. To the best of our knowledge, no work has as of yet been provided which tackles the problem in its complete form.

Several works provide solutions to parts of the DMOP [KOT18, SLT18]. However, as of yet, no work has been put forth that provides a complete solution to all of the defined objective functions (refer to Section 4). Most of the other presented works combine a selected number of objectives into a single objective function using a weighted sum with arbitrarily chosen weights. Table 2 shows the coverage of the various objective functions among the works evaluated in Table 1.

4 Formal Problem Description

In this section, we establish a formal model for the DMOP as described in Section 2. In particular, we develop a generic model for task and network structure of a wide range Iot or WSN networks while incorporating the challenges and providing mathematical definitions for the optimisation metrics.

The goal of Task Allocation for IoT networks is to distribute a set of tasks $T_i \in V_{Task}$ to a set of nodes $N \in V_{Nodes}$. To represent the dependencies between the tasks a directed acyclic graph (DAG) $G_{Task} = (G_{Task}, E_{Task})$ is used, as it models relevant problem aspects with little loss of generality. In this model, each vertex of the graph is a task T_i and each edge in the DAG $e_{ij} \in E_{Task}$ represents a directional dependency between the two tasks T_i and T_j. Additionally, weight values w_{ij} are assigned to each edge, which represents the data

communication cost between the two tasks if these tasks are executed on different nodes. Another weight value q_i is attached to each vertex, which represents the processing cost of the task T_i. Finally, a task may be assigned additional spatial constraints S_i to specify where in the network it may be executed. In this model, we consider three distinct task types:

Sensing Tasks create information to be relayed to other nodes. These do not possess predecessors in the DAG. Due to the raw sensor information, these tasks typically have either high communication or high processing costs.
Processing Tasks possess both predecessors and descendants and have typically high processing costs and no spatial constraints.
Relaying Tasks are internal tasks automatically created to forward information and have typically high communication costs and no spatial constraints. These tasks can either be generated by the task allocation or created implicitly by a routing protocol.
Actuating Tasks only possess predecessors typically incurring low processing costs.

An example for such a network with two sensing tasks, one processing task accumulating the information from both sensing tasks and one actuating task is shown in Figure 1.

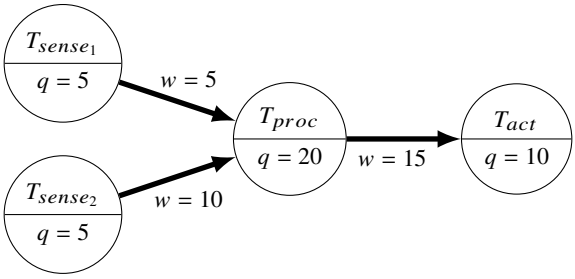


Fig. 1: Basic task graph example

The IoT network is modelled as an undirected graph $G_{Net} = (V_{Nodes}(t), E_{Com}(t))$. Each node N_i is associated with a battery and its specific amount of energy $E_i(t)$ at time t as well as its position $\vec{x}(t)$. Two nodes can communicate if there is an edge $e_{ij} \in E_{Com}(t)$. The set of edges and vertices is time-dependent due to the dynamic nature of the network as nodes and links between nodes may transiently or permanently fail. Each edge is associated an energy consumption $E_{ij}(t)$ representing the amount of energy necessary to forward one unit of information and a latency value $L_{ij}(t)$ modelling the current latency when transmitting data along this edge. Additionally, a function for each node is assumed, which transforms processing cost to latency $L_i(c_j)$.

The problem of task allocation is to find an allocation of tasks to nodes. Formally, this allocation is an injective function $A : V_{Task} \rightarrow V_{Nodes}$, which assigns each vertex in the task graph G_{Task} to a non-empty subset of vertices N_{T_i} of the network $N_{T_i} \subset G_{Net}, N_{T_i} \neq \emptyset$. An allocation is valid iff for each Task T_i in the Task-Set V_{Task}, there is at least one node $N_j \in V_{Nodes}$ allocated, which fits the tasks' sensor and actuator requirements. Additionally, the node N_j needs to be connected through communication edges $e_{jl} \in E_{com}$ with all

nodes $N_l = A_t(T_k)$ executing directly connected tasks $e_{ik} \in V_{Tasks}$. Since the Network may change its structure over time and communication conditions are unlikely to remain the same, the goal is to find a series of valid allocations $A = A_0, A_1, \ldots, A_n$ with associated start times t_i^{start} and end times t_i^{end}. For each time t the respective allocation is defined as $A_t = A_i, t_i^{start} \leq t \leq t_i^{end}$. Overall the allocation series shall maximise Network Lifetime $NL(A)$, Network Energy Balance $C(A)$, Reliability $R(A)$ and Quality of Information $I(A)$ while minimising Latency $L(A)$ and Energy Consumption $E(A)$.

The Network Lifetime $NL(A)$ is defined as the maximum time where a valid Allocation A_i exists:

$$NL(A) = \max(t_i^{end}), \text{where } A_i \text{ is valid} \tag{1}$$

The latency L_{ij} for each connected pair of nodes i, j is given by the edge weights L_{ij} connecting them. The latency between dependent tasks T_k, T_l can thus be defined as the sum of all edge weights along the path $P_{kl}(A)$ plus the latency $L_k(q_i)$ generated through processing of Task T_k on Node $N_i = A_t(T_k)$ on their assigned nodes. The latency $L(A_t)$ of the allocation A_t is defined as the maximum latency of all connected sub-tasks, see Equation 2.

$$L(A_t) = \max_{T_k, T_l \in V_{Tasks}} \sum_{e_{ij} \in P_{kl}(A_t)} L_{ij} + L_i(q_k) \tag{2}$$

The load balancing in the network directly relates to the distribution of energy consumption between the nodes. This energy consumption distribution can be calculated as shown in Equation 3, where $E_j(A_t)$ is the total energy consumption of Node N_j based on the assigned tasks and $\overline{E}(A_t)$ is the estimated average energy consumption for all the nodes at time t.

$$C(A_t) = \sqrt{\frac{1}{N_N} \sum_{N_i \in V_{Nodes}} \left[E_i(A_t) - \overline{E}(A_t) \right]^2} \tag{3}$$

The energy consumption of the task set $E(A_t)$ at time t is defined as the sum of the energy consumption of all nodes in the network:

$$E(A_t) = \sum_{N_i \in V_{Nodes}} E_i(A_t) \tag{4}$$

For the reliability of the communication links, we refer to Zhang et al. and their work on a reliable Task Allocation Method [Zh19]. The proposed reliability metric $R(A_t)$ incorporates both the reliability of nodes and the links between them. The reliability for a task allocation is then calculated by combining the reliability of all assigned nodes and all transmissions along the network graph.

Quality of Information (QoI) can be measured through the amount and quality of sensors providing information for a sensing task. A simple approach is to define QoI $I(A)$ as the sum of the quality of the sensors I_j of the nodes N_j for each sensing task T_i, see Equation 5.

$$I(A_t) = \sum_{T_i} \sum_{N_j \in A_t(T_i)} I_j \tag{5}$$

The different objective functions need to be combined with the network and task specification to the full Dynamic Multi-Objective Optimisation Problem (DMOP), see Equation 6.

$$\begin{aligned} &min\,(L(A), C(A), E(A), -NL(A), -R(A), -I(A)) \\ &s.t.\ \forall t \in [t_0, NL(A)], A_t \text{ is valid} \end{aligned} \tag{6}$$

With all the above elements incorporated, the resulting DMOP combines all aspects of the state-of-the-art literature into a unified model capable of representing a wide range of Task Allocation Optimisation Problems and the associated challenges.

5 Conclusion

This paper provides an overview of the challenges in IoT systems incorporating OC mechanisms to enable resilience and performance through online optimisation. We have addressed the OC characteristics with IoT features regarding energy consumption, which imposes the major limitation for both approaches. The task allocation problem for IoT networks imposes a variety of challenges to overcome. In its entirety, it represents a Dynamic Multi-Objective Optimisation Problem (DMOP), which poses a major challenge itself. We propose a model to formally evaluate and analyse algorithmic solutions for this DMOP, which is general enough to cover most conceivable networks and task structures. For future work, it is necessary to develop and evaluate algorithms which tackle the full DMOP using the proposed model and compare their results and performance with existing solutions in centralised and decentralised approaches. Due to the complexity of the problem, a framework of algorithmic building blocks needs to be developed to allow the adaptation of the algorithms to the practical problem to minimise the overhead in terms of communication, processing and reallocation.

Acknowledgement: This work is funded by the German Federal Ministry of Education and Research with project number 01IS18071A.

Bibliography

[CPA14] Colistra, G.; Pilloni, V.; Atzori, L.: The problem of task allocation in the Internet of Things and the consensus-based approach. Elsevier Comp. Net., 73:98–111, 2014.

[ETX12] Edalat, N.; Tham, C.; Xiao, W.: An auction-based strategy for distributed task allocation in wireless sensor networks. Computer Communications, 35(8):916 – 928, 2012.

[Gu15] Guo, W.; Li, J.; Chen, G.; Niu, Y.; Chen, C.: A PSO-Optimized Real-Time Fault-Tolerant Task Allocation Algorithm in Wireless Sensor Networks. Trans. Parallel Distrib. Syst., 26(12):3236–3249, 2015.

[HX11] Hu, Xiaoqing; Xu, Bugong: Task Allocation Mechanism Based on Genetic Algorithm in Wireless Sensor Networks. In: Applied Informatics and Communication. Springer, pp. 46–58, 2011.

[Ji13] Jin, Y.; Vural, S.; Gluhak, A.; Moessner, K.: Dynamic Task Allocation in Multi-Hop Multimedia Wireless Sensor Networks with Low Mobility. Sensors, 13(10):13998–14028, 2013.

[Ka14] Kafi, M.; Djenouri, D.; Ben-Othman, J.; Badache, N.: Congestion Control Protocols in Wireless Sensor Networks: A Survey. IEEE Commun. Surveys Tuts., 16(3):1369–1390, May 2014.

[KOT18] Khalil, E.; Ozdemir, S.; Tosun, S.: Evolutionary task allocation in Internet of Things-based application domains. Future Generation Computer Systems, 86:121–133, 2018.

[MSSU11] Müller-Schloer, C.; Schmeck, H.; Ungerer, T.: Organic Computing — A Paradigm Shift for Complex Systems. Springer, 2011.

[Ou14] Ouadjaout, A.; Lasla, N.; Bagaa, M.; Doudou, M.; Zizoua, C.; Kafi, M.; Derhab, A.; Djenouri, D.; Badache, N.: DZ50: Energy-efficient Wireless Sensor Mote Platform for Low Data Rate Applications. Procedia Computer Science, 37:189–195, 2014.

[PAG09] Penella, M.; Albesa, J.; Gasulla, M.: Powering wireless sensor nodes: Primary batteries versus energy harvesting. In: Instrumentation and Measurement Technology Conference. IEEE, pp. 1625–1630, 2009.

[Ph10] Pham, N.; Ganti, R.; Uddin, Y.; Nath, S.; Abdelzaher, T.: Privacy-Preserving Reconstruction of Multidimensional Data Maps in Vehicular Participatory Sensing. In: Wireless Sensor Networks. Springer, pp. 114–130, 2010.

[Ra02] Raghunathan, V.; Schurgers, C.; Park, S.; Srivastava, M.: Energy-aware wireless microsensor networks. IEEE Signal Process. Mag., 19(2):40–50, 2002.

[SLT18] Sun, Z.; Liu, Y.; Tao, L.: Attack localization task allocation in wireless sensor networks based on multi-objective binary particle swarm optimization. Journal of Network and Computer Applications, 112:29 – 40, 2018.

[XZ20] Xu, M.; Zhou, J.: Elite Immune Ant Colony Optimization-Based Task Allocation for Maximizing Task Execution Efficiency in Agricultural Wireless Sensor Networks. Journal of Sensors, 2020, 2020.

[Ya14] Yang, J.; Zhang, H.; Ling, Y.; Pan, C.; Sun, W.: Task Allocation for Wireless Sensor Network Using Modified Binary Particle Swarm Optimization. IEEE Sensors J., 14(3):882–892, 2014.

[Yi17] Yin, X.; Dai, W.; Li, B.; Chang, L.; Li, C.: Cooperative task allocation in heterogeneous wireless sensor networks. Int. J. of Distributed Sensor Networks, 13(10), 2017.

[Yu14] Yuan, M.; Jiang, C.; Li, S.; Shen, W.; Pavlidis, Y.; Li, J.: Message passing algorithm for the generalized assignment problem. In: Int. Conf. on Network and Parallel Computing. Springer, pp. 423–434, 2014.

[Yu17] Yu, J.; Wan, S.; Cheng, X.; Yu, D.: Coverage Contribution Area Based k -Coverage for Wireless Sensor Networks. IEEE Trans. Veh. Technol., 66(9):8510–8523, Sep 2017.

[Zh19] Zhang, D.; Hu, C.; Zhu, X.; Xu, R.: Reliable and Efficient Task Allocation Method in Wireless Sensor Networks. In: Genetic and Evolutionary Computing. Springer, pp. 461–470, 2019.

[ZLH13] Zheng, Y.; Liu, F.; Hsieh, H.: U-Air: When Urban Air Quality Inference Meets Big Data. In: 19th SIGKDD. ACM, pp. 1436—1444, 2013.

R. Reussner, A. Koziolek, R. Heinrich (Hrsg.): INFORMATIK 2020,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2021 1215

Combined Certificate and Resource Discovery for Dynamically (Dis-)Aggregating IoT Processes

Frank Engelhardt[1], Mesut Güneş[2]

Abstract: The concepts of Microservices and Organic Computing contribute to a fully distributed architecture of the Internet of Things (IoT), avoiding single points of failure through massive service distribution. The distribution and the lack of structure, however, come with large communication overheads. We discuss the necessity of structure in IoT networks focusing on the problem of trust handling, specifically analyzing the certificate chain discovery problem. Moreover, we provide an argument towards solving the certificate chain discovery problem in the same manner as the service discovery problem in a combined, semi-structured approach. By numerical analysis we show that the introduction of a hierarchy can avoid scalability problems and that resource directories used for service discovery can serve as hierarchical entities.

Keywords: Web of Trust; Certification; Internet of Things

1 Introduction

The modern, yet permanently evolving IoT imposes many changes to our everyday lives as the pervasiveness of intelligent gadgets enriches many sectors at once. The trust that we grow in smart things stems from the benefits that we gain from their intelligence as they improve services and systems that we interact with every day. But as computation becomes more ubiquitous and the interconnection between devices more complex, the verification of communication and computations run by smart devices becomes a huge problem. Trust in services is easy to compromise, especially when networks and systems become so complex that no single controlling instance has complete knowledge about every single system entity. Service aggregation using Organic Computing may play a key role in managing growing complexity in IoT systems [Ro16]. We speak of such IoT systems that use Organic Computing and Microservices to manage their intrinsic complexity as dynamically (dis-)aggregating IoT systems.

It is, however, obvious, that establishment of trust in such systems is far more complex than the verification and validation of each sub-component of an organic system. Communication and aggregation can also be compromised, such that the aggregation as a whole needs to be

[1] Otto-von-Guericke Universität Magdeburg, Communication and Networked Systems (ComSys), Faculty of Computer Science, Universitätsplatz 2, 39106 Magdeburg, Germany, frank.engelhardt@ovgu.de
[2] Otto-von-Guericke Universität Magdeburg, Communication and Networked Systems (ComSys), Faculty of Computer Science, Universitätsplatz 2, 39106 Magdeburg, Germany, mesut.guenes@ovgu.de

doi:10.18420/inf2020_113

trustworthy as well. Exchanging symmetric keys is not an option in huge networks, thus asymmetric encryption needs to be implemented in an efficient and scalable way.

In this paper, we investigate the certification process in IoT networks and discuss the necessity of structure to reduce communication complexity. Since in IoT devices are most commonly resource-constraint, storing certificates in order to establish trust networks is a problem that should be delegated to devices with bigger storage capacity. In numerical analysis, we show that the utilization of such resource directories enables the certificate chain discovery to be solved with reasonable overhead compared to a completely distributed approach.

The rest of this paper is structured as follows. Section 2 gives an introduction to the structure of future of dynamically (dis-)aggregating IoT processes. Section 3 introduces the trust chain management problem. Section 4 summarizes related work. In Section 5, we present a solution considering common communication approaches and the concept of resource directories. Section 6 contains a numerical analysis, and Section 7 concludes with a discussion.

2 Dynamically (Dis-)Aggregating IoT Processes

Microservice architectures are dominating in software development nowadays [AAE16; Na16]. The concept allows for aggregation of applications from small entities of software which are easier to maintain and develop. It is also easy to scale these services as migration cost is low. In the IoT market there is rapid development towards small and flexible services. Applications also migrate towards the end-user in order to decrease latency and improve scalability. Many edge cloud approaches are on the market for that purpose, for example Microsoft's Azure IoT Edge, Akamai IoT Edge Cloud, and Google's Cloud IoT.

The development of these approaches also follows the goal to decentralize server infrastructures and gain independence from big data warehouses. More and more stakeholders in industry aim to have local infrastructure close to their property. Together with the microservice approach, this continuous development improves flexibility, scalability and responsiveness of huge systems. However, the configuration and maintenance overhead for such distributed systems becomes significant. For example, the failure of a single sensor, router, or server can require the migration and duplication of many services in order to replace the failed subsystems. This process must be automated in order to keep complexity low on microservice levels.

Fig. 1 demonstrates such a scenario where services have to be migrated due to a system fault on some entities. To allow automatic migration of services and restoring of functionality, the system under observation and control must be based on a flexible service description and allow a controller to re-assemble functionality from small parts. With organic computing, for example, the overall system can restructure itself according to global rules. This self-organizing behavior allows for managing the growing complexity of several hundreds or

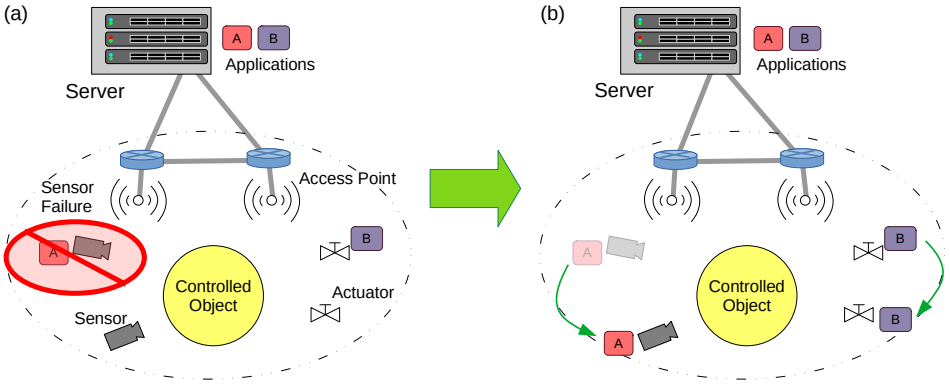


Fig. 1: Migration of a service in a redundant control system due to failure of a sensor (a). The sensor driver A migrates in order to recover from the failure and re-establish system functionality (b). Due to dependencies in the controlled object, also the actuator software has to migrate to another point in the system.

thousands of microservices dispatched on sensors, actuators, routers, and servers. However, such a restructuring demands strict system definitions with capabilities, requirements, and trust models in order to give organic controllers all the necessary information.

3 Trust in Internet of Things (IoT) Networks

Trust is a category that extends beyond security issues [YZV14]. The reliability, availability, resilience and persistence of a system are also contributing to its overall trustworthiness. Systems which generally are considered trustworthy are more often relied upon for in terms of information exchange, relaying, reasoning, or taking actions. For sensors and actuators, for example, their accuracy, precision and proper fault modeling may be considered more important than privacy or confidentiality, as the proper operation of a system depends primarily on those former properties.

In terms of security, however, trustworthiness is an important issue and covers authenticity, confidentiality, integrity and access control. Due to the heterogeneity and huge complexity of IoT networks, key exchange imposes a great challenge to applications. Configuring keys per hand comes with prohibitively high labor cost. Asymmetric encryption via certificates can be a solution to the key exchange problem without sharing secret keys, but opens up the new problem of certificate exchange. A certificate is hereby granted by a signing entity v, containing the information that another entity u has public key P_u.

$$\text{Cert}(u, v) = A_u | P_u | e(h(u|P_u), S_v)$$

Where S_v is the secret key of the issuer v, $e(m, k)$ is an asymmetric encryption function for message m using key k, and $h()$ is a hash function. A_u is the address (or *common name*) associated with u. | denotes the string concatenation.

3.1 Public Key Infrastructure

Public Key Infrastructure (PKI) solves the certificate exchange problem by hierarchical structures and is well-established in the World Wide Web. With PKI, a globally trusted Certificate Authority (CA) issues certificates for users u. The CA thereby has to check the identity of u and then associates P_u with it by issuing the certificate. Every entity wishing to check the identity of u can then take the publicly available certificate and verify it with the public key of the globally trusted CA.

The approach has drawbacks, however. The CA is a single point of failure in several ways. First, it has to be designated and maintained with special effort, imposing configuration maintenance overhead. Second, it must be trusted by everyone, which is especially problematic for heterogeneous IoT networks. Third, the central organization can also impose scalability issues.

3.2 Web of Trust

The Web of Trust (WoT) which was introduced with PGP [Ca07] dismisses the hierarchical PKI idea. Instead, nodes can issue certificates for each other in a peer to peer manner, creating a non-hierarchical network of trust. Trust in an entity increases with more peers have granted certificates for it. So there are nodes that are potentially more trustworthy than others, depending on the heterogenity of the peer group that had issued certificates for them. As this approach imposes the problem of non-binary trust, it is scalable and applicable for heterogeneous IoT networks.

Let (V, E) with $E \subset V \times V$ be the certificate graph of the IoT network of the nodes V. The certificate graph stores the certification relations between nodes, i.e. $(u, v) \in E \Leftrightarrow \text{Cert}(u, v)$ exists in the network. For simplicity we further assume a mutual certification process, meaning any node v that trusts u and issues a certificate $\text{Cert}(u, v)$ also receives a certificate $\text{Cert}(v, u)$ from u. Hence the graph is undirected.

A source node s can assume trust for a destination node d if there exists a certificate chain

$$s := v_0, v_1, ..., v_{c-1}, v_c =: d \qquad v_1, ..., v_c \in V, (v_i, v_{i+1}) \in E$$

with $\text{Cert}(v_i, v_{i+1})$ for $0 \leq i < c$. To begin with, every node may have only its neighbors included in its own certificate list when it enters a network. The problem of certificate chain discovery [Ki05; Mo07] evolves, as not every node is able to have direct access to every certificate present in the network.

4 Related Work in IoT PKI Solutions

The PGP standard [Ca07] defines a WoT model for the World Wide Web that could potentially be adapted for IoT [Ki05; Mo07]. Decentralized approaches exist, e.g. [DI10] that mitigate the configuration problem. But the issuing of a root certificate is always a problem.

Blockchain-based PKI approaches have the potential to eliminate the CA as the single point of failure [LSM17; PDF18; SB18]. These approaches, however, have huge drawbacks. Each node must store at least a subcopy of the global blockchain, and there must be a mining procedure as credit source between nodes. Both add storage and computational demands that are expensive for IoT nodes and limit scalability and energy efficiency.

5 Certificate Handling in (Dis-)Aggregating IoT Networks

In our work we suggest to use resource directories that are part of the Constrained Application Protocol (CoAP) specification [SKA19] to aid in certificate exchange process.

5.1 Service Discovery and Certificate Chain Discovery

Both the service-oriented architecture and the certification process suffer from respective discovery problems. Service discovery is necessary for aggregating microservices, because the complexity of applications is so high that they can not be managed statically. Similarly, since trust can not be statically configured with every IoT device, networks have to autonomously discover certificate chains at runtime. Both mechanisms can, therefore, utilize the same infrastructure, as we show in the following sections. Our idea is basically to make use of CoAP service directories, which compose a solution to the resource discovery problem, and re-use them to additionally store and discover certificate chains in IoT networks.

We base our solution on the certificate chain discovery algorithm proposed by Kitada et al. [Ki05], which we briefly introduce before presenting our variation. Suppose every IoT node $v \in V$ stores the set of certificates

$$\begin{aligned} C_v = &\{\text{Cert}(v,u) | u \text{ has signed a certificate for } v\} \cup \\ &\{\text{Cert}(u,v) | v \text{ has signed a certificate for } u\} \end{aligned}$$

that it either signed itself for another node u or that were issued for v by u. The certificate graph then contains an edge between nodes u and v if and only if there is a certificate $C \in C_v \cup C_u$. A certificate chain between nodes s and d is found using the Algorithm 1.

Mohri et al. [Mo07] calculated the mean communication overhead $S_1(k)$ for finding such a path as follows.

$$S_1(k) = h \times k \times S_{\text{res}} + h \sum_{i=1}^{k} m^i \times S_{\text{req}}(i-1) \tag{1}$$

Algorithm 1 Find a certificate chain $v_0, v_1, ..., v_c$ between $s = v_0$ and $d = v_c$ [Ki05]

$(V, E') \leftarrow \text{SpanningTree}(V, E, s)$
$v_0, v_1, ..., v_c \leftarrow$ Path from $d = v_c$ to $s = v_0$ in E'
return $v_0, v_1, ..., v_c$

Where h is the average number of hops between two nodes that share a certificate, m is the average node degree in the graph (V, E), k is the height of the spanning tree, and $S_{\text{req}}(i)$, S_{res} are the packet sizes of the request and response packets in bytes. They are given as follows (including header sizes $s_{\text{req}}, s_{\text{res}}$) [Mo07]:

$$S_{\text{req}}(i) = \text{sizeof}(\text{Cert}(u, v)) \times i + s_{\text{req}},$$
$$S_{\text{res}} = \text{sizeof}(\text{Cert}(u, v)) + s_{\text{res}}.$$

Note that k is at the same time the average path length of a certificate chain. By increasing the number of edges in the graph (thus increasing m), the path length k is reduced. Because k is in the exponent, increasing m to reduce k is a good choice, but is often not possible, because the number of certificates that can be stored on a resource-constrained IoT node is limited. Kitada et al. suggest $m = 4$ as a minimum value to ensure a closed graph is formed, but that might result in long paths and high overhead.

5.2 Variation by Introducing Structure with Resource Directories

As a solution to the exponential overhead we suggest the introduction of resource directories with bigger storage. CoAP specifies such resource directories to address the resource discovery problem [Sh20]. As the standard does not restrict their use, they can serve to support the certificate chain discovery problem as well. In IoT networks, these resource directories may consist of small server nodes, with memory capacities in the Gigabyte range, thus being able to store a significant amount of certificates. Examples of these could be edge nodes, like network routers or bridges, that are common entities in IoT networks.

Assuming that a single directory can store up to l certificates, we propose a network structure where the n nodes are divided into $\lceil n/l \rceil$ groups, each group being assigned to one resource directory. The certificate graph can then be ordered as indicated in Fig. 2.

The $\lceil n/l \rceil$ directories have to form a trust network among one another, which can be a fully connected graph, or itself a meshed network similar to that proposed in [Mo07] or [Ki05]. The former method is preferable, since it reduces the path length k to 3 at maximum. We however show by numerical analysis that also a meshed connection between the resource directories reduces the overhead to a practically feasible amount.

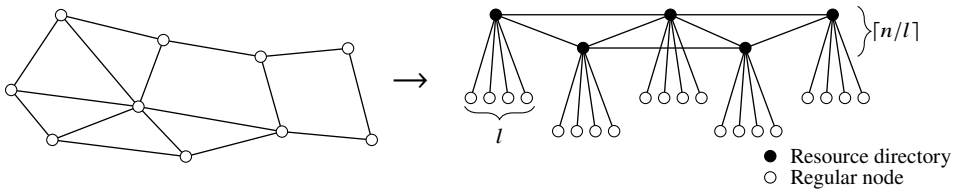


Fig. 2: Left: Unstructured certificate graph after [Ki05; Mo07]. Right: Structured graph through resource directories.

6 Numerical Analysis

We compare the communication overhead of the proposed resource directory based discovery algorithm using Eq. (1). We assume an IoT network in the form of a Wireless Multi-Hop Network (WMHN) with a variable node count n and a proactive routing scheme, that has fully populated routing tables as required by [Ki05]. The respective meshed topologies shall be chosen such that their diameter d is $O\left(\sqrt{n}\right)$. This is, for example, the case for a Manhattan grid. We estimate $d = \sqrt{n}$ for the sake of simplicity, which is for example the case for an 8-neighborhood.

Having $r = \lceil n/l \rceil$ resource directories (as part of the n regular nodes), we assume them to be interconnected with a certificate graph of diameter $d_{\text{structured}} = \sqrt{r}$. The height of the spanning tree then can grow up to a maximum of $k_{\text{structured}} = d_{\text{structured}} + 2$ between two regular nodes. We estimate the average node degree $m_{\text{structured}}$ as follows.

$$m_{\text{structured}} = 2 + \frac{r^2 - 3r}{n}.$$

The proof is given in the appendix. We compare the results to those that were assumed by Mohri et al. [Mo07] with a general network structure that does not distinguish between resource directories and normal nodes. For that, we then assume $m_{\text{unstructured}} = 4$, $k_{\text{unstructured}} = \lceil \sqrt{n} \rceil$. The structured approach thus reduces the value of k approximately by a factor of $\sqrt{l}$ compared to the unstructured approach.

Parameter	Symbol	Value
Request header size	s_{req}	16 Byte
Response header size	s_{res}	16 Byte
Certificate size	sizeof(Cert(u, v))	256 Byte
Mean hop count	h	1

Tab. 1: Common parameters used for analysis.

Moreover, the parameters in Tab. 1 are chosen for both methods.

Fig. 3 shows the results for both network structures, where $l = 100$ is chosen for the structured network. The structured network clearly outperforms the unstructured one. This is because the height of the spanning tree k and thus the communication path length is limited to the relatively small number of resource directories compared to the overall node count. In the structured approach, the flooding can omit the relatively big number of leaf nodes that are known to hold no useful information.

A disadvantage is the increased storage capacity that a resource directory needs to store certificates. They do not only have to store all certificates of their l leaf nodes, but also need additional storage to form a trust network among each other. In case of a fully connected network between the resource directories, the total number of stored certificates is $(r + l)\times\text{sizeof}(\text{Cert}(u, v))$.

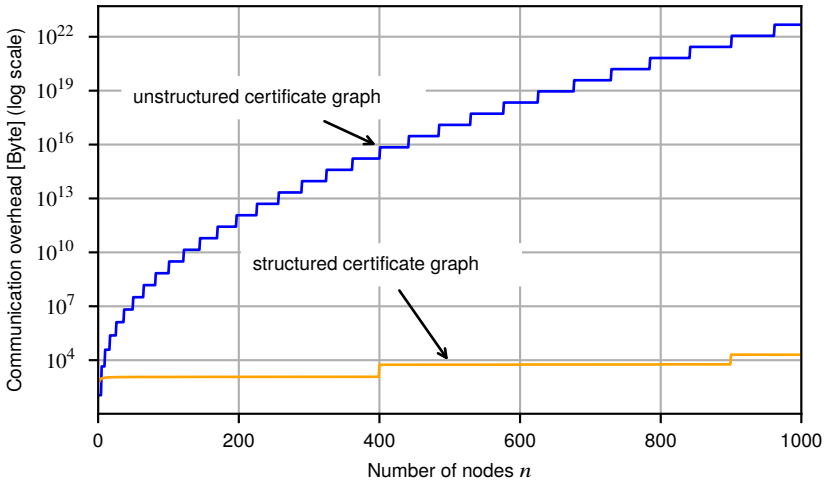


Fig. 3: Overhead analysis.

7 Discussion

Both the certificate chain discovery problem and the service discovery problem are important milestones on the pathway to realize microservice-based (dis-)aggregation of large applications in the IoT. In this paper, we provide an argument for solving both problems in the same manner in a combined, semi-structured approach. The benefits of the combination are the reduction of redundancies as well as strong synergy effects. By using specialized devices with a certain memory capacity, that act as combined service and certificate directories, the communication complexity is shown to be reduced by a significant amount. In our numerical analysis, resource directories, as they are proposed for example in IETF drafts [SKA19], show to reduce communication overhead. Naive approaches, on the

other hand, can easily overburden the capacities of devices and networks even in small-sized problems.

Although the semi-structured approach is accepted and already widely adopted for the resource discovery problem, it is not so popular for the certificate chain discovery problem, where the idea of having a completely distributed WoT, in absence of any hierarchies, dominates. The semi-structured approach is here an in-between solution that utilizes elements of both WoT and PKI approaches.

Acknowledgments. This work is funded by the German Federal Ministry of Education and Research (BMBF) as part of the project "Doriot" (Dynamic runtime for organically (dis-)aggregating IoT-processes) under grant number 01IS18071A. See also the project website `www.doriot.net`.

References

[AAE16] Alshuqayran, N.; Ali, N.; Evans, R.: A Systematic Mapping Study in Microservice Architecture. In: 2016 IEEE 9th International Conference on Service-Oriented Computing and Applications (SOCA). Pp. 44–51, Nov. 2016.

[Ca07] Callas, J.; Donnerhacke, L.; Finney, H.; Shaw, D.; Thayer, R.: OpenPGP Message Format. Internet Engineering Task Force (IETF)/, RFC 4880, Nov. 2007.

[DI10] Dahshan, H.; Irvine, J.: An Elliptic Curve Distributed Key Management for Mobile Ad Hoc Networks. In: 2010 IEEE 71st Vehicular Technology Conference. Pp. 1–5, May 2010.

[Ki05] Kitada, Y.; Takemori, K.; Watanabe, A.; Sasase, I.: On Demand Distributed Public Key Management without Considering Routing Tables for Wireless Ad Hoc Networks. In: 6th Asia-Pacific Symposium on Information and Telecommunication Technologies. Pp. 375–380, Nov. 2005.

[LSM17] Lin, J.; Shen, Z.; Miao, C.: Using Blockchain Technology to Build Trust in Sharing LoRaWAN IoT. In: Proceedings of the 2nd International Conference on Crowd Science and Engineering. ICCSE'17, Association for Computing Machinery, Beijing, China, pp. 38–43, July 2017.

[Mo07] Mohri, H.; Yasuda, I.; Takata, Y.; Seki, H.: Certificate Chain Discovery in Web of Trust for Ad Hoc Networks. In: 21st International Conference on Advanced Information Networking and Applications Workshops (AINAW'07). Vol. 2, pp. 479–485, May 2007.

[Na16] Nadareishvili, I.; Mitra, R.; McLarty, M.; Amundsen, M.: Microservice architecture: aligning principles, practices, and culture. O'Reilly Media, Inc., 2016.

[PDF18] Pinto, G. V.; Dias, J. P.; Ferreira, H. S.: Blockchain-based PKI for crowdsourced IoT sensor information. In: International Conference on Soft Computing and Pattern Recognition. Springer, pp. 248–257, 2018.

[Ro16] Roca, D.; Nemirovsky, D.; Nemirovsky, M.; Milito, R.; Valero, M.: Emergent Behaviors in the Internet of Things: The Ultimate Ultra-Large-Scale System. IEEE Micro 36/6, pp. 36–44, Nov. 2016.

[SB18] Singla, A.; Bertino, E.: Blockchain-Based PKI Solutions for IoT. In: 2018 IEEE 4th International Conference on Collaboration and Internet Computing (CIC). Pp. 9–15, Oct. 2018.

[Sh20] Shelby, Z.; Koster, M.; Bormann, C.; van der Stok, P.; Amsüss, C.: CoRE Resource Directory draft-ietf-core-resource-directory-24. IETF Draft/, Mar. 2020.

[SKA19] van der Stok, P.; Koster, M.; Ansüss, C.: CoRE Resource Directory: DNS-SD mapping draft-ietf-core-rd-dns-sd-05. IETF Draft/, July 2019.

[YZV14] Yan, Z.; Zhang, P.; Vasilakos, A. V.: A survey on trust management for Internet of Things. Journal of Network and Computer Applications 42/, pp. 120–134, 2014.

Appendix

Theorem 1. $m_{\text{structured}} = 2 + \frac{r^2-3r}{n}$ *is an upper bound for the average node degree in the structured network.*

Proof. The $n - r$ regular nodes each have a degree of 1. For the r directory nodes, we assume that the regular nodes are balanced among them. So edges exists from each directory node to (on average) $(n - r)/r \leq l$ regular (leaf) nodes. Furthermore, assuming a fully connected network between directory nodes as worst-case assumption, an edge to each other of the directory nodes exist, which yields an average degree of $r - 1 + (n - r)/r$ for each directory node. That yields

$$m_{\text{structured}} = \frac{(n-r) \times 1 + r \times (r - 1 + (n-r)/r)}{n} = 2 + \frac{r^2 - 3r}{n}.$$

□

R. Reussner, A. Koziolek, R. Heinrich (Hrsg.): INFORMATIK 2020,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2021 1225

IAL: An Information Abstraction Layer for IoT Middleware

Andrei Günter[1,4] Christopher Schwarzer[2,4] Matthias König[3,4]

Abstract: The internet of things is an ever-expanding world of connected devices and services. Through observation of analog and digital processes, plenty of information is continuously produced. More than often, the information thus obtained is tailored to serve one separate purpose. Existing architectures omit the fact that partial results can be enriched with meta information and can be shared among network participants while information is processed. We show that enriched and simplified information packages can be shared to facilitate cooperation between constrained and smart devices as well as to serve for system optimization. In this work, we propose an implementation of the so-called information abstraction layer which serves as collective resource for querying mechanisms to provide information. To emphasize and illustrate the need for a standardized information abstraction layer into existing middleware, we outline a realistic example of use and introduce concepts to build and evaluate shared information in networks.

Keywords: internet of things; iot; middleware; architecture; abstraction layer; sensor redundancy; sensor networks; multi modality; light detection and ranging; lidar

1 Introduction

Billions of devices become connected with the emergence of smart environments and the Internet of Things (IoT). Processing data collected by these devices enables software developers to provide useful services. A smart environment offers various possibilities to develop a service since a broad range of IoT devices can be used to gather helpful information.

For instance, a presence detector could be realized with the following two approaches: 1. by using a passive infrared (PIR) sensor or 2. by using a depth camera which recognizes motion. Both devices sense helpful information for this specific task, but each device provides a different *modality* of information, in this case a typical PIR sensor provides a voltage and a depth camera a point cloud. These modalities are too different to allow a reasonable comparison between them, but both can be interpreted uniquely to achieve the higher level information *presence detected*.

This example is illustrated in Fig. 1, where a service robot and a smart light are controlled by IoT devices in the environment to provide services. Therefore, two sensors gather

[1] andrei.guenter@fh-bielefeld.de
[2] christopher.schwarzer@fh-bielefeld.de
[3] matthias.koenig@fh-bielefeld.de
[4] Bielefeld University of Applied Sciences, Campus Minden, Artilleriestraße 9, 32427 Minden, Germany

doi:10.18420/inf2020_114

information of different modalities: 1. a depth camera provides points clouds, which are processed by a smart device with high performance to support controlling the robot and 2. a PIR sensor is used to deduce presence information with a constrained device to control a light. The depth data also carries helpful information for controlling the light, since a people detection algorithm is able to provide a presence detection as well. However, the constrained device is not able to process depth data due to a lack of performance. For the smart device, it is of little effort to share presence information in a level of detail which is appropriate for the constrained device, since it is already processing the depth data.

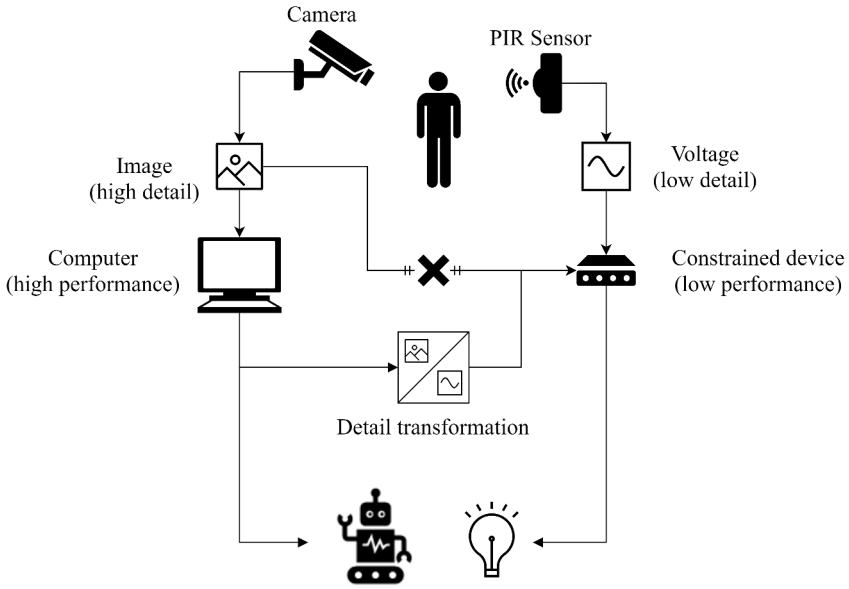


Fig. 1: A simplified example of providing information in different levels of detail in a smart environment.

A common approach in software development is to process a particular modality of information to provide a service. In turn, this restricts a software to rely on a certain type of device. In other words, the software is coupled to this certain type of device. The advantage of our proposed concept is a decentralized structure of devices and a loose coupling between hardware and software components.

Current architectures for closed systems and IoT networks provide several abstraction layers, e.g. hardware abstraction layers to support integration of different hardware, so that information can be exchanged seamlessly. This work focuses on enhancing the process of exchanging information in existing architectures by introducing an Information Abstraction Layer (IAL), which allows services to process information of different modalities to decouple software services from hardware. This task is part of a project named *"Dynamic runtime for organically (dis-)aggregating IoT-processes"* (DoRIoT). One goal of DoRIoT is to improve

the adaptability and resilience of IoT services by complementing unavailable devices with suitable alternative devices. Considering the example above, this means that the service for detecting presence remains functional or partially functional for as long as a PIR sensor, a depth camera or any other device capable of detecting presence provides correct data. This work contributes to this challenge by:

1. providing an architectural overview to handle multiple modalities of information,
2. implementing a use case, which is similar to the example given with Fig. 1,
3. evaluating the implementation with resulting improvements for further work.

The following sections are structured as follows: Section 2 summarizes related work, Section 3 presents a detailed formulation of the concept, Section 4 shows an actual implementation, Section 5 evaluates our implemented system, and finally Section 6 concludes this work.

2 Related work

The idea of IoT attracts researchers from multiple computer science domains since the diversity of devices involves many aspects: energy management, networking, safety, security, data persistence and many more. Over the past two decades, each of these research domains introduced comprehensive requirements for designing adaptable, intelligent, and resource friendly network architectures. This lead to a multitude of approaches for abstraction layers in IoT networks, i.e. IoT middleware.

Existing architectures for IoT middleware can be categorized into three layers: service-based, cloud-based and actor-based [Ng17]. The requirements of paradigms like big data and neural networks can be seen as cornerstones for service-based [Ca14; So15] and cloud-based [Ng17] architectures, since both architectures can be described as heavyweight in terms of computational power or data persisting capabilities.

One goal of this work is to develop an abstraction layer for IoT middleware allowing cooperation between constrained and smart devices. Service-based architectures will not be further discussed in this work, since service-based architectures are not designed for integrating constrained devices, and the same applies for cloud-based architectures, since these provide limited support for constrained devices [Ng17].

The development of communication protocols [BCS12; BG18] and operating systems [Ba13; Ba18; Le05] designed for constrained devices allowed researchers to setup actor-based architectures [PA15], which enable constrained devices to cooperate with smart devices.

It was shown that actor-based architectures improve existing software development approaches, since a heterogeneously designed application running on a smart device can be

executed on a network of constrained devices [Me17]. In other words, actor-based architectures decouple the execution of services from the performance of single connected devices, but did not focus on decoupling of software components from information modalities gathered by physical devices.

In this work, we focus on the development of a data-driven actor-based architecture to decouple services from modalities of information by abstracting information from raw data and translating it to a shared modality. This shared modality can be consumed to provide services independently from actual hardware.

3 Concept

The fundamental idea of our concept is that nearly every information can be used to serve multiple purposes and can be extracted into different levels of detail. Extracting information into a smaller level of detail allows constrained devices to process information which originally was achieved by devices acquiring information with a high level of detail, as shown with Fig. 2.

Smart devices produce or consume detailed information of models very close to the real world as opposed to constrained devices which produce or consume only simplified models. For instance, a very complex model might be a three-dimensional (3D) model of depth data in high resolution enhanced with colored data representing a full space and contained objects. While a comparable very abstract model would be a single distance from one point to another in that same space.

As the detail of an information decreases, it becomes feasible to process the information with a constrained device, since less performance is required to process smaller data sizes. In contrary, the requirement for reasoning the data and mapping it to meta information increases. For instance, a single distance is too abstract to make reasonable decisions in a process if there is no meta information on what exactly this distance is representing.

Information can be distinguished based on whether it is in a raw form, as it was acquired, or it is in a synthesized form enriched with meta information describing its context. In most cases, sharing information is of great interest, when it was synthesized before and represents a higher level context. The IAL is responsible to enrich information with context and enable information to be reused by multiple network participants. In other words, the integration of an IAL into existing IoT middleware allows services to use and share data extended with meta information for providing services decoupled from raw information modalities and consequently from its hardware.

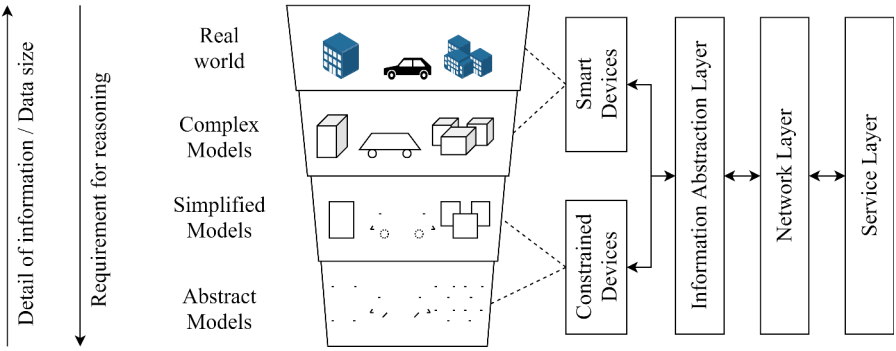


Fig. 2: Concept for different levels of detail in representations of information.

3.1 Problem formulation

A detailed formulation of our concept is given in the following. We assume that an environment contains agents which provide services by processing information. Agents are computing units optionally equipped with sensors or actors to interact with the environment. We denote a set of agents in an environment as:

$$\mathbb{A} = \{a_1, ..., a_{m-1}, a_m\} \tag{1}$$

Every agent uses information as input and optionally generates information as output. We define n inputs and outputs for an agent a_i as:

$$\mathbb{Z}_{a_i} = \{z_1, ..., z_{n-1}, z_n\} \tag{2}$$

A proper generation of outputs requires agents to presume particular data types for inputs. We define modalities for inputs and outputs to distinguish between different data types of information and to describe the context of information. With the following equation we define k modalities for an agent a_i, whereas k is less or equal to n since multiple pieces of information might be described with the same modality:

$$\mathbb{M}_{a_i} = \{m_1, ..., m_{k-1}, m_k\} \mid k \leq n \tag{3}$$

Information of high detail can be reduced or divided into multiple pieces of information, where each piece of information can be classified differently. For instance, a voice record can be classified based on information about number of different voices, duration of each occurring voice, gender of voices, or whether the record is a speech, a reading, or a song. Each classification might be described within a different context and might be represented with a different data type. In other words, information of high detail can be reduced, divided,

or transformed into pieces of information, where each piece might be described with a different modality. We define that a modality m_i can be transformed to another modality m_j if there is an existing set of process steps $\mathbb{P}_{i,j}$ which allow a deduction of m_j from m_i.

$$\mathbb{P}_{i,j} = \{p_1, ..., p_{l-1}, p_l\} \tag{4}$$

We define that a service s_i can be provided by an agent a_i if a necessary subset of information $\mathbb{Z}_{s_i} \subseteq \mathbb{Z}_{a_i}$ is available and up to date. Further, we denote that all modalities of $\mathbb{Z}_{s_i}$ can be described with $\mathbb{M}_{s_i} \subseteq \mathbb{M}_{a_i}$. In this work we show that services in IoT networks can be optimized by choosing proper modalities for $\mathbb{M}_{s_i}$ which enable to reuse information and allow a cooperation between agents equipped with different sensors leading to a loose coupling between hardware and software. In the remainder of this section, we propose a system architecture to provide an overview of necessary components to implement the presented concept. Fig. 3 depicts our proposed decentralized system architecture, which is composed of multiple participants in a network, namely: agents and brokers.

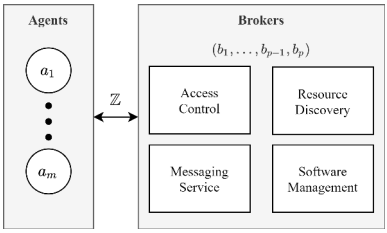


Fig. 3: Proposed system architecture.

Agents $\mathbb{A}$ generate information $\mathbb{Z}$ which is shared via networking by communicating with brokers. Brokers are responsible for managing and distributing information among network participants and can be described as IoT middleware, which consists of four components: a messaging service, an access control, a resource discovery, and a software management. A messaging service is required to send and receive messages through the network. A typical setup for a messaging service would be a publish and subscribe scenario where each participant is able to send information by publishing to a topic or to receive information by subscribing to a topic.

If a topic contains incomplete, outdated or corrupted data, then other related topics should be available for use. Therefore, a resource discovery allows consumers and producers to query for information. A resource discovery also supports aggregation of new devices to the system, since devices can be directed to existing topics. Newly aggregated devices from other networks or devices containing processing units also contribute to the system by running local software. Detecting and synchronizing updated software can be beneficial for network participants, especially in case of required security updates. In contrary, software updates also involve security issues. Therefore, an access control is used to allow only permitted participants to access messaging services, resource discovery, and software management.

4 Implementation

We did not implement every aspect of our presented concept, since the following components are handled as distinct research topics in the DoRIoT project: access control, resource discovery, and software management. We focused on implementing four agents and one broker with a messaging service. An overview of the implementation is depicted in Fig. 4.

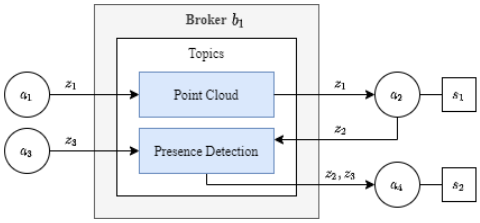


Fig. 4: An overview of the implemented components.

The system provides two services s_1 and s_2, where s_1 is provided by agent a_2 and s_2 is provided by agent a_4. The agent a_2 provides a multi-object tracking based on point cloud data with a graphical user interface, where each moving object is shown with a distinct color in a map relative to real world coordinates. The agent a_4 controls a smart light based on presence information. Therefore, the system processes five different types of information, whereas we omit two raw information types for simplicity in Fig. 4 since those are directly processed by agent a_1 and agent a_3. A raw point cloud is acquired via a solid state LiDAR sensor and initial noise is removed by agent a_1. The result z_1, a noise reduced point cloud represented by an array of 3D coordinates (x,y,z) and annotated with a sensor identifier as context, is send to a broker b_1 and can be described with modality m_1. The other raw information is a voltage, which is observed and processed with a PIR sensor by agent a_3. The result z_3, a binary value representing presence or absence of people in a room with an identifier, is also send to the broker b_1 and can be described with modality m_2.

The agent a_2 consumes information of modality m_1 and provides the service s_1 by applying a change detection to a sequence of point cloud frames. If a change is detected, then a clustering and a tracking is executed and visualized. The change detection is implemented via a software library [RC11] and the multi object tracking is realized via an open source algorithm [Pa19]. The described change detection is part of the algorithm for processing the point cloud and it directly relates to the function of the PIR sensor. In other words, the change detection also detects presence of objects and humans. Therefore, the agent a_2 is also used to publish presence information z_2 to the broker b_1. As already described with modality m_2, information z_2 can also be described with a binary modality representing presence or absence of humans in a room with the same identifier as described with m_2. In other words, z_2 and z_3 share the same modality m_2.

The service provided by a_4 makes use of agents with completely different hardware, since the light is controlled with presence information initially acquired with a PIR sensor or acquired with a solid state LiDAR sensor. Thus, service s_2 allows optimization of several

criteria, such as availability, redundancy, and network load. The hardware characteristics of all agents are listed in Tab. 1, which highlights the differences in terms of performance between all used devices.

Agent	Sensor	Processor	Memory	Connectivity	OS
a_1	Solid state LiDAR	1.5 GHz	4 GB DDR-4	Ethernet	Rasp. Buster Lite
a_2	none	1.9 GHz	16 GB DDR-4	Ethernet	Ubuntu 18.04
a_3	PIR	216 MHz	512 KB SRAM	Ethernet	RIOT-OS 2020.04
a_4	none	216 MHz	512 KB SRAM	Ethernet	RIOT-OS 2020.04

Tab. 1: Characteristics of employed agents.

5 Evaluation

Based on the previously presented implementation, we can further specify the dependencies of a service by analyzing the flow of information $\mathbb{Z}$ which is supplied to services by agents. Fig. 5 depicts the dependencies for service s_2 provided by a_4 allowing the statement, that the service s_2 is available for as long as one element of $\mathbb{Z}_{a_4} = \{z_2, z_3\}$ is delivered frequently. Further, it can be seen that there is a transformation $\mathbb{P}_{1,2}$ from modality m_1, which is described with point cloud data, to modality m_2, which is a binary value representing *presence detected* or *absence detected*.

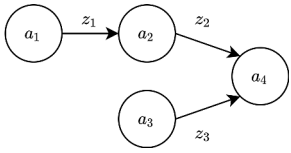


Fig. 5: A graph showing the dependencies for service s_2 provided by agent a_4.

We identify missing descriptions of occurring modalities in our implementation by analyzing each process step of our implementation. Fig. 6 shows each process step for the agents a_2 and a_3 and illustrates resulting outputs of information. By checking the communication with the broker b_1, it can be seen that clustered data is not shared and subsequently not described with a modality. Clustered depth data is interesting for services based on point clouds, since clustering is a major procedure in a multitude of algorithms used in depth data, e.g. a classification of objects.

Fig. 6 also serves as appropriate illustration to motivate the software management component presented in Section 3. Homogeneously designed applications could be distributed among network participants with lower computing capabilities by sharing the source code for each process step of a_2 to multiple other agents $a_f, ..., a_{g-1}, a_g$. An implementation could look similar to an already proposed solution in recent work [Me17].

In this paragraph we are going to outline the capability of implementing a resource discovery component on top of the existing implementation. As defined with the proposed concept, there are existing meta annotations describing modalities $\mathbb{M}_{a_i}$ of agents. These modalities

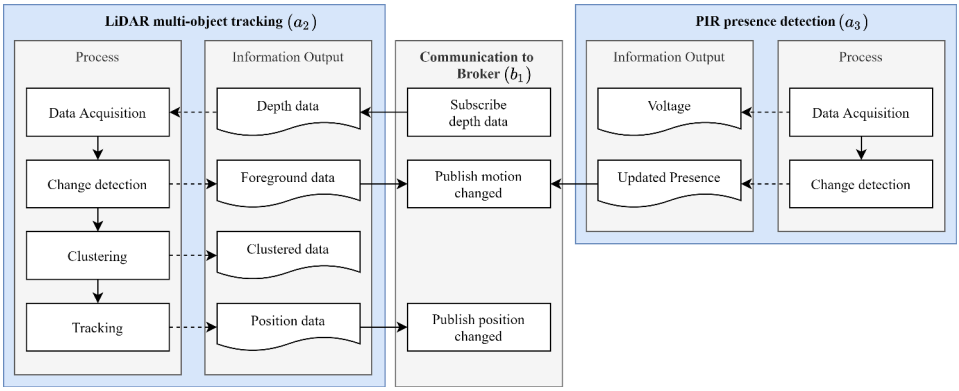


Fig. 6: Process steps and respective information outputs of agents a_2 and a_3.

can serve for querying certain types of information and brokers are able to provide a list of available topics which match the query. For now, modalities are represented by a simple identifier and a datatype and can be configured by developers. We recommend to use meaningful identifiers for modalities, since this enables to use semantic querying techniques while keeping the architecture lightweight.

6 Conclusion

Services in existing architectures often consume raw information and produce highly abstracted information for one specific application. Partial results are not shared which increases a system's redundancy because intermediate processing steps have to be carried out for each requirement on several occasions.

Our presented concept allows constrained devices and smart devices to cooperate even when running completely different sensors or actors. This was accomplished with an IAL component to transfer original raw information achieved from real world environments into a shared modality.

To highlight a use case for different services with different devices, we implemented a multi-object tracking with a LiDAR sensor and a presence detection using a PIR sensor. Then, we abstracted a modality describing raw information achieved by the LiDAR sensor into a binary shared modality *presence detected* or *absence detected*. The implementation was enabled by abstracting the presence information from depth data with a change detection step, which is a common procedure to apply in tracking services. Thus, and due to the proposed decentralized architecture, it is of little effort to share this abstracted presence information with other IoT devices.

Further work offers a multitude of use cases. Networks can be optimized by discovering redundant devices or by rating the services with quality metrics like availability, scalability, or resilience. With future work, we are going to focus on the implementation of remaining components, which were proposed with our architecture: access control, service discovery, and software management.

References

[Ba13] Baccelli, E.; Hahm, O.; Günes, M.; Wählisch, M.; Schmidt, T. C.: RIOT OS: Towards an OS for the Internet of Things. In: IEEE Conference on Computer Communications Workshops. Pp. 79–80, 2013.

[Ba18] Baccelli, E.; Gündoğan, C.; Hahm, O.; Kietzmann, P.; Lenders, M. S.; Petersen, H.; Schleiser, K.; Schmidt, T. C.; Wählisch, M.: RIOT: An Open Source Operating System for Low-End Embedded Devices in the IoT. IEEE Internet of Things Journal 5/6, pp. 4428–4440, 2018.

[BCS12] Bormann, C.; Castellani, A. P.; Shelby, Z.: CoAP: An Application Protocol for Billions of Tiny Internet Nodes. IEEE Internet Computing 16/2, pp. 62–67, 2012.

[BG18] Buschsieweke, M.; Güneş, M.: Access Control for Medical Devices: Tweaking LCap for Health Informatics. In: 2018 IEEE Globecom Workshops (GC Wkshps). Pp. 1–7, 2018.

[Ca14] Calbimonte, J. P.; Sarni, S.; Eberle, J.; Aberer, K.: XGSN: An Open-source Semantic Sensing Middleware for the Web of Things. In: International Workshop on Semantic Sensor Network. 2014.

[Le05] Levis, P.; Madden, S.; Polastre, J.; Szewczyk, R.; Whitehouse, K.; Woo, A.; Gay, D.; Hill, J.; Welsh, M.; Brewer, E.; Culler, D.: TinyOS: An Operating System for Sensor Networks. In: Ambient Intelligence. Springer, pp. 115–148, 2005.

[Me17] Mehta, A.; Baddour, R.; Svensson, F.; Gustafsson, H.; Elmroth, E.: Calvin Constrained — A Framework for IoT Applications in Heterogeneous Environments. In: 2017 IEEE 37th International Conference on Distributed Computing Systems (ICDCS). Pp. 1063–1073, 2017.

[Ng17] Ngu, A. H.; Gutierrez, M.; Metsis, V.; Nepal, S.; Sheng, Q. Z.: IoT Middleware: A Survey on Issues and Enabling Technologies. IEEE Internet of Things Journal 4/1, pp. 1–20, 2017.

[PA15] Persson, P.; Angelsmark, O.: Calvin – Merging Cloud and IoT. Procedia Computer Science 52/, pp. 210–217, 2015.

[Pa19] Palanisamy, P.: Multiple Object Tracking from Point Clouds v1.0.2, 2019, URL: https://github.com/praveen-palanisamy/multiple-object-tracking-lidar.

[RC11] Rusu, R. B.; Cousins, S.: 3D is here: Point Cloud Library (PCL). In: IEEE International Conference on Robotics and Automation. Pp. 1–4, 2011.

[So15] Soldatos, J.; Kefalakis, N.; Hauswirth, M.; Serrano, M.; Calbimonte, J. P.; Riahi, M.; Aberer, K.; Jayaraman, P. P.; Zaslavsky, A.; Žarko, I. P.; Skorin-Kapov, L.; Herzog, R.: OpenIoT: Open Source Internet-of-Things in the Cloud. Springer, 2015.

R. Reussner, A. Koziolek, R. Heinrich (Hrsg.): INFORMATIK 2020,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2021 1237

Application Layer Security for the IoT

Information Security for CoAP with LCap, Payload Encryption, and HMACs

Marian Buschsieweke[1] Mesut Güneş[2]

Abstract: Security for the mostly constrained devices forming the IoT is an active field of research. In this paper, we propose two CoAP Options, *HMAC1*/*HMAC2* and *Crypt1*/*Crypt2* complementing our previous work on Lightweight Capability Based Access Control (*LCap*). This results in a lightweight, flexible, and complete solution for application layer security for CoAP nodes with severely limited memory. In our evaluation, we show that a pure software implementation without cryptographic hardware acceleration is feasible for practical use on highly constrained IoT devices. Due to mostly idiomatic use of cryptography, existing security analyses apply to our proposal. Our security framework was designed with ample focus on reducing the complexity of the system, which allows lean implementations and simplifies security reviews. This makes *LCap* based security a good fit for security in the IoT.

1 Introduction

An ever increasing number of daily objects are equipped with a microcontroller and network connectivity to form SmartX environments. The architecture of the Internet of Things (IoT), in which any given IoT node can communicate with any other, is highly promising for such smart environments: This ubiquitous availability paves the way for spontaneous cooperation between any willing IoT nodes and allows for emergent systems to be formed. The other side of the coin is that the risk of abuse of IoT nodes is capital. In order to exploit the potential of this architecture while controlling the risk, efforts on providing security and privacy for the IoT are being made in both academia and industry [TB19]. In this paper, a complete application layer security stack is proposed to join this effort.

The remainder of this paper is structured as follows: Promising work on application layer security for IoT nodes is presented in Section 2. In Section 3 we propose the *HMAC1*/*HMAC2* and the *Crypt1*/*Crypt2* CoAP Options to provide integrity, authenticity, data freshness, and confidentiality for messages send using CoAP [SHB14]. These options are complementary to our previously proposed solution for access control, *LCap* [BG18], and combined provide a complete security framework for CoAP nodes. The impact the use of our proposed security framework has on response time and required CPU instruction is evaluated in Section 4. In Section 5, a security evaluation is performed. Finally, a conclusion is drawn in Section 6.

[1] Otto-von-Guericke University Magdeburg, Communication and Networked Systems (ComSys), Faculty of Computer Science, Universitätsplatz 2, 39106 Magdeburg, Germany, marian.buschsieweke@ovgu.de
[2] Otto-von-Guericke University Magdeburg, Communication and Networked Systems (ComSys), Faculty of Computer Science, Universitätsplatz 2, 39106 Magdeburg, Germany, mesut.guenes@ovgu.de

doi:10.18420/inf2020_115

2 Related Work

In the following, the most promising work on access layer information security targeting IoT nodes of class C2 (using RFC 7228 [BEK14] terminology) and is presented below.

2.1 OAuth

The IETF working group Authentication and Authorization for Constrained Environments (ACE) is strongly engaged in work on information security for IoT devices. One approach to access control is the adaption of OAuth for the IoT [Se20]. For this, efficient encodings of OAuth entities such as the CBOR Web Token [Jo18] to encode claims are used. Claeys et al. used these building blocks to employ OAuth 1.0a on IoT nodes [CRT18]. They were able to implement core components of their proposal on a class C2 device (using RFC 7228 [BEK14] terminology). However, without a complete implementation the feasibility of their proposal cannot be conclusively verified. In addition, applicability to the more constrained classes C1 and C0 IoT nodes is crucial for an access control solution for the IoT. So far, it remains unclear whether OAuth is lightweight enough to be run on the more constrained IoT nodes below class C2.

2.2 Object Security for Constrained RESTful Environments (OSCORE)

Object Security for Constrained RESTful Environments (OSCORE) [Se19] implements application layer security for CoAP relying on CBOR Object Signing and Encryption (COSE) [Sc17]. It primarily targets CoAP nodes communicating over proxies with either CoAP or CoAP-mappable HTTP endpoints. In this context secure communication channels on the transport layer have to terminate at the proxy to enable it to perform required modification of the forwarded messages, such as the conversion between CoAP and HTTP. OSCORE moves the actual request and response into the payload using Concise Binary Object Representation (CBOR) for serialization and COSE for security. Only header fields and options required for conformance with CoAP/HTTP or required to be accessible by proxies are additionally exposed in the message carrying an OSCORE request/response. For access control, OSCORE relies on supplementary approaches such as OAuth.

2.3 Capability Based Access Control

The basic idea of Capability Based Access Control (CBAC) is the separation of the decision on and the enforcement of access. It puts the client in charge of proactively obtaining an unforgeable capability token that proves access rights. The authority issuing this token is thus deciding on the access. A server receiving a capability token is only enforcing this decision by verifying the validity, authenticity, integrity, and genuineness of the token. The basic idea was first introduced by Saltzer et al. [SS75] in the context of processes requesting access to resources managed by an operating system. Mahalle et al. [Ma12] proposed transferring this approach as described above to network communication. Chen et al. [CGH16] proposed using a CoAP Option to embed the capability token in the request.

2.4 Delegated CoAP Authentication and Authorization Framework (DCAF)

Gerdes et al. [GBB15] proposed to let constrained servers and clients delegate authorization to a Server Authorization Manager (SAM) and a Client Authorization Manager (CAM), respectively. In the Delegated CoAP Authentication and Authorization Framework (DCAF), an access ticket is issued to the client once the authorization managers agree on granting access. Exactly like a capability token, this ticket proves granted access to a requested resource. But in addition, DCAF delegates key negotiation to the authorization managers, so that a DTLS channel will be established using this pre-shared key. Thus, DCAF provides a full security stack with access control being implemented on the application layer, and integrity, authenticity and confidentiality on the transport layer.

2.5 Lightweight Capability Based Access Control (*LCap*)

Buschsieweke and Güneş introduced the Lightweight Capability Based Access Control (*LCap*) [BG17; BG18], a lightweight, scalable and flexible access control for IoT nodes. Unlike other approaches for CBAC, *LCap* relies solely on symmetric cryptography. In *LCap*, individual keys are exchanged between CoAP servers and *Token Authorities*. These keys are used to sign *LCap Tokens* using a classic HMAC scheme. In addition, every *LCap Token* contains an individual *Token Key* that is encrypted using the keys shared with the *Token Authorities*. This zero round trip key exchange of the *Token Key* allows the use of symmetric cryptography while the number of pre-shared keys a CoAP server needs to store remain constant with growing number of authorized CoAP clients; and can be as low as one. CoAP clients have to provide a valid *LCap Token* in order to access a resource. As proof of possession of the *LCap Token*, a client proves knowledge of the plain text of the *Token Key*, whose cipher text is embedded in the *LCap Token*. The server can decrypt the embedded encrypted *Token Key* using the keys exchanged with the issuing *Token Authority*. *LCap* additionally provides authenticity and integrity by using the *Token Key* to attach an HMAC protecting the CoAP request. Replay attacks are prevented and data freshness is enforced by adding a slow ticking time stamp, the *LCap Epoch*: Within the validity of an *LCap Epoch* CoAP's duplicate detection will reject replays as duplicates. *LCap* uses CoAP Options nested within the *LCap Option* to encode additional side conditions that need to be met in order to gain access. These Suboptions can be marked as critical and elective in the same manner as CoAP Options, so that new side conditions can be defined in a backward compatible manner.

3 *LCap* Based Application Layer Security

LCap by itself implements access control and additionally provides authenticity, integrity, and data freshness for the request. This paper introduces two additional groups of CoAP Options: The *HMAC* Options and the *Payload Encryption* Options. The former provides the same security guarantees for the response, the latter enables confidentiality (for both requests and responses).

Name	#	Type	Description
Time[1]	1	`uint`[2]	Timestamp the HMAC was calculated
Key ID	9	`uint`[3]	ID of key used in the HMAC
Algo	11	`uint`[3]	ID of the HMAC algorithm used

1. This Suboption is not used for HMAC Options in negotiation mode.
2. Unix time stamp.
3. Value is 0 if Suboption is missing.

Tab. 1: List of Suboptions used in *HMAC1* and *HMAC2* Options

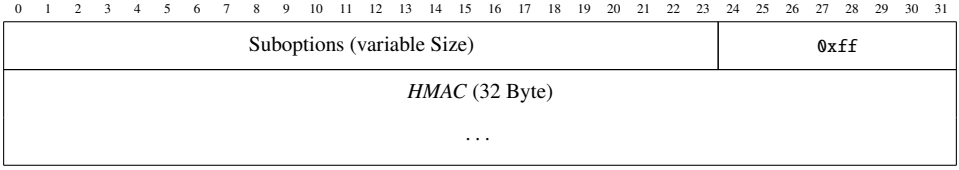


Fig. 1: Format of the *HMAC1*/*HMAC2* Option when carrying an HMAC

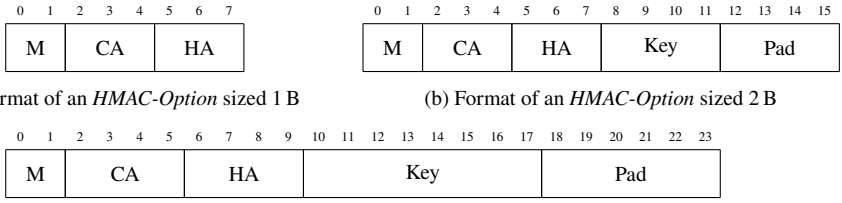
3.1 Integrity and Authenticity Protection

The *HMAC* Options use a simple HMAC-Scheme [KBC97] to protect the integrity and authenticity of the message carrying them. Similar to the terminology in the CoAP Block-Wise Transfer [BS16], the *HMAC2* Option always refers to the response and the *HMAC1* Option refers to the request. The *HMAC1* Option is incompatible with the use of *LCap* in requests and is intended as a lightweight alternative to *LCap* in scenarios access control is not needed. The *HMAC2* Option complements the use of *LCap* by protecting the response.

An *HMAC2* Option in a response or an *HMAC1* Option in a request is referred to as an *HMAC* Option in HMAC mode. In this mode, *HMAC* Options contain an HMAC value to protect the integrity of the IP address and port of both sender and receiver, as well as the whole content of the CoAP message. The cryptographic parameters used for the calculation of the HMAC and the time stamp when the message was created are given using the Suboptions in Fig. 1. These Suboptions use the CoAP Option Format [SHB14], but are stored within the data section of a regular CoAP Option. The encoding of an *HMAC* Option in HMAC mode is depicted in Tab. 1.

An *HMAC2* Option in a request or an *HMAC1* Option in a response is referred to as an *HMAC* Option in negotiation mode. As this name implies, these options are used to negotiate cryptographic parameters of the HMAC, rather than carrying an HMAC. The negotiation mode is particularly useful if the server (or client) defaults to a cryptographic hash algorithm not supported by the client (or server). For obvious reasons, neither the *Time* Suboption nor an HMAC value is added when the HMAC Options are used for negotiation.

The BLAKE2s-256 [Au14] cryptographic hash function is referred to with ID 0 as *Algo*.



(a) Format of an *HMAC-Option* sized 1 B

(b) Format of an *HMAC-Option* sized 2 B

(c) Format of an *HMAC-Option* sized 3 B

M *Mode ID*, specifies the block cipher mode of operation
CA *Cipher ID*, specifies the block cipher algorithm to use
HA *Hash ID*, specifies the cryptographic hash to use to generate the initialization vector (IV)
Key *Key ID*, specifies the key to use for encryption (defaults to zero)
Pad *Padding*, specifies the number of padding bytes added to the input (defaults to zero)

Fig. 2: Formats of the *Crypt1*/*Crypt2* Option depending on the size of the CoAP Option

Alg. 1: Algorithm used to derive the IV

Data: shared key as *key*, message as *msg*, LCap Epoch or HMAC time stamp as *nonce*
Result: derived IV

```
return cryptoash(key, nonce, msg.token, msg.message_id)
```

A value of 0 in the *Key ID* is only allowed in the *HMAC2* Option and refers to the same key used to protect the integrity of the request. In case the request used an *LCap Token*, the *Token Key* is referred to with *Key ID* 0.

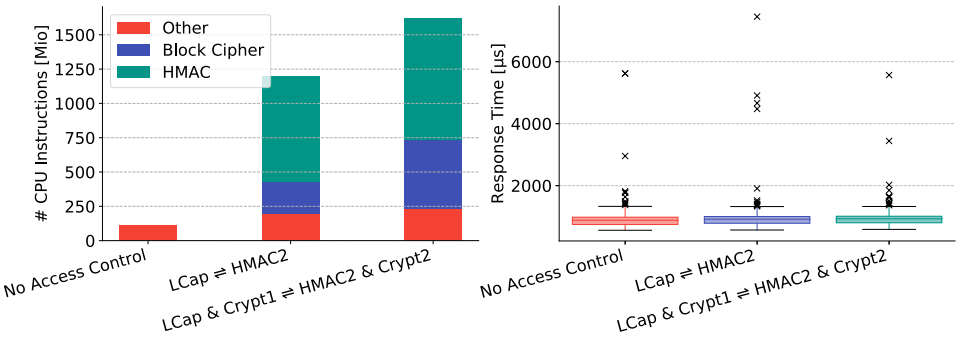
3.2 Confidentiality of the Payload

Payload Encryption of CoAP messages is provided by the *Crypt1* and *Crypt2* Options. The naming convention is the same as used for the HMAC Options: *Crypt1* always refers to the request and *Crypt2* to the response. Again, both CoAP Options are allowed for both request and response.

A *Crypt1* Option in a request or a *Crypt2* Option in a response is referred to as Crypt Option in encryption mode. In this mode, the option specifies the cryptographic parameters used for encrypting the payload of the CoAP message. The encoding of *Crypt1* and *Crypt2* Options is depicted in Fig. 2.

A *Crypt1* Option in a response or a *Crypt2* Option in a request is referred to as Crypt Option in negotiation mode. Again, these are used to indicate the preferred configuration to the communication partner. The format of the Crypt Option in encryption mode as shown in Fig. 2 is also used in negotiation mode. However, the padding is always set to zero during negotiation.

Using *Payload Encryption* implies the use of either the *LCap* Option or the *HMAC1*/*HMAC2*. The motivation for this is on the one hand that there is practically no use case for confidentiality without integrity and authenticity. On the other hand, this allows reusing the



(a) Total CPU instruction required by the server

(b) Response time

Fig. 3: Benchmark for different security settings when using AES-128 as block cipher, BLAKE2s-256 as cryptographic hash function, and CFB as mode of operation

replay protection of the *LCap* Option or the *HMAC* Options to compute IVs, rather than increasing the size of CoAP messages by explicitly transferring it. The cryptographic hash function specified using *Hash ID* is used to compute the IV as shown in Alg. 1. The result is truncated as needed for the selected block cipher.

A *Cipher ID* of 0 refers to AES-128, a *Mode ID* of 0 specifies cipher feedback (CFB) as mode of operation, and a *Key ID* of 0 refers to the same key that was used in the *HMAC* Option or in the *LCap Token* that was used to protect the integrity of the message.

4 Performance Evaluation

4.1 Benchmark Setup

A desktop class PC was used to run both the CoAP client and server, so that they could communicate using the local network device. Thus, the communication is not affected by network congestion and channel properties. The client sent 32 767 `PUT` requests to the URI-Path `/led` with the payloads `1` and `0` in turns. For each of the following security parameters benchmarks were run:

1. No application layer security for both request and response
2. *LCap*-Option in the request, HMAC2-Option in the response
3. *LCap*- and Crypt1-Option in the request, HMAC2- and Crypt2-Option in the response

For each configuration a benchmark was run with different combinations of the used block cipher, the used mode of operation, and the used cryptographic hash function. The response time was measured and the total number of required CPU instructions the server required to handle all 32 767 requests were recorded using `callgrind`.

4.2 Benchmark Results

As shown in Fig. 3a, the required CPU instructions increase by a factor of ≈ 11.7 when protecting the request using the *LCap*-Option and the response using the HMAC2-Option. When additionally the payload of both request and response are encrypted, the CPU instruction increase by a factor of ≈ 14.5. The additionally required CPU instructions are mostly spent on the cryptographic primitives, namely the used block cipher and the used cryptographic hash function. The remaining additional CPU instructions are required to construct and parse the additional CoAP Options as well as the verification of the *LCap* attributes such as whether the token is used within its period of validity.

In Fig. 3b the response time depending on the used security options is shown. These box plots indicate that on desktop class hardware the impact caused by the use of application layer security is completely negligible. Apparently, the response time is I/O bound for every setting on the used hardware.

5 Security Evaluation

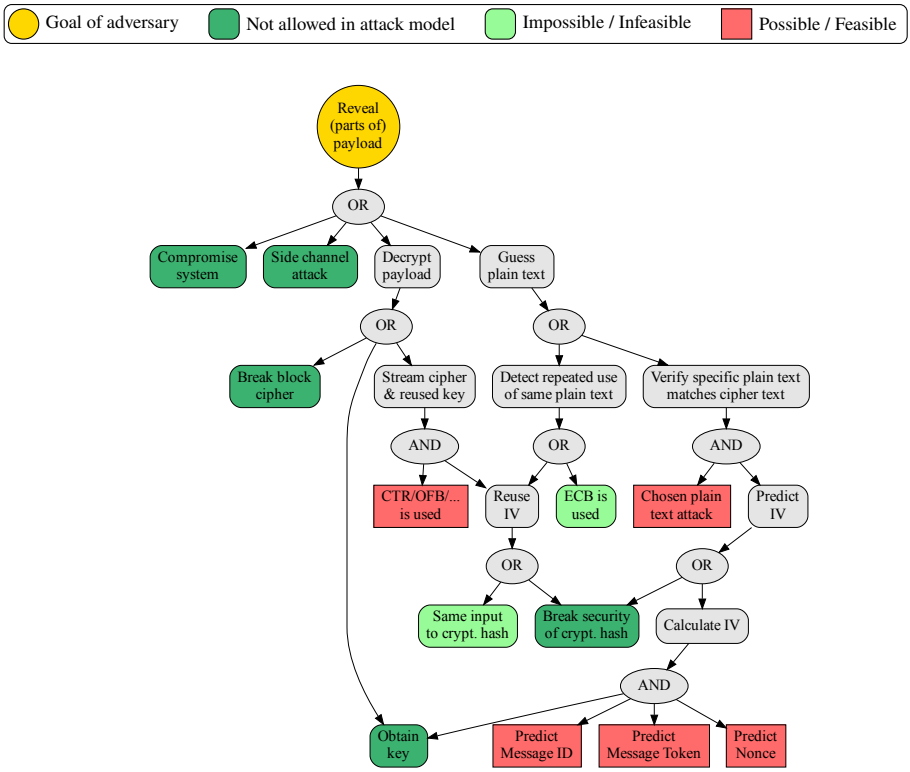
5.1 Attack Model

In the following, the security of the *Payload Encryption* and the *HMAC1*/*HMAC2* Options proposed in the paper are analyzed. For that it as assumed that the adversary is able to intercept any message, record any message, and inject and alter messages at will. It is however assumed that the adversary is unable to compromise any of the communicating nodes and no implementation flaws are present in communicating systems, including those leaking details on side channels. Finally, it is assumed that it is infeasible for the adversary to break the security of the used cryptographic hash function and the used block cipher.

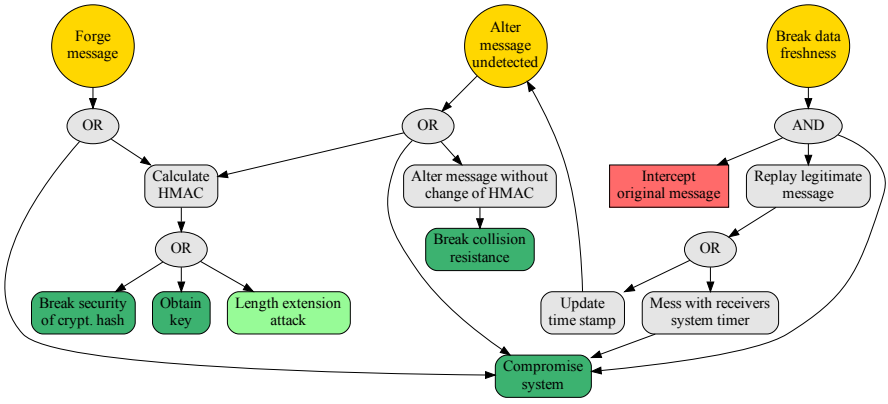
This attack model, thus, rules out the increasingly relevant attack vector of side channel attacks as well as any attack on the used cryptographic primitives. The reason for this is that side channel attacks inherently target a specific implementation rather than the actual specification this paper presents. Hence, these attacks are out of scope of this paper. Similarly, the security of cryptographic building blocks is out of scope of this paper. Instead we refer to the plethora of cryptanalysis published for the widely used block ciphers and cryptographic hash functions.

5.2 Security Evaluation of the *Payload Encryption*

In Fig. 4a an attack tree is shown that analyzes potential attacks on the confidentiality of the payload when using *Payload Encryption*. As *Payload Encryption*, with the exception of how the IV is obtained, is an idiomatic use of symmetric encryption, plenty of existing security analysis is present. Thus, most attack vectors apply to symmetric cryptography in general and are excluded from the adversary abilities in Section 5.1 as out of scope. The major difference to textbook use of symmetric cryptography is the computation of the IV, rather



(a) Attack tree for *Payload Encryption*



(b) Attack tree for *HMAC1*/*HMAC2* Options

Fig. 4: Attack trees analyzing attack vectors on the *Payload Encryption* and *HMAC1*/*HMAC2* Options

than choosing it randomly (using a high entropy source of randomness) and transferring it explicitly. With the use of ECB as block cipher mode of operation in *Payload Encryption* being forbidden, all remaining attack vectors listed in 4a depend on either the same IV to be generated more than once or being predictable by the adversary. Predicting the IV requires predicting the cryptographic hash of the concatenation of the key, the "Nonce" (either the *Epoch* of the *LCap* Option or the time stamp of the *HMAC*-Option used together with the *Payload Encryption*), the Message Token, and the Message ID. With the exception of the key, all parts of the input are easy to predict by an adversary. Assuming enough bits in the key are unknown to the adversary, it still is infeasible for the adversary to predict the hash due to the security guarantees of the cryptographic hash function.

As CoAP relies on the Message ID for duplicate detection, the same Message ID cannot be reused for the duration the communication partner is expecting duplicates. By the time the same Message ID can be reused again, the "Nonce" (the *LCap Epoch* or the *HMAC*) time stamp is different. Thus, the IV of different messages is never calculated using the same input. The chance of a cryptographic hash function with a high collision resistance yielding the same hash value again for these distinct input values is therefore close enough to zero, that it becomes infeasible for an adversary to count on this.

5.3 Security Evaluation of the *HMAC1/HMAC2* Options

Possible attacks on the message integrity, authenticity or data freshness are analyzed in the attack tree in Fig. 4b. The *HMAC* Option relies on idiomatic use of a cryptographic hash function that relies on the classic HMAC [KBC97] construct to prevent length extension attacks, so that e.g. the SHA-256 hash function can securely be used. The addition of a time stamp that is also covered by the HMAC, a receiver of a message protected by an *HMAC* Option is able to determine and enforce the freshness of the received data.

6 Conclusion

In this paper we have complemented our existing work on *LCap*, a lightweight implementation of Capability Based Access Control (CBAC) that additionally provides integrity, authenticity and data freshness for CoAP requests: We introduced the *HMAC2* Option to extend the same security to CoAP responses and proposed the *Crypt1* / *Crypt2* Option to provide confidentiality by encrypting the payload of CoAP messages. As both constructs are idiomatically applying symmetric cryptography to protect CoAP messages, preexisting security analyses mostly apply to the introduced *Payload Encryption* and *HMAC2* Option. Hence, the security implications and best practises are well understood. In addition, we provide a novel scheme to securely compute the initialization vector (IV) used for the *Payload Encryption*, which frees communication partners of explicitly sending IVs. Our performance evaluation shows that pure software implementations without cryptographic accelerators perform well, making it suitable for use in highly constrained IoT devices. In combination, *LCap*, *HMAC2*, and *Payload Encryption* form a complete application layer security framework for constrained CoAP nodes.

References

[Au14] Aumasson, J.-P.; Meier, W.; Phan, R. C.-W.; Henzen, L.: BLAKE2. In: Information Security and Cryptography. Springer Berlin Heidelberg, pp. 165–183, 2014, URL: https://doi.org/10.1007/978-3-662-44757-4_9.

[BEK14] Bormann, C.; Ersue, M.; Keränen, A.: Terminology for Constrained-Node Networks, RFC 7228, May 2014, URL: https://rfc-editor.org/rfc/rfc7228.txt.

[BG17] Buschsieweke, M.; Günes, M.: Securing Critical Infrastructure in Smart Cities: Providing Scalable Access Control for Constrained Devices. In: PIMRC'17 - Workshop on "Personalised Mobile Applications for Smart Cities and Smart Citizens (PMA 2017)". Montreal, Canada, Oct. 2017.

[BG18] Buschsieweke, M.; Günes, M.: Access Control for Medical Devices: Tweaking LCap for Health Informatics. In: IEEE Global Communications Conference (GlobeCom), Workshop on Wireless Energy Harvesting Communication Networks. Abu Dhabi, UAE, Dec. 2018.

[BS16] Bormann, C.; Shelby, Z.: Block-Wise Transfers in the Constrained Application Protocol (CoAP), RFC 7959, Internet Engineering Task Force, Aug. 2016, URL: http://www.ietf.org/rfc/rfc7959.txt.

[CGH16] Chen, B.; Güneş, M.; Huang, Y.-L.: CoAP Option for Capability-Based Access Control for IoT-Applications. In: Proceedings of the International Conference on Internet of Things and Big Data. Scitepress, 2016, URL: https://doi.org/10.5220%2F0005950902660274.

[CRT18] Claeys, T.; Rousseau, F.; Tourancheau, B.: Securing Complex IoT Platforms with Token Based Access Control and Authenticated Key Establishment. In: International Workshop on Secure Internet of Things (SIOT). Oslo, Norway, Feb. 2018, URL: https://hal.archives-ouvertes.fr/hal-01596135.

[GBB15] Gerdes, S.; Bergmann, O.; Bormann, C.: Delegated CoAP Authentication and Authorization Framework (DCAF), draft-gerdes-ace-dcaf-authorize-04, Internet Engineering Task Force, Oct. 2015, URL: https://tools.ietf.org/html/draft-gerdes-ace-dcaf-authorize-04.

[Jo18] Jones, M.; Wahlstroem, E.; Erdtman, S.; Tschofenig, H.: CBOR Web Token (CWT), RFC 8392, May 2018, URL: https://rfc-editor.org/rfc/rfc8392.txt.

[KBC97] Krawczyk, H.; Bellare, M.; Canetti, R.: HMAC: Keyed-Hashing for Message Authentication, RFC 2104 (Informational), Updated by RFC 6151, Internet Engineering Task Force, Feb. 1997, URL: http://www.ietf.org/rfc/rfc2104.txt.

[Ma12] Mahalle, P. N.; Anggorojati, B.; Prasad, N. R.; Prasad, R.: Identity driven capability based access control (ICAC) scheme for the Internet of Things. In: ANTS'12. Institute of Electrical & Electronics Engineers (IEEE), Dec. 2012, URL: http://dx.doi.org/10.1109/ANTS.2012.6524227.

[Sc17] Schaad, J.: CBOR Object Signing and Encryption (COSE), RFC 8152, Internet Engineering Task Force, July 2017.

[Se19] Selander, G.; Mattsson, J.; Palombini, F.; Seitz, L.: Object Security for Constrained RESTful Environments (OSCORE), RFC 8613, July 2019, URL: https://rfc-editor.org/rfc/rfc8613.txt.

[Se20] Seitz, L.; Selander, G.; Wahlstroem, E.; Erdtman, S.; Tschofenig, H.: Authentication and Authorization for Constrained Environments (ACE) using the OAuth 2.0 Framework (ACE-OAuth), Internet-Draft draft-ietf-ace-oauth-authz-33, Work in Progress, Internet Engineering Task Force, Feb. 2020, URL: https://datatracker.ietf.org/doc/html/draft-ietf-ace-oauth-authz-33.

[SHB14] Shelby, Z.; Hartke, K.; Bormann, C.: The Constrained Application Protocol (CoAP), RFC 7252, Updated by RFC 7959, Internet Engineering Task Force, June 2014, URL: http://www.ietf.org/rfc/rfc7252.txt.

[SS75] Saltzer, J.; Schroeder, M.: The protection of information in computer systems. Proceedings of the IEEE 63/9, pp. 1278–1308, 1975, URL: https://doi.org/10.1109%2Fproc.1975.9939.

[TB19] Tschofenig, H.; Baccelli, E.: Cyberphysical Security for the Masses: A Survey of the Internet Protocol Suite for Internet of Things Security. IEEE Security & Privacy 17/5, pp. 47–57, 2019.

R. Reussner, A. Koziolek, R. Heinrich (Hrsg.): INFORMATIK 2020,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2021 1247

Programming IoT applications across paradigms based on WebAssembly

Karl Fessel[1], André Dietrich[1], Sebastian Zug[1]

Abstract: The key to IoT applications' success is the opportunity to exploit data generated by one node for various applications. Solutions for this are either centralized server systems, which aggregate the data and answer corresponding requests from different clients, or the concepts of edge computing, in which individual nodes take over the provision and processing of data directly. Although the advantages of immediate processing are obvious, edge computing concepts have so far been limited to more powerful nodes. Embedded in the DoRIoT project, we transfer the idea to low performance devices. This includes challenging questions related to security, scheduling and coordination issues. Additionally, we have to support the programming process itself. In order to achieve sufficient acceptance in the programming community we have to ensure that "freely programmable" is not bounded by hardware oriented programming paradigms and languages. Furthermore, the developer should be able to implement IoT-requests based on standard building blocks in a programming language of his choice.

In this paper we introduce the architecture and a tool-chain to cope with these challenges based on a WebAssembly-interpreter (WAMR) embedded in the DoRIoT software stack. The prototypical integration provides the applicability of WASM compiler tool-chain, originally focused on web-applications, and supports the orchestration of multiple requests in parallel.

Keywords: IoT; RIOT; DoRIoT; WASM

1 Motivation

The Internet of Things (IoT) [AIM10], Ubiquitous-Computing [We93], Industry 4.0 [HPO15], or Cyber-Physical Systems [Sh11], etc. is a collection of terms, which more or less share the same fundamental idea: in which an assembly of temporal and regional fluctuating heterogeneous systems share their information and capabilities to achieve a certain goal. Capabilities in this case means either sensing, acting, or computational resources.

Although the idea is pretty straight forward, it comes with a variety of yet unsolved problems, such as security and privacy issues, connectivity, the integration of hardware, diverging standards, performance, etc. Especially for low performance embedded nodes these open questions limit flexibility. Due to performance and security issues small sensor/actuator nodes show a closed structure that offers little scope for individual adjustments. In contrast to more powerful edge computing nodes, their behavior cannot be updated or adjusted

[1] Technische Universität Bergakademie Freiberg, Informatik, Germany, {Karl.Fessel, Andre.Dietrich, Sebastian.Zug}@informatik.tu-freiberg.de

doi:10.18420/inf2020_116

according to specific use-cases. Based on a fixed firmware, nodes transmit their measurements unfiltered with a predefined sample rate. The generic configuration intends to balance required communication bandwidth and update frequency to cover the requirements of all applications. But of course, individual messages generated according to tailored requests promise a higher utilization of the node and better system performance[Sh16].

The DoRIoT project[2] intends to overcome this separation between different node performance classes, related to their capability to execute user-specific requests. The project focuses on methods and tools for building self-organized systems, ranging from small sensor nodes (classes C0, C1, and C2 according to RFC7228 terminology) to server solutions. Users specify data aggregation and processing methods, the distributed intelligence assigns the requests to a specific node or a set of nodes according to communication bandwidth, accessible interfaces, timing constraints, etc. Consequently, each request has to be executable on different node architectures and operating systems. Virtual Machines (VM) or interpreters are commonly used to ensure hardware independence of applications. We evaluated their concepts and implementations, related to the chosen node classes, as well as multi-threading and multi-user capabilities, required performance capacities, supported languages, security issues, etc. In parallel, we investigate programming abstractions offered by the provided programming languages and paradigms.

A promising new approach in the field of interpreters are projects that try to transfer WebAssembly concepts to small IoT nodes. WASM is a binary instruction format for a stack-based virtual machine. It was designed as a portable compilation target for programming languages focused on client and server applications. WASM code runs natively in browsers; it is usually run by a combination of a interpreter and different optimizing levels of just-in-time (JIT) and ahead-of-time (AOT) compilation. WASM-bytecode can be generated by a huge number of compilers (from different source languages like C(++), Rust, Go, and many more) most of these compilers are build on top of LLVM-toolchain[3]. In the context of Web and C(++) code, the Emscripten SDK and tool-chain is often used to adapt preexisting C(++) code to the browser, by providing a libc-like API.

Based on the selection process described in Sec. 2 we integrated the WebAssembly Micro Runtime[4] interpreter into our project architecture, which is also supported by the Bytecode Alliance[5]. We identify three basic types of programs/tasks/usage-patterns (request, process, and function) that are typical for sensor networks and by providing generall WASM interfaces we will make these available to many languages.

[2] Dynamic runtime environment for organic (dis-)aggregating IoT-processes
DoRIoT project website `http://www.doriot.net/`

[3] originally "Low Level Virtual Machine", project website `https://llvm.org/`

[4] `https://github.com/bytecodealliance/wasm-micro-runtime`

[5] `https://bytecodealliance.org/`

2 State of the Art

2.1 Programming Languages and Paradigms

According to the IoT Developer Survey held in 2019 by Eclipse Foundation (cf. [Ecl20]), the highest ranked IoT programming languages on constrained devices in 2019 were C, C++, Java, and surprisingly JavaScript. The available programming languages and paradigms are determined by the underlying Operating System (OS) running on the node. For low performance systems with tailored embedded OS (FreeRTOS, Contiki, RIOT OS[6][Ba18]) C is still the dominating language. As an alternative, TinyOS[7] offers a component-based, event-driven task model implemented in nesC[8], a specific C dialect[Ka07; OB09].

In contrast to previous examples, TinyDB offers a more declarative approach. It is a distributed query processing system for extracting information from a network of (smart) TinyOS sensors ([Ma05]). As the name suggests, it interprets a network of sensors similar to a database and, therefore, also applies a SQL-like syntax to collect data from a heterogeneous network of sensors. It borrows the semantics of `SELECT`, `FROM`, `WHERE`, and `GROUP BY` clauses from SQL, but it also offers further features, which have been especially developed to minimize the power consumption in sensor networks, such as life-time queries, dealing with events, or the creation of Semantic Routing Trees (SRT) for power-efficient information and query propagation.

SelectScript [DZK14] was developed while struggling with the dominant imperative programming paradigm in order ease the development effort for embedded systems and their access. It supports Python's data-types and operations, Lua's object-orientation based on prototypes and dictionaries, LISP's higher-order functions, lazy evaluation and tail-recursion. This was combined with a three valued logic (to simplify error handling) and SQL-like query capabilities, that can also be applied to solve reasoning problems. The key idea thereby was, not to be forced to switch between a program and an interface, so that one syntax or notion can be applied for programming but in the same way also be used for querying, no matter how complex or divergent a query might be. It exists an implementation of an VM[9] that had also been tested on 8-Bit microcontrollers.

Although there are approaches to use multi-paradigm languages, it seems to be more sufficient to support different languages on one device, based on the problem and the developer experience. There are other concepts such as miniKanren (see [By09]), for example. miniKanren is yet another relational programming language, but what makes it important in the context of IoT is, that it allows applying temporal logic [Ru18] and since it can run in "both" directions, it can be used either as a theorem-checker or -prover. For example, given the task to a sensor of measuring for ever or as long as possible under certain

[6] RIOT OS project website: `https://riot-os.org/`
[7] TinyOS project website: `http://www.tinyos.net`
[8] Network embedded systems C project website: `http://nescc.sourceforge.net`
[9] SandhillSkipperVM project website: `https://github.com/andre-dietrich/SandhillSkipper`

timing constraints, the sensor could figure out the optimal schedule based on local energy consumption, or at least give the answer why this goal cannot be fulfilled. This information can be either used to pass this task to another sensor, which meets the defined requirements or to re-plan the global task.

These examples show that tasks might need different tools and a flexible tool-chain for developing IoT-applications. This separates the selection of programming languages from the applied OS, and thus, liberates the user to chose the best solution for a given problem.

2.2 Virtual machines and Interpreters

Virtual Machines (VM) implement this request, but are mostly focused on one language. Java VMs support a large bandwidth of devices including embedded devices implementing a write-once-run-everywhere property based on a generic bytecode. Nevertheless, embedded Java code and run-times differ from desktop VMs. The general concept enables code mobility and thus to move code dynamically to different devices. This was utilized for example by JINI (cf. [Wa99]) and the OSGi framework (cf. [LNH03]), which are both so-called "service delivery platforms" that are used to tackle modularization, collaboration and service discovery in distributed systems. In contrast to fixed services traditionally realized in IoT-nodes, JINI and OSGi allow services to be dynamically installed, started, stopped, updated and even uninstalled. Additionally, services and clients can join or leave a federation anytime. Whereby JINI can also load functionality into a process (locally) even while the process is running.

Next to these not so commonly known examples there are of course also ports of other (mostly imperative) languages to realm of embedded systems, such as Python MicroPython[10], PyMite [PNS09], or Zerynth[11]. Unfortunately Zerynth is not mentioned in the Eclipse IoT survey, but next to an embedded Python VM it also offers an OS[12] abstraction and combines with ChibiOS or FreeRTOS, which allows to running Python and C programs in parallel.

There are a number of Projects that support running JavaScript on embedded devices like Espruino[13] (supporting microcontrollers at C2 level) and Tessel[14],which requires significant more resources than C2 in RFC7228-Scale[15]. Alternative solutions use NodeJS on systems starting from RaspberryPi level. They run JavaScript code either by just in time compilation or interpretation, which means the code has to be transported over the network and be either jit-compiled or interpreted on the target, this code can be preprocessed for compactness and/or execution speed making JavaScript a pseudo assembly and VM.

[10] MicroPython project website: `https://micropython.org/`
[11] Zerynt project website: `https://www.zerynth.com/`
[12] ZeryntOS project website: `https://www.zerynth.com/zos`
[13] Espruino project website: `http://www.espruino.com`
[14] Tessel project website: `https://tessel.io/`
Tessel 1 project website: `http://web.archive.org/web/20150213073259/https://tessel.io/`
[15] Tessel 1: ARM-M3 (180MHz, 32MB RAM); Tessel 2: MIPS (580MHz, 64MB RAM)

2.3 WASM implementations for embedded systems

If a VM is too restricive and taylored for one programming language only, the application programmer is not able to choose the most suitable solution for the job anymore. Another approach is to select a general VM to which multiple languages compile, such as the VM-Model for WASM. It exists a huge number of open-source WASM interpreters[16] some of them are applicable to and/or target small embedded systems. Small systems need the VM to run as an interpreter for bytecode without any JIT compilation. We tested a Rust based approach (wasmi[17]), which at the moment does not integrate well with RIOT and its tool chain, in contrast to C that is well supported by RIOT that targets embedded systems and has a good and active developer community. We applied the WebAssembly Micro Runtime[18](WAMR), which is based on a modular approach that makes it very adaptable, furthermore it already supports multiple embedded operating systems. Other interpreters that may fit such systems are WAC[19], which has been ported to ESP32, or WASM3[20] that relies on tail-call optimization, which might be problematic with some compilers. The WASM3 documentation states that is able to run on system starting at ~64Kb for code and ~10Kb RAM. WAMR claims to use 85Kb for the Interpreter and use a low amount of memory, it also provides a greater set of Post-MVP Features than WASM3.

3 Application concept

3.1 Building blocks of the aggregation

The range of potential user codes reaches from aperiodic single shoot accesses on current measurements over periodic data aggregations in combination with smoothing algorithms up to complex event driven aggregation functions, providing domain-specific data formats. For realizing and monitoring the execution of dynamically assigned user byte-codes we have to structure them in predefined building blocks, implemented on a set of abstract interfaces. Hence, we need to identify common patterns in user code and provide related interface implementations in the WASM tool-chain. System functions like `wait()`, `readSensorData()`, or `display()` close the gap between the OS and WASM interpreter.

Based on the requirement analysis of the DoRIoT-Project we identified the three query types representing abstract building blocks of actual user codes, they are described in a pseudo-code semantic:

Request ... part of a program that run once and in most cases answers one question or triggers one action. All resources (RAM, program memory) are free again after a single run.

[16] `https://github.com/appcypher/awesome-wasm-runtimes`
[17] `https://github.com/paritytech/wasmi`
[18] `https://github.com/bytecodealliance/wasm-micro-runtime`
[19] `https://github.com/kanaka/wac`
[20] `https://github.com/wasm3/wasm3`

```
request(){ send me the temperature at sensor 3 };
```

Process ... part of a program that will stay running and process data that is needed for the app to provide its service or part of this.

```
process(){ wait(10 seconds);
query data and save to database };
```

Function ... part of a program that can provides a specific functionality that can be called from nodes within the network. A function extends the interface of the node for all queries, according to permissions.

```
fuction_avgtemp(){return average of saved temperature-data};
```

These conceptional elements are independent from each other and may be combined in an app as they provide orthogonal functionality. Access to these building blocks and to the SystemAPI will be managed by access control and capability management.

3.2 Implementing multi-paradigm aggregation requests

How can basic components support real world applications, coping with a variable number of different aggregation and processing chains? Let's consider a system of IoT-nodes equipped with different sensors (temperature, humidity). Different users transmit queries for data aggregation and processing.

The first example implements an isolated process, continuously calling the system functions `wait()`, `getTemperature()` and `display()`. These functions are implemented in separate headers and linked during compile time. We assume that individual IoT-nodes provide a specific collection of these functions. Continuous updates of a temperature-display that directly mounted to the node is realized by an imperative programming paradigm.

```
process(){
  wait(10 seconds);
  var x = getTemperature();
  display(x);
};
```

Lst. 1: Process configuration for a single sensor data aggregation and output

If we extend the scenario with an additional node that controls a ventilation system, the aggregation process may consider external sensory data too. Hence, an abstract `query()` replaces the system function call from the previous example. The `query()` criteria is evaluated at run-time and references local system functions or/and remote aggregation methods. If external sensor data is relevant, the interpreter spreads out corresponding

`request()` queries to surrounding nodes. This way, the second example implements a spatial criteria (`room`) and amplitude (`>22C`) for filtering in Lst. 2, line 3. In order to realize such potentially complex filters, the query concept has to integrate declarative programming concepts.

```
1  process(){
2      wait(1 minute);
3      if query(num of all room temperatures where temp > 22C)> 0:
4          switchhigh();
5      else:
6          switchlow();
7  };
```

Lst. 2: Ventilation system: instead of requesting it to switch its airflow, a process is installed that automates this task

By adding humidity sensors to the system, its performance can be further improved. The calculation of absolute humidity requires relative humidity information and air temperature, which should be located in-between.

```
1  process(){ //virtual sensor, calculating absolute humidity
2      wait(10 seconds);
3      var sat_humidity = calcSatHumidityatNormalPressure( query(temperature where room is same as
           this ) );
4      var rel_humidity = getHumidity();
5      store abs_humidty = rel_humidity/sat_humidity;
6  };
7
8  function_absHumidity(){ //precalculated value
9      coap_return(abs_humidty);
10 }
11
12 process(){ //ventilation controller
13     wait(1 minute);
14     if query(num of all temperatures where temp > 22C) > 0:
15         switchhigh();
16     else if query(num of all temperatures where temp < 16C) > 0:
17         switchlow();
18     else if coap_request(outdoor, absHumidity) < coap_request(indoor, absHumidity):
19         switchhigh();
20     else
21         switchlow();
22 };
```

Lst. 3: Extending the temperature controlled ventilation system by a humidity sensor

The sensors act either as a function or as a process node, to keeping this information up to date. A single shot calculation would be a request, thus it would require to transport the code every time. This information and the air quality tracked at the outside and the inside over time, to discover trends in combination with a weather service, may be a good measure to decide when to open and close ventilation. A system like this may not need a full climate

control and thus save energy. The required calculation may reside on a node that has access to all of this data and may publish control information for the ventilation system. Such a compilation needs memory and access to the weather service and therefore may reside on an edge node. It spawns multiple processes and functions which send requests to other nodes.

The three pseudo code examples illustrate the vision of the project, an intuitive combination of context-sensitive building blocks with complex logical statements. The user code accesses actual hardware functions through system-API-functions for timing, memory access, periphery access, and communication. This functionality gets combined and distributed via self-defined functions.

4 Integration and Implementation

DoRIoT integrates the multi-user approach on three layers into the general system architecture on top of RIOT OS. Actual user queries are realized by so-called DoRIoT-Apps, combining the mentioned building blocks. `WAMR` implementation of a WASM interpreter connects the app level and the OS, supervised by the Runtime Access Control layer. An integrated supervisor creates and manages modules as well as VM instances and generates monitoring information. The general access to OS interfaces is controlled by an internal a Run-time Access Control Unit, control information will be provided by LCap-Lightweight Capability Based Access Control [BG17].

The transfer of WASM code is managed by CoAP. The CoAP message may contain further attributes that enable the evaluation of LCap's access rules and performance restrictions of an app.

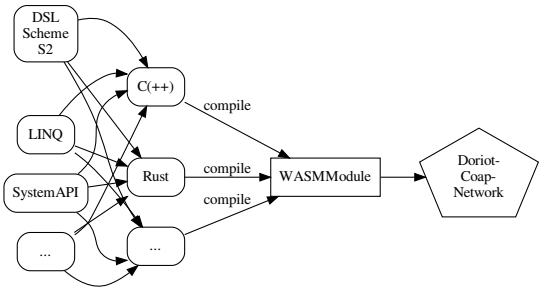


Fig. 1: DoRIoT compile chain

Based on WASM integration and corresponding tools the system already support the inclusion of multiple programming languages. The availability of WASM as a target within the LLVM-tool-chain further help the adaption of its massive number of llvm-front-ends that implement the translation of source-code and therefor programming languages within the LLVM toolchain. Other languages or combinations of multiple may be compiled in a multi-step process. Fig. 1 illustrates this compile-chain.

We emphasized the need of a multi-paradigm concept while developing code for IoT-systems and intend to realize the combination of imperative and declarative program parts following

the LINQ (Language Integrated Query) concept. While LINQ[21] itself is a specific group of implementations within .Net, its approach was ported to many other languages. The example shows how the pseudo-code `query` might be adapted to a LINQ programming style, that is already parseable for (pre)compilers with numerous adaptations of this concept.

```
request(){
    TermperaturSensors.update();
    coap.return( Query( from(s, TermperatureSensors).where( s.temp > 22).orderby(s.temp)));
}
```

Lst. 4: Query and filter temperature sensors based on Lst. 2 (similar but extended)

5 Outlook & Summary

We believe that there is a strong desire in porting different programming paradigms to IoT applications, even to the smallest devices, since it liberates the development process. Some standard tasks can and shall be realized in C while others can concentrate onto functional or logical programming, or execute snippets in order to enable complex queries, and thus shift some of the "global" task's execution logic down to the end devices.

The basis, of course, is a working API that allows to access an all-of-systems-service for all tasks (written in different languages). The application of a VM furthermore enables some form of service control, that is not common in this particular case. On the one hand, it is possible to restrict the memory consumption, which is vital. On the second hand, the usage of an API and the VM's possibility to enforce restrictions, it is also possible to define more fine granular access and execution control for different tasks.

Acknowledgements

This work is funded by the German Federal Ministry of Education and Research (BMBF) as part of the project "DoRIoT" under grant number 01IS18071A.

References

[AIM10] Atzoria, L.; Ierab, A.; Morabitoc, G.: The Internet of Things: A survey. Computer Networks 54/15, pp. 2787–2805, 2010.

[Ba18] Baccelli, E.; Gündogan, C.; Hahm, O.; Kietzmann, P.; Lenders, M.; Petersen, H.; Schleiser, K.; Schmidt, T. C.; Wählisch, M.: RIOT: An Open Source Operating System for Low-End Embedded Devices in the IoT. IEEE Internet of Things Journal, Vol. 5, No. 6, pp. 4428-4440/, Dec. 2018.

[21] .Net LINQ Manual `https://docs.microsoft.com/en-us/dotnet/standard/using-linq`

[BG17] Buschsieweke, M.; Güneş, M.: Securing critical infrastructure in smart cities: Providing scalable access control for constrained devices. In: 2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC). 2017.

[By09] Byrd, W. E.: Relational programming in miniKanren: Techniques, Applications, and Implementations, PhD thesis, Indiana University, 2009.

[DZK14] Dietrich, A.; Zug, S.; Kaiser, J.: SelectScript: A query language for discrete simulations. ACM Transactions on Programming Languages and Systems (TOPLAS)/, 2014.

[Ecl20] URL: https://iot.eclipse.org/community/resources/iot-surveys/assets/iot-developer-survey-2019.pdf, visited on: 06/08/2020.

[HPO15] Hermann, M.; Pentek, T.; Otto, B.: Design Principles for Industrie 4.0 Scenarios: A Literature Review, tech. rep., Technical University Dortmund, Fakulty of Mechanical Engineering, 2015.

[Ka07] Kabadayi, S.; Julien, C.; O'Brien, W.; Stovall, D.: Virtual sensors: a demonstration. In: The 26th international conference on computer communications: demonstrations track (Infocom). Pp. 10–12, 2007.

[LNH03] Lee, C.; Nordstedt, D.; Helal, S.: Enabling Smart Spaces with OSGi. Pervasive Computing, IEEE 2/3, pp. 89–94, 2003.

[Ma05] Madden, S. R.; Franklin, M. J.; Hellerstein, J. M.; Hong, W.: TinyDB: An acquisitional query processing system for sensor networks. ACM Transactions on Database Systems (TODS) 30/1, pp. 122–173, 2005.

[OB09] OBrien, W. J.; Julien, C.; Kabadayi, S.; Luo, X.; Hammer, o.: An architecture for decision support in ad hoc sensor networks. Electronic Journal of Information Technology in Construction 14/, pp. 309–327, 2009.

[PNS09] Pedersen, R. U.; Nørbjerg, J.; Scholz, M. P.: Embedded programming education with lego mindstorms nxt using java (lejos), eclipse (xpairtise), and python (pymite). In: Proceedings of the 2009 Workshop on Embedded Systems Education. Pp. 50–55, 2009.

[Ru18] Rudavsky-Brody, N.: Temporal Logic, μKanren, and a Time-Traveling RDF Database. ACM on Programming Languages/, 2018.

[Sh11] Shi, J.; Wan, J.; Yan, H.; Suo, H.: A Survey of Cyber-Physical Systems. In: International Conference on Wireless Communications and Signal Processing. IEEE, pp. 1–6, 2011.

[Sh16] Shi, W.; Cao, J.; Zhang, Q.; Li, Y.; Xu, L.: Edge computing: Vision and challenges. IEEE internet of things journal 3/5, pp. 637–646, 2016.

[Wa99] Waldo, J.: The JINI Architecture for Network-Centric Computing. Communications of the ACM 42/7, pp. 76–82, 1999.

[We93] Weiser, M.: Ubiquitous computing. Computer 26/10, pp. 71–72, 1993.

R. Reussner, A. Koziolek, R. Heinrich (Hrsg.): INFORMATIK 2020,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2021 1257

Service Migrations in TSCH Network using Wireless Channel Estimation and Prediction

Ali Nikoukar[1], Saleem Raza[1], Tharakeswara Rao[1], Mesut Güneş[1], Behnam Dezfouli[2]

Abstract: Industrial Wireless Sensors and Actuators Networks (IWSANs) are gateway to the Industrial 4.0, which promises to realize smart factory leading to the Industrial Internet of Things (IIoT). It employs Cyber-Physical Systems (CPSs) to enhance operational efficiency and flexibility while reducing cost. IWSANs are delay-sensitive and always require low latency and reliable connection from sensor to actuator to successfully perform a physical action. Reliability and low-latency complement each other to prevent expected failures in wireless medium. In this way, detecting and predicting failure before it actually occurs is key to actually avoid it well in time. Detection and predictions are imperative in locating faults and failures. The causes of failures in a sensor or actuator can include hardware malfunction, poor battery life, interference, accident, and short term wireless connectivity problems. Although, industrial environment mostly undertakes redundant resource to circumvent such issues, yet poor coordination among multiple resources and inaccurately predicting failures may result in losses. In such a scenario, migration of services come to be a rescue, where an intermediary can migrate service from one device, which cannot complete a task due to resource exhaustion, to a more resource-rich device.

Thus, in this paper, we focus on wireless connectivity failures caused by interference in the 2.4GHz frequency band. We do it by designing an Multi Channel Sniffing Setup (MCSS) testbed, that acts as a spectrum observer and is deployed in different locations in industrial WSAN. Alongside, we use the concept of Cognitive Radio (CR) to predict interference and noise level in the spectrum by proposing an Intelligent Low-power Wireless Spectrum Prediction (ILPWSP) based on Deep Q Network (DQN). The MCSS testbed and the ILPWSP coordinate in assessing wireless connectivity risks, predict failures in sensor and actuator nodes and then make efficient decisions on the migration of services from one device to another device. Our results show the feasibility of spectrum prediction with an acceptable ratio for reliable IWSN.

Keywords: IWSAN; TSCH; Channel Prediction; Migration service

1 Introduction

The proliferation of low-cost processors along with low-power wireless technologies and advancement in the production of small high-performance microprocessors have enabled the Internet of Things (IoT) [Bl10; Gü09; Ni18]. It is predicted that the number of devices connected to the Internet will increase by 75 billion devices [Na18], by 2025. It includes

[1] Otto-von-Guericke Universität, Communication and Networked Systems, Universitätsplatz 2, 39106 Magdeburg, ali.nikoukar@ovgu.de, saleem.raza@ovgu.de, tharakeswara.paddolkar@ovgu.de, mesut.guenes@ovgu.de
[2] Internet of Things Research Lab Department of Computer Science & Engineering Santa Clara University Santa Clara, USA, bdezfouli@scu.edu

doi:10.18420/inf2020_117

applications such as smart homes, smart cities, and smart factories. Each application has its unique Quality of Service (QoS) requirements. For example, while video surveillance requires high-throughput, timeliness is critical in applications such as autonomous driving to avoid a fatal accident. Industry 4.0 is one of the main domains benefiting from IoT by employing Cyber-Physical Systems (CPSs). It is predicted that by 2026, the worldwide market for industrial wireless will reach 7 Billion dollars. Furthermore, over 3 million robots will operate in industries by 2020. In some cases, robots and actuators will be responsible for a critical task that has to be executed in real-time. According to International Society of Automation (ISA) based on the QoS requirements three categories are defined. Fig. 1 explains the importance of timeliness in safety and control applications. Because wired networks suffer from issues such as scalability, mobility, and high cost, there is a need for reliable wireless solutions to guarantee low-cost, flexibility, and packet delivery in real-time.

Category	Class	Application	Description
Safety	0	Emergency action	Always critical
Control	1	Closed-loop regulatory control	Often critical
	2	Closed-loop supervisory control	Usually non-critical
	3	Open-loop control	Human in loop
Monitoring	4	Alerting	Short-term operational consequence
	5	Logging and downloading/uploading	No immediate operational consequence

TSCH
IEEE 802.15.4
Increasing timeliness requirement

Fig. 1: Different classes of industrial applications defined by ISA

Despite such benefits, wireless solutions need to be energy efficient as the sensor/actuator nodes are battery-powered. Because radio is the major cause of energy consumption in IoT devices. An extended radio operation time will reduce battery life, thus an unexpected death of nodes will harm network reliability. Medium Access Control (MAC) protocols are designed to manage and schedule wireless communication, but the static nature of MAC protocols fail to predict a highly dynamic wireless spectrum.

In this regard, an efficient protocol called Time-Slotted Channel Hopping (TSCH) has been proposed as part of the IEEE 802.15.4 standard, which dynamically involves channel hopping to overcome channel impairments such as interference and fading. However, such random channel hopping still suffers from dynamic channel conditions at different transmission times and locations, which makes some nodes highly prone to transmission failures at one location while other nodes having a higher likelihood of transmission success at a different location. In many cases, it can be beneficial to assign the task of another sensor/actuator nodes, who can not complete their task optimally, and then reassign the same task to nodes who are more capable. This can be achieved by migrating the code to another device with similar resources and adopting its functionality to the task requirements. Failure in Industrial Wireless Sensors and Actuators Networks (IWSANs) can be due to many reasons such as defected parts, accidents, or poor network connectivity.

In this work, we focus on the lack of a reliable wireless connection because of interference in an operational location. Most of the time, the industrial environment is harsh for wireless transmissions due to the operation of various wireless networks such as surveillance cameras and Wi-Fi access points, as they may cause major interference for Industrial Wireless Sensor

Networks (IWSNs). Interference may cause a node to fail its transmission, and this not only wastes energy for re-transmission but also causes increased latency which may be a threat to real-time operation for IWSNs applications. In such cases, migration of service can play a major role, in which the case at a neighboring distance, a device with similar capability might be available to provide the same services whose interference level is lower. In this way, service migration can ensure that the entire system works reliably and with optimal resource usage. However, decision making an important part of the migration of services. The system has to compute the cost of migration based on application requirements such as latency, energy efficiency, and reliability in wireless transmission. In addition, in real-time networks, failure needs to be predicted intelligently to meet the task deadline constraints. Accuracy in interference prediction is critical to decide if because of connectivity conditions, the device is capable to accomplish the task or not. In such cases, we witness the use of Cognitive Radio (CR) to observe, learn, predict, and provide link quality estimations.

To this end, many researchers have suggested embedding machine learning in network design. Consequently, with the integration of Software-Defined Radio (SDR) and machine learning, CR algorithms are developed to control network parameters intelligently. The idea is to have a cycle of sensing, learning, and decision making by considering the consequence of decided actions as feedback for the learning process. However, training of algorithms in machine learning is an extremely time-consuming process, which makes it an undesirable solution for time-sensitive wireless networks. Approaches, such as cloud radio, are proposed to overcome this limitation by handing over the process to powerful servers located in the cloud. But even then, because of communication distance between the cloud and wireless transceivers, there is a significant delay in exchanging data. Recently, intermediary solutions such as fog and edge computing are proposed to fulfill the latency gap. Still, due to freshness of advancement in designing high power processing units (i.e., GPU) on single board computers and complexity of implementation of lightweight (i.e., energy-efficient and low bandwidth occupancy), knowledge transfer from the end node to fog server is missing.

In this paper, we develop a Intelligent Low-power Wireless Spectrum Prediction (ILPWSP) model based on *Q-learning* algorithm to predict the interference in the wireless spectrum. To provide the training data set for ILPWSP, we design a Multi Channel Sniffing Setup (MCSS) to sense the wireless spectrum concurrently and constantly. Consequently, the network manager will have real-time information about the interference in each location and it can determine the high-risk neighborhood for wireless transmission in terms of packet loss. Packet loss risk identification helps the network manager to migrate the service to the less risky location by assigning the task to a device with similar capability and resources. In this way, it can potentially increase the chance of successful transmission at a new node and location. Our results show that using ILPWSP we can achieve a reliable degree of accuracy in noise prediction in the wireless channel.

The rest of the paper is organized as follows. A brief overview and introduction is given in Section 2. The methodology of research is explained in Section 3. Section 4 presents results and finally, we conclude the paper in Section 5.

2 Background

Real-time wireless communication in IWSAN: IWSANs are the integral parts of the industry ecosystem, they could be deployed in many industrial applications resulting in concepts like CPSs, smart factory, etc. IWSANs integrates sensor networks as well as actuator networks, they complement each other in sensing and then performing required actions. This sensor-actor integration helps achieve the autonomy of many industrial processes and control systems which results in less human intervention. A rising trend of using IWSANs is seen owing to the compelling benefits of wireless networks such as low-cost of deployment and maintenance, and the flexibility and self-organizing features of sensors and actuators networks. The communication between sensors and actuators require reliable transmission of data to successfully support a mission-critical industrial application. Often such reliability is compromised due to wireless channel impairments like fading, interference, collisions, and noise. Industrial applications inherently require low-delay (real-time) and high reliability, otherwise, a successful transmission from the sensor to the actuator is difficult to maintain. Further, the industrial environment is harsh owing to the presence of heavy machinery, high-temperature conditions, high voltage induction, electrical motors, and drives operating at high voltage. Alongside this, there could be other wireless networks operating in the unlicensed 2.4 GHz spectrum. Such an environment could pose threats to reliable communication for IWSANs and hence it may compromise on the required QoS for a given industrial application. A typical scenario of wireless link failure due to the interference between an actuator and a network manager is depicted in Fig. 3(a). Mostly TSCH employs centralized network architecture, as it is widely preferred in the industrial environment due to its ease of management compared to distributed architecture [De14]. In such architecture, a network manager is responsible to assesses the overall network and takes care of the scheduling of nodes, updates the list of best channels, and selects best routes, and security measures. Employing multi-channel operation and random selection of channels instead of a single operation reduces the risk of collision because of interference. This architecture is also commercialized and used in industrial wireless network technologies such as WirelessHART. Although, channel hopping has the potential to circumvent the effects of fading and interference, yet many wireless technologies and standards share 2.4 GHz band. This sharing of the spectrum with technologies like WiFi and Bluetooth makes 2.4 GHz band crowded resulting in degrading each other's performance. Many researchers studied the impact of interference on link failure in TSCH network. For example, authors in [ZPD18] studied the cost a benefit of channel blacklisting in TSCH network. The study analyzes the local or the global implementation of channel blacklisting suggested by many researchers [Ko17]. However, the concept of channel blacklisting in TSCH is involved with the cost of delay because of channel observation and negative impact on timeliness as a result. Therefore, researchers propose the concept of CR [Mi02] using artificial intelligence and smart radio to find the interference-free time slots in the wireless spectrum. Especially with the advancement in powerful processors and lightweight machine learning algorithms, the idea of deployment of CR is becoming more practical.

Cognitive radio: Cognitive Radio (CR) in a wireless network consists of three main parts: sensing the wireless spectrum to provide a dataset that can be fed to the machine learning algorithm, a desirable machine learning model to predict interference, and making the decision to tune transmitter parameters to avoid collision and achieve required Packet Error Rate (PER). Fig. 2 shows the basic concept of CR. Below we describe important actions performed by CR.

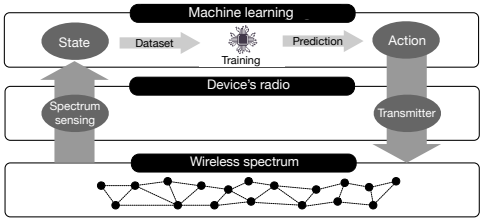


Fig. 2: The basic concept of cognitive radio for wireless network.

A) Sensing. In wireless communication and specifically, MAC layer channel sensing is an efficient approach to determine channel conditions. Generally, it is assumed that the device itself is responsible to sense channel and make transmission decisions. On one hand, this method has the advantage of higher accuracy, because the same radio in the same location is listening to the wireless channel and the interference level in the neighbor location may vary. But, this is costly in terms of power consumption due to the longer radio operation time. Cooperative and external sensing is proposed to solve this issue. In this way, wireless devices can share their observation and can have a more accurate estimation of wireless channel condition. Another benefit of this method is to assign the sensing task to devices with a constant power source to save energy for low-power devices.

B) Prediction. An optional interference prediction algorithm is critical to avoid collisions and transmission in free time slots in the spectrum. Because IEEE 802.11 has a higher data rate, transmit power, and wider channels, it is the main cause of interference in the 2.4 GHz frequency band for IEEE 802.15.4 networks. As a result, for IEEE 802.15.4 transmission in interference-free time is more desirable to save power while maintaining acceptable PER and efficiently use the shared wireless spectrum. In this direction researchers in [DT18] use Reinforcement Learning (RL) as a machine learning and prediction algorithm to optimize transmission success by decision making for channel selection in TSCH network. Sensing the spectrum helps find these interference-free gaps, however, the spectrum is very dynamic, and sensed data can lose their validity over time. In addition, to increase the accuracy of the prediction results, the machine learning algorithm demands more training samples. Low-power wireless devices are designed to save power by minimizing the radio operation and they are not capable of providing continuous high frequency sensed samples. Besides, they need to transfer these samples to more powerful computers such as fog or cloud to avoid wasting energy because of the training process. Although CR is an intelligent solution to optimize spectrum efficiency, yet in a highly crowded spectrum, it may fail to maintain

a reliable connection link. Still, CR can help understand the risk of transmission on a certain wireless link. In high risky scenarios for wireless transmission, central management networks such as TSCH can assign the task to devices with a similar capability and resources placed in interference-free locations [RFG19]. This concept is called the migration of service and it is achievable using the virtualization machine to be operating system agnostic to execute any codes written for different platforms.

The need for service migration in IWSAN: In traditional IWSNs scalability is challenging because of the direct connection of the end node and cloud service. The reasons for this challenge are bandwidth limitation and response delay because of physical distance of cloud service from operational node and increasing the size of collected data to process and analyze. The concept of IoT at the edge or introducing the intermediate fog nodes helps to increase the scalability and timeliness by dis-aggregation of services.

3 Migration service architecture using MCSS testbed and ILPWSP

In our design we deploy several multi-channel sniffing devices alongside the operating nodes in IWSNs. In this case, MCSS continuously observes and monitors the interference conditions in the target location and provides feedback to the central manager. When a transmission in the location A has a high risk in terms of packet loss, the central manager can assign the task to another device in location B with similar resources. We consider a typical industrial network architecture as shown in Fig. 3, which comprises of existing architecture in Fig. 3(a) and our proposed architecture in Fig. 3(b). Industrial environments are full of wireless devices such as surveillance cameras, WiFi access points, and smartphones. The operation of these devices in the neighborhood of the industrial sensors and actuators causes interference in the wireless channel. Lack of reliable connection for real-time IWSNs may cause a severe impact on the functionality of the entire network. Our proposed solution for this problem is presented in Fig. 3(b), where we introduce the MCSS and fog radio for interference prediction. In this scenario, MCSS is continuously monitoring the wireless channels and feeding the machine learning based prediction model deployed in the fog computing.

Multi Channel Sniffing Setup (MCSS): As shown in Fig. 3(c), MCSS consists of 40 nRF52840 USB dongles, each dongle is responsible to sniff a single channel with 2MHz width in 2.4 GHz frequency band. The nRF52840 includes an ARM Cortex-M4 processor with 1MB flash memory, 256KB RAM, and has -95dBm antenna receiver sensitivity. This setup allows us to collect noise samples of all the channels with 9μs interval for major low-power wireless technologies operating in 2.4 GHz, such as different versions of Bluetooth Low Energy (BLE), IEEE 802.15.4, and 2.4 GHz proprietary protocols. The observation provides real-time feedback about the noise in wireless spectrum in a defined location. Using the collected dataset ILPWSP can predict the channel condition in the future. ILPWSP helps to reduce the risk or PER caused by interference, by assigning the task to

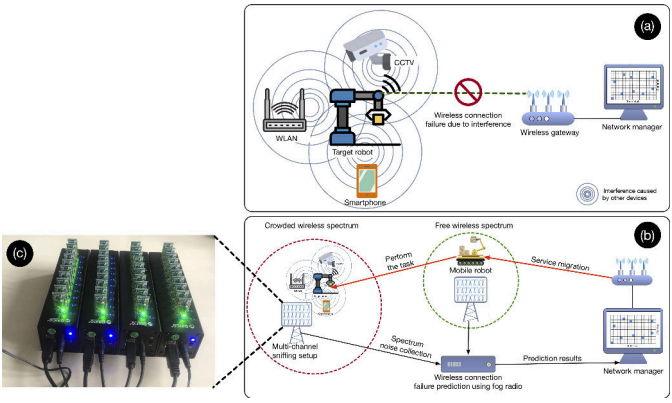


Fig. 3: The proposed network scenario

other devices with similar resources and capability. The output recorded by this tested setup is fed to the fog radio.

Intelligent Low-power Wireless Spectrum Prediction (ILPWSP): Reinforcement Learning (RL) is one of the active research areas in machine learning for time series prediction. In RL, the environment, and defining adaptive policy using perceived states of the environment helps to improve the accuracy of decision making and continuous adaptation with the environment. RL is also known as a semi-supervised machine learning algorithm, because it receives feedback from its previous actions to improve accuracy over time. In the wireless spectrum, due to the high variation of interference level in a short period, RL algorithms are suitable to continuously observe and predict. The key entities in RL are *agent*, *environment*, *actions*, *rewards*, and *states*. The *agent* interacts with the *environment* and takes *action*. In return, it retrieves the *rewards* or *penalty* from the effect of previous *action* in the *environment*. This feedback helps to improve decision accuracy. In ILPWSP, the *agent*, which is our Q-learning model, interacts with the wireless spectrum environment based on the configuration parameters given through actions. In each corresponding state, the agent receives the reward based on the action. In our simulation experiments, the agent keeps track of all the errors based on the actions taken, and the rewards received, through this online learning, it generates an optimal policy to minimize those errors. The agent minimizes errors by comparing error values with the available training data set, which we feed as input to the model. In the next step, the model is validated through testing data set so as to examine if the optimal parameters selection is performed with reasonable prediction accuracy. The optimal configuration parameter is expressed as Markov Decision Process (MDP) [SB18]. Among many variants of RL, *Q-learning* is a unique approach of online learning. It arrives at a policy based on a Q-table which stores the results of actions taken from a given state. ILPWSP is based on *Q-learning* and it can navigate high dimensional configuration parameter space depending on strategy value function Q. Equation (1) explains the *Q-learning* where Q is a state-action value denoted by $Q(s_t, a_t)$ and works as follows.

Agent observes the environment and performs an action in it and aims towards maximizing the expected reward. For real-time spectrum prediction in a complex wireless environment, *Q-learning* is efficient to find out the best policy of RL based on value function. *Q-Learning* is an incremental algorithm that determines optimal policy in a step by step process at each step t, agent observes current state s_t, selects and performs an action a_t and observes the next state s_{t+1} in the process, receives reward r_t, and finally updates the Q values $Q(s_t, a_t)$.

$$Q_{t+1}(s_t, a_t) = Q_t(s_t, a_t) + \eta \left[r_t + \gamma \min_{a_t} Q_t(s_{t+1}, a_{t+1}) - Q_t(s_t, a_t) \right] \quad (1)$$

Process repeats until Q value function converges to an optimal value as $Q_{t+1}(s_t, a_t) \rightarrow Q_*(S, A)$.

In this paper, we implement ILPWSP model based on DQN network which is a variant of *Q-Learning*. The ILPWSP uses grid search method to find optimal configuration parameters such as testing and training data sizes. In ILPWSP, we limit the range of configuration parameters to batch size, epochs, hidden layer nodes, input dimension, and difference order. The batch size is the number of samples to input into the model. An epoch is defined as passing of data set forward and backward once in the whole network. Hidden layers nodes act as intermediate nodes which add weights to the inputs and perform an activation function on them to produce outputs. We pass the data set multiple times into the same neural network. We collected the interference samples in an office environment operating with IEEE 802.11 enabled APs. The samples were collected on channel 11 of IEEE 802.15.4. In the dataset, we witness non-stationary behavior in different timestamps, therefore, we need to convert the dataset to stationary series by using the difference transformation technique. The main goal of ILPWSP is to predict the interference level on channel 11 in IEEE 802.15.4 with the help of DQN, where we try to find optimal policy by tuning the available parameters. The number of steps taken is 10 as selected by the grid search.

4 Results and Discussions

In this Section, we present and discuss the interference prediction results using ILPWSP. Furthermore, we compare our results with State Action Reward State Action (SARSA) as an implemented baseline algorithm to evaluate the performance of ILPWSP.

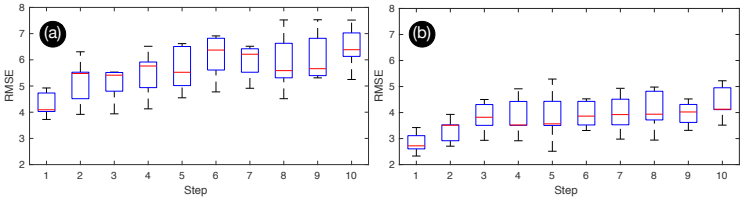


Fig. 4: (a) SARSA and (b) ILPWSP results for channel 11 in IEEE 802.15.4.

The results for the evaluation is presented in Fig. 4, where Fig. 4(a) shows the prediction results for SARSA and Fig. 4(b) demonstrates the prediction results for ILPWSP. As can be observed in the Y-axis, we use the RMSE metric to evaluate the performance of the prediction. In each part, we train and predict 0.5 million, and the samples are divided into train and test size of 0.30 and 0.20 million at each step. The optimal sample size is selected by the grid search and agent navigates through the environment and along with each step size until a terminal state. In Fig. 4(b), at the starting of the steps, the agent left free to randomly explore the environment and learning takes place by considering all possible configuration parameters of batch size, epochs, input dimension, difference order, and hidden layer node. At the second step the agent starts learning, and the error is relatively high, where the ratio of exploration is balanced and the agent by selecting all configuration parameters that result in a high error. At the third step the error has gradually decreased the agent avoided selecting the similar parameters. From the fourth step the model gradually minimize the error. At the final step, policy by the ILPWSP agent selects the optimal configuration by gaining the confidence of parameter selection and achieves low prediction error. However a detailed look at the Fig. 4(b), we notice that several trails and the number of samples length respectively the error is minimized, which determines agent selecting the right actions that result in a low error. Where in the SARSA, error increases constantly because it only considers local optimal value as the best value. While the ILPWSP follows the greedy policy. In SARSA, it takes the policy strategy into account and joins into its updates and refreshes by considering the approach of previous actions. In Fig. 4(a) shows the values of SARSA approach and concludes the result it is unable to converge the values and shows high variance. Although, in times of low interference, ILPWSP and SARSA show almost the same performance, however, when interference increases ILPWSP shows its strength over SARSA. The little differences in performance can impact the timeliness of the network and inaccuracy in the prediction which may cause collision. In this way, ILPWSP serves critical feedback that helps decide the link reliability to predict failure and efficiently make decisions of service migrations.

5 Conclusion

In this paper, we introduced ILPWSP model based on CR network for IEEE 802.15.4 to enhance the reliability and timeliness of the network. Our design is a hybrid approach that consists of three major elements: First we design a MCSS for external cooperative sensing method in CR. Second, we use fog radio to dis-aggregate the computing for machine learning in ILPWSP to achieve low latency. Last, for prediction part in ILPWSP we develop a DQN model. Our results for ILPWSP prove the feasibility of spectrum prediction for decision making to migrate the service in case of a high risk of interference.

Acknowledgment: This work is funded by the German Federal Ministry of Education and Research (BMBF) as part of the project “Doriot” (Dynamic runtime environment for organic (dis-)aggregating IoT-processes) under grant number 01IS18071A. See the project website at `http://www.doriot.net`.

References

[Bl10] Blywis, B.; Güneş, M.; Juraschek, F.; Hofmann, S.: Gossip routing in wireless mesh networks. In: 21st Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications. IEEE, pp. 1572–1577, 2010.

[De14] De Guglielmo, D.; Seghetti, A.; Anastasi, G.; Conti, M.: A performance analysis of the network formation process in IEEE 802.15. 4e TSCH wireless sensor/actuator networks. In: 2014 IEEE Symposium on Computers and Communications (ISCC). IEEE, pp. 1–6, 2014.

[DT18] Dakdouk, H.; Tarazona: Reinforcement Learning Techniques for Optimized Channel Hopping in IEEE 802.15. 4-TSCH Networks. In: Proceedings of the 21st ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems. Pp. 99–107, 2018.

[Gü09] Güneş, M.; Juraschek, F.; Blywis, B.; Mushtaq, Q.; Schiller, J.: A testbed for next generation wireless network research. PIK-Praxis der Informationsverarbeitung und Kommunikation 32/4, pp. 208–212, 2009.

[Ko17] Kotsiou, V.; Papadopoulos, G. Z.; Chatzimisios, P.; Theoleyre, F.: Label: Link-based adaptive blacklisting technique for 6tisch wireless industrial networks. In: Proceedings of the 20th ACM International Conference on Modelling, Analysis and Simulation of Wireless and Mobile Systems. Pp. 25–33, 2017.

[Mi02] Mitola, J.: Cognitive radio. An integrated agent architecture for software defined radio./, 2002.

[Na18] Nawaratne, R.; Alahakoon, D.; De Silva, D.; Chhetri, P.; Chilamkurti, N.: Self-evolving intelligent algorithms for facilitating data interoperability in IoT environments. Future Generation Computer Systems 86/, pp. 421–432, 2018.

[Ni18] Nikoukar, A.; Raza, S.; Poole, A.; Güneş, M.; Dezfouli, B.: Low-Power Wireless for the Internet of Things: Standards and Applications. IEEE Access 6/, pp. 67893–67926, 2018.

[RFG19] Raza, S.; Faheem, M.; Guenes, M.: Industrial wireless sensor and actuator networks in industry 4.0: Exploring requirements, protocols, and challenges—A MAC survey. International Journal of Communication Systems 32/15, e4074, 2019.

[SB18] Sutton, R. S.; Barto, A. G.: Reinforcement learning: An introduction. MIT press, 2018.

[ZPD18] Zorbas, D.; Papadopoulos, G. Z.; Douligeris, C.: Local or global radio channel blacklisting for ieee 802.15. 4-tsch networks? In: 2018 IEEE International Conference on Communications (ICC). IEEE, pp. 1–6, 2018.

Methoden und Anwendungen der Computational Humanities

Methoden und Anwendungen der Computational Humanities

3. Workshop der Fachgruppe Informatik und Digital Humanities (InfDH)

Manuel Burghardt[1], Claudia Müller-Birn[2]

Im dritten Jahr des InfDH-Workshops wird eine aktuelle Entwicklung innerhalb der Digital Humanities-Community aufgegriffen, bei der sich unter dem Schlagwort der „Computational Humanities“[3] in zunehmendem Maße ein eigener Teilbereich entwickelt, der primär auf statistische und algorithmische Analyseverfahren in den Geistes- und Kulturwissenschaften abzielt. Computational Humanities bedeutet dabei im Wesentlichen eine Spezialisierung und Profilierung innerhalb des *big tent*[4] der Digital Humanities, dessen breitgefächertes Spektrum von Digitalisierungs- und Modellierungsverfahren, digitalen Ansätzen für das Publizieren, Kommunizieren und Lehren bis hin zur Beschäftigung mit digitalen Kulturphänomenen (bspw. Computerspiele und eBooks) reicht.

Für die Informatik ergeben sich für den Bereich der Computational Humanities genuine Herausforderungen im Spannungsfeld von Data Science, Visual Analytics und Research Software Development:

- Entwicklung von Algorithmen und Analysetools zur Nutzung in originär geistes- und kulturwissenschaftlichen Forschungskontexten;
- Entwicklung von computergestützten Visualisierungen oder Ansätzen aus dem Bereich der Visual Analytics und Integration in die geistes- und kulturwissenschaftliche Forschung;
- Entwicklung und Evaluation maschineller Lernverfahren und statistischer Methoden für die Analyse von Text / Bild / Audio / Video-Daten in geistes- und kulturwissenschaftlichen Forschungskontexten;

[1] Computational Humanities, Universität Leipzig, burghardt@informatik.uni-leipzig.de
[2] Human-Centered Computing, Freie Universität Berlin, clmb@inf.fu-berlin.de

[3] Vergleiche auch die Aktivitäten der Computational Humanities Research Community (`https://cohure.github.io/CoHuRe/`), die sich 2019 informell gegründet hat und 2020 einen gleichnamigen internationalen Workshop durchführen wird. Ebenso das bereits im Jahre 2014 durchgeführte Dagstuhl-Seminar zum Thema „Computational Humanities - bridging the gap between Computer Science and Digital Humanities“ (`https://www.dagstuhl.de/14301`).

[4] Terras, M. (2011). Peering Inside the Big Tent: Digital Humanities and the Crisis of Inclusion. `http://melissaterras.blogspot.com/2011/07/peering-inside-big-tent-digital.html`

doi:10.18420/inf2020_118

- Anpassung und Weiterentwicklung von Verfahren des Data Mining, des Information Retrieval und des Natural Language Processing für originär geistes- und kulturwissenschaftlichen Forschungskontexte;
- Adaption und Weiterentwicklung von computergestützten Verfahren aus anderen Bereichen (bspw. Bioinformatik oder Signalverarbeitung) für die Geistes- und Kulturwissenschaften;

Der vorliegende Workshop-Band „Methoden und Anwendungen der Computational Humanities“ dokumentiert zahlreiche Fallstudien und Erfahrungen zum Einsatz von informatischen Methoden im Sinne der Computational Humanities. Die Beiträge decken dabei ganz unterschiedliche Themengebiete ab, die von „Computer Vision“ und „Optical Character Recognition“ bis hin zu „Text Mining“ und „Geo-Visualisierung“ reichen. Insgesamt liegen neun Beiträge vor (Annahmequote 60 %), die jeweils von drei unabhängigen Gutachter_innen des Programmkomitees[5] anonym bewertet wurden. Unser herzlicher Dank gilt den Autor_innen und Gutachter_innen.

[5] Eine vollständige Liste des Programkomitees sowie weitere Informationen zum Workshop findet sich online: https://fg-infdh.gi.de/infdh-worskshop-2020

R. Reussner, A. Koziolek, R. Heinrich (Hrsg.): INFORMATIK 2020,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2021 1271

SubRosa: Determining Movie Similarities based on Subtitles

Jan Luhmann[1] Manuel Burghardt[2] Jochen Tiepmar[3]

Abstract: For streaming websites, media shopping platforms and movie databases, movie recommendation systems have become an important technology, where mostly hybrid methods of collaborative and content-based filtering on the basis of user ratings and user-generated content have proven to be effective. However, these methods can lead to popularity-biased results that show an underrepresentation of those movies for which only little user-generated data exists. In this paper we will discuss the possibility of generating movie recommendations that are not based on user-generated data or metadata, but solely on the content of the movies themselves, confining ourselves to movie dialog. We extract low-level features from movie subtitles by using methods from Information Retrieval, Natural Language Processing and Stylometry, and examine a possible correlation of these features' similarity with the overall movie similarity. In addition we present a novel web application called *SubRosa* (`http://ch01.informatik.uni-leipzig.de:5001/`), which can be used to interactively compare the results of different feature combinations.

Keywords: Movie Similarity; Subtitles Processing; Information Retrieval; Stylometry; Natural Language Processing

1 Introduction

With a rapidly increasing number of movies produced each year, a growing number of film industries worldwide[4] and new possibilities of distribution via streaming websites, such as *Netflix* and *Amazon Prime*, or on-demand services like *Vimeo* and *Youtube Movies*, recommendation systems for movies have become an essential tool to enhance the user experience: Users who are eager to discover and watch movies unknown to them would be completely lost at the attempt to single-handedly pick the ones they are interested in from the mass of released movies. Currently used movie recommendation systems are largely based on collaborative filtering [BL07; SK09]. A major challenge for collaborative filtering recommendation systems is the so-called cold start problem. A cold start, i.e. the case that a movie does not yet have enough user ratings or that a user did not yet rate enough movies to calculate any recommendations using collaborative filtering, is mostly handled using content-based approaches:
A basis of movie similarities can be determined for initial recommendations using metadata

[1] Leipzig University, Computational Humanities Group, Leipzig, Germany, jan.luhmann@gmx.net
[2] Leipzig University, Computational Humanities Group, Leipzig, Germany, burghardt@informatik.uni-leipzig.de
[3] Leipzig University, Computational Humanities Group, Leipzig, Germany, jtiepmar@informatik.uni-leipzig.de
[4] UNESCO Institute of Statistics. (2016) Record number of films produced. `http://uis.unesco.org/en/news/record-number-films-produced` (date accessed: 2019-08-29)

doi:10.18420/inf2020_119

provided by the movie distributor such as genre tags, list of cast and plot synopsis, as well as tags regarding plot, style and mood of a movie [Sa01].
However, recommendation systems using this approach still suffer from a popularity bias: Any movie that has not been sufficiently tagged by users or whose metadata is only fragmentary is a lot less likely to survive a cold start. To increase diversity and novelty in movie recommendations, it would be greatly beneficial to be able to estimate movie similarities independently of any human-supplied attributional data but based on the content of movies themselves.
In addition of the actual video and audio data of a movie, a third resource which can represent one aspect of a movie, i.e. its dialog, is its subtitles. Of course, subtitles can only contain a fraction of the information which a movie's dialog provides, completely missing information about speakers, intonation, facial expressions etc. But since it still may contain information about dialog topics and manner of speaking, and since English subtitles – even for little known films – are widely available today through online platforms such as *OpenSubtitles*[5] and can be processed inexpensively and efficiently in comparison to video or audio data, we will here discuss and explore the possibility of detecting movie similarities using feature extraction from subtitles. The applied methods of feature extraction are related to Natural Language Processing (NLP), Information Retrieval (IR) and Stylometry.

1.1 Related Work

Before we present the experimental setup, we discuss a number of works that use subtitles and movie scripts as a basis to calculate similarities between movies.
Blackstock; Spitz [BS08] propose a method for classifying 399 movies by NLP-related features extracted from movie scripts. Their method is partly stylometric, examining the ratios and distributions of occurrences of grammatical word forms using Part-of-Speech tagging (POS), partly statistical using features derived from speaker annotations which are present in movie scripts, and partly based on Named Entity Recognition (NER), for the analysis of identical named entities. Movies are classified by genre using Maximum Entropy Markov Models and Naive Bayes. While stylometric features achieve the best results, the overall accuracy is relatively low. The authors conclude that a larger and more diverse dataset would have improved their results. However, freely and digitally available movie scripts are much harder to obtain than subtitle files, and they are also more difficult to process.
Nessel; Cimpa [NC11] propose a movie recommendation engine which uses an Inductive Inference-based method of calculating similarities among subtitle texts of 290 pre-selected movies. The results of the evaluation experiments look promising, although it is difficult to say how their approach would perform on a more diverse dataset.
Bougiatiotis; Giannakopoulos [BG17] examined the correlation of movie similarities and features extracted from subtitles and audio, using a dataset of subtitle files of 160 movies.

[5] http://www.opensubtitles.org

Bag-of-Words (BOW) representations of subtitle text were used for calculating topic models. Segments of audio data were classified by event (music, speech, noises etc.) and in case of music classified by music genre. In evaluations, extracted audio features only yield very low accuracy scores. The two most accurate results are generated by topic modeling using Latent Dirichlet Allocation, and by simple tf-idf weighting of BOW representations.

2 Experimental Setup

2.1 Dataset

Our dataset consists of English subtitles for 5,914 movies. These movies are all among the 10,000 most rated movies on *IMDb*[6]. Despite our motivation to tackle a popularity bias, we chose rather well-known movies to be able to better assess the results of our experiments. We decided to use such a large and diverse dataset because it may improve the quality of some models and more accurately represent a later real-world application.
Subtitles were kindly made available by *OpenSubtitles*. They claim that their database only offers files that can be freely and legally distributed. For each movie they provide us with several versions of subtitles. Frequently, subtitles contained OCR errors (optical character recognition), encoding errors, and for our purposes unwanted data like speaker annotations, music lyrics, authorship tags by the subtitle creator and HTML markup. To minimize the occurrence of such data, subtitles for all movies undergo a heuristic cleaning, validation and selection process, which leaves us with one selected subtitle file for each of 5914 movies, formatted as a SubRip file (*.srt). Additionally, metadata (*IMDb*-ID, title, release year, genres, runtime, number of ratings, rating) for these movies are obtained from *IMDb*.

2.1.1 Preprocessing

For some of our feature extraction methods which are described in Sect. 2.2 it is necessary to convert a movie's subtitles into a continuous text. This is done by simply joining all dialog lines to a single string, separated by whitespace. The text is then further processed using *spaCy*'s language processing pipeline[7]. In our case, this pipeline consists of the following stages: tokenization, POS tagging, dependency parsing, NER, lemmatization. For each movie, we store its sequence of tokens, lowercased and without punctuation, additionally in lemmatized form, and their corresponding POS tags. We filter out interjections, and named entities which refer to persons, organizations or places because these would heavily distort the results of the Bag-of-Words Model discussed in Sect. 2.2.1.

[6] http://www.imdb.com
[7] http://www.spacy.io/

2.2 Methods for Feature Extraction

In this section the applied methods for feature extraction are presented. Their configurations and the adjustments of their parameters were determined experimentally, i.e. by testing them on randomly selected samples of our dataset and examining the results.
There is one method for topical analysis of documents, in this case with the motivation to address the question: *What are the characters of a movie talking about?* (see Sect. 2.2.1). Four methods are aimed at stylistic analysis, in an attempt to address the question: *How are the characters of a movie talking?*, considering aspects of lexicality, syntax and speech rhythm (see Sect. 2.2.2 to 2.2.4 and 2.2.6). A sixth method is aimed at an analysis of emotions: *Which emotions are the characters expressing in their speech?* (see Sect. 2.2.5)

2.2.1 Bag-of-Words Model (BOW)

The Bag-of-Words model is a simple approach for representing text documents in Information Retrieval and Natural Language Processing. In our model, the subtitle text for a movie (a document) is represented by its set of unigrams of lemmatized tokens, weighted by sublinear-tf-idf scaling [MRS08] which is a logarithmic variation of tf-idf scaling. By logarithmizing the term frequency factor, the actual number of occurrences of a term in the document has a less drastic effect on the weighting. To filter out stop words and to reduce dimensions of our document representation, all terms which occur in more than 95% of all documents are ignored. This limit is adjusted to the occurrence of "traditional" stop words. Likewise, all terms which occur in less than 2.5% of all documents are ignored. This limit is adjusted to the occurrence of terms containing spelling mistakes. Finally remaining are 4,952 terms, so that documents are represented by this model as vectors of dimension 4,952.

2.2.2 Distribution of Stop Words (SWD)

One method that has long been successfully used in stylometry, particularly in studies of author attribution, is to measure the rates of occurrences of stop words in a given document and to understand the distribution of these as a fingerprint of the author's writing style. This was mainly developed by John Burrows during the late 1980s [Bu87]. This method is applied on our dataset of unlemmatized tokens. Based on the NLTK stop word list and the most common terms in our corpus, we manually generate a set of 87 stop words. Document representations are then computed in the same way as our BOW-Model but considering only the terms of the stop word set, weighted simply by their term frequency.

2.2.3 POS Tag Trigrams (PTT)

The method of extracting features related to syntactic structures is commonly used for genre classification of text documents. Genres in this case are usually fiction, academic text, news text, conversation, etc., where the syntactic structure is a discriminating factor between genres, regardless of the topic of texts. Santini [Sa04] proposes the use of trigrams of POS tags which "are large enough to encode useful syntactic information, and small enough to be computationally manageable ", so we choose to do the same in this work. The POS tags emitted by *spaCy* are from the "OntoNotes 5 version of the Penn Treebank tag set." [8] Only trigrams within sentence boundaries are considered. Similarly to the method presented in Sect. 2.2.2, we are interested in the distribution of the frequencies of globally frequent, common features among our documents. For this reason we ignore all trigrams that occur in less than 90% of our documents, resulting in 417 trigrams, weighted by their frequency.

2.2.4 Stylometric & Statistical Measurements (SSM)

Next we calculate a model of document representations based on various ratios and measurements which are popular in stylometric studies and which are once again often successfully used for genre classification and authorship attribution. These include Shannon Entropy [Sh48] of term probabilities, standardized type-token ratio [Jo44; TC13], average length and ratios of lengths of sentences and words.

2.2.5 Sentiment Analysis (SEA)

Using sentiment analysis of a document of our dataset we may estimate the emotional arc of the respective movie, assuming that this arc is in some way reflected in the dialog. We use the well evaluated open-source tool *VADER Sentiment*[9] [HG14] for calculating a *compound sentiment score* for a given text, which reflects positive or negative sentiment on a scale from 1 to -1. Since emotions usually evolve very intensely over the course of a movie, it is not useful to calculate the sentiment of an entire document of movie dialog at once. Instead, for each second of a movie we calculate the sentiment of the dialog spoken, resulting in a sentiment curve (Fig. 1). Smoothing by simple moving average (window size 10) is applied. As features of a curve, we calculate its mean value, its first and third quartile, its rate of zero crossings and the rates of zero crossings of the first and second derivative. The latter metrics are derived from signal processing [Ch88] and may be replaced in future work by more robust metrics to measure entropy, volatility and oscillation of time series.
Fig. 1 shows the highly different sentiment curves for the horror movie "Evil Dead" (2013) and the romantic musical "La La Land" (2016).

[8] `https://spacy.io/api/annotation#pos-tagging` (date accessed: 2019-08-29)
[9] `https://github.com/cjhutto/vaderSentiment` (date accessed: 2019-08-29)

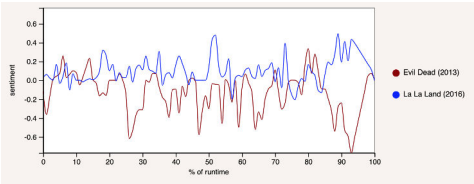


Fig. 1: Detail view of Sentiment Analysis curves in *Sub Rosa*.

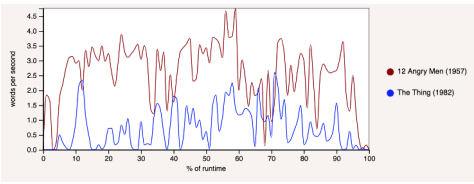


Fig. 2: Detail view of Speech Tempo Analysis curves in *Sub Rosa*.

2.2.6 Speech Tempo Analysis (STA)

It is intuitively understandable that distinctive features of a movie may be the tempos at which characters speak as well as the frequency and duration of speech pauses. These may correlate with the rhythm of a movie's editing which is crucial to its style and atmosphere [Va85]. For each second of a movie, we calculate the speech tempo by approximating how many words are spoken during the second, resulting in a speech tempo curve (Fig. 2). Smoothing by simple moving average (window size 10) is applied to the curve. We extract the same features as from the SEA curves.
Fig. 2 shows the highly different speech tempo curves for the slow-paced horror movie "The Thing" (1982) and the dialog-driven courtroom drama "12 Angry Men" (1957).

2.3 Similarity Measurement

For each of the mentioned models we have generated a unique feature vector. To determine the similarity between two movies' subtitles, we calculate the distances between the six respective vectors (BOW, SWD, PTT, SSM, SEA, STA). For the BOW model, cosine distance is used. For all other models, cosine delta is used as a distance metric. Compound distance scores are the weighted averages of all models' distance values. The weights can be set manually in the web application that is described in the next section.

2.4 Web Application for Interactive Evaluation

So far, we have been using various parameters to determine the similarity of movies based on their subtitles. Our primary research goal in this project and future work is to learn more about the expressiveness and validity of the different features, i.e. which feature or combination of features (and different weights of the features) provides the best similarity measure for movies? And accordingly: Are there differences for movies from different directors, genres or dates? To investigate these questions in an exploratory way, we designed a web application called *Sub Rosa*, which is available via `http://github.com/bbrause/subrosa`. We also provide a live demo of the application that can be found

at http://ch01.informatik.uni-leipzig.de:5001/. *SubRosa* allows users to adjust the weighting of feature models based on which compound distance scores are calculated. Users can request the nearest neighbors of any movie which are visualized in a graph in which each node represents a movie and the length of the edge between each two movies is proportional to the square of the compound distance score calculated between them. In addition, specific movies can be compared in more detail by providing information about the most frequent 200 tokens in the BOW model, POS and stop word distributions, stylometric features, sentiment and speech tempo scores.

3 Results

Beyond the exploratory testing of the *SubRosa* web interface, we can estimate whether our models succeed to determine similarity between movies by analyzing two-dimensional projections of all vectors of the models. These were made by first reducing to 50 dimensions using Truncated SVD[10] and then further reducing to two dimensions using t-SNE for visualization [MH08]. I The color of each data point matches the genre of the movie it represents. Interactive plots for each single model as well as for some weighted combinations of models are available via https://chart-studio.plot.ly/~bbrause/#/

Fig. 3 shows a plot of all six unweighted features. Although some genres, e.g. Comedies and Horror movies, seem to form (partial) clusters, most of the other genres are highly scattered in the feature space. In other words: movie similarity based on dialogs does not manifest itself clearly in the existing genre definitions, i.e. a Comedy with a war theme (e.g. "The Men Who Stare at Goats", 2009) is rather rated similar to other War movies than to other comedies.

Fig. 4 provides an overview of plots that were created for each singular model. It shows that each models produces rather different similarity patterns. In a way the BOW model stands out, as it produces at least some visible clusters of movies. Taking a closer look at those clusters, Fig. 5 underlines the previous observation: although some genres tend to form clusters (e.g. Western), most clusters are formed by either a common cultural theme (e.g. Indian movies) or some other kind of theme (e.g. Religion or War).

An obvious next step would be a more detailed evaluation of the quality of our models, which will be possible if a ground-truth dataset of human-estimated similarities of movies can be found.

4 Conclusions

We presented an experimental setup to determine the similarity of movies based on different features that can be extracted from the subtitles. This is not only important to generate objective recommendations for movie consumers, but can also help to aid computer-based

[10] https://scikit-learn.org/stable/modules/decomposition.html (date accessed: 2019-08-29)

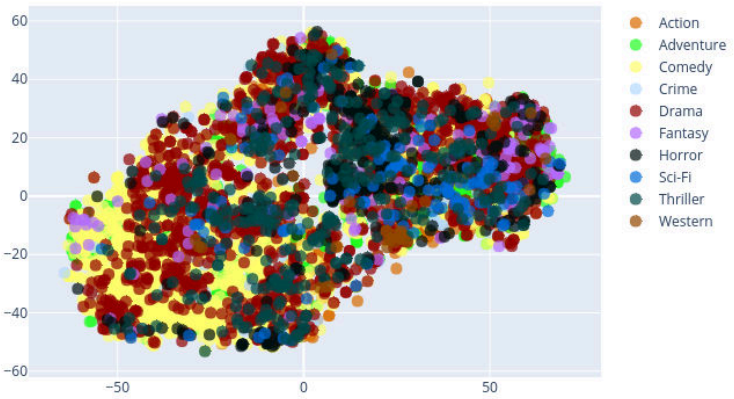


Fig. 3: All feature models concatenated (unweighted), 2D projection using Truncated SVD and t-SNE.

film studies, which are currently a trend in the Digital Humanities[11]. *SubRosa* is intended as an interactive tool that allows users to experiment with with different features and weights, to see how different parameter settings have an effect on the results. In addition, we created some very basic plots that show that obvious similarities (beyond genre classifications) can be identified with our different models. However, the question remains: which features or combination of features yields the best results for the detection of similar movies?
As was indicated in the previous section, we are planning to do a systematic evaluation with all of the parameter configurations as an obvious next step, as we would like to know what is the best setting to detect movie similarity on the dialog level. A possibility for proper evaluation of our experiment may be provided by the "MovieLens" dataset by the research lab *GroupLens* (University of Minnesota, USA)[12]. It offers user-generated tags, mostly related to style, mood, plot or setting, for 58,000 movies. Similarities of movies regarding their tags may be compared with their similarities regarding our models.

5 Acknowledgements

We would like to thank the *OpenSubtitles* team for providing part of their subtitle database, which made this work possible in the first place.

[11] See the SIG AudioVisual material in Digital Humanities, `https://avindhsig.wordpress.com/`.
[12] `http://grouplens.org/datasets/movielens/`
(date accessed: 2019-08-29)

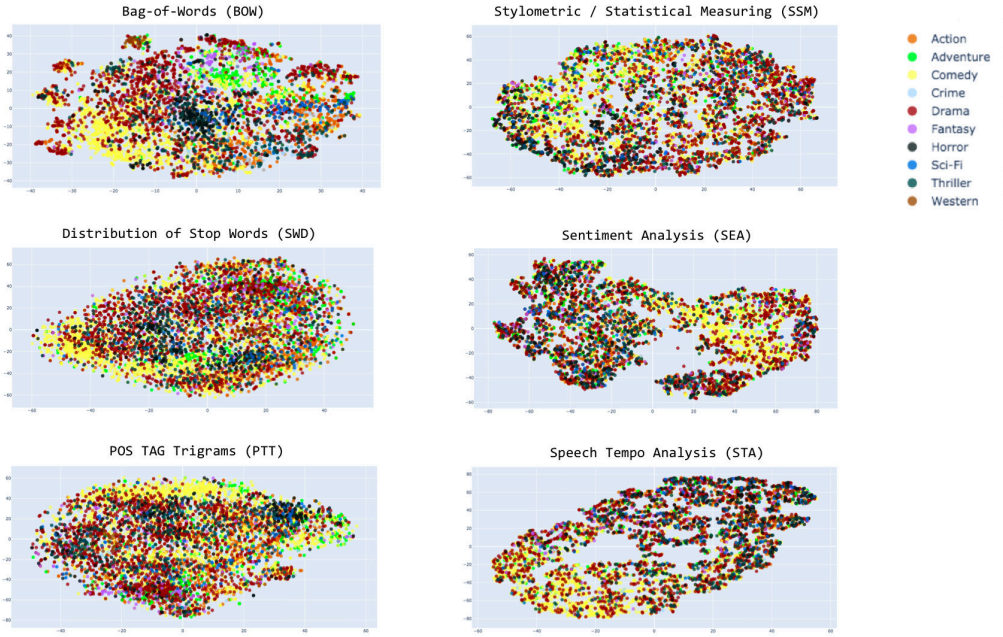


Fig. 4: Plots for the six separate feature models.

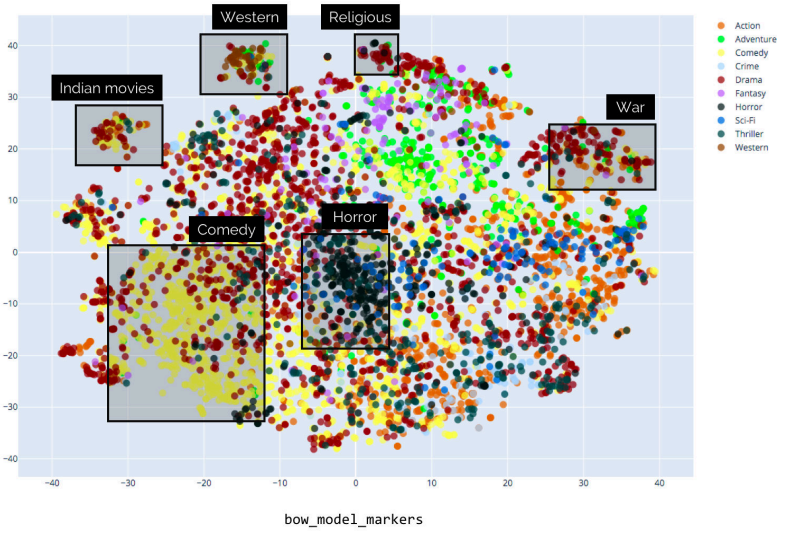


Fig. 5: Annotated BOW plot.

References

[BG17] Bougiatiotis, K.; Giannakopoulos, T.: Multimodal Content Representation and Similarity Ranking of Movies, arXiv preprint arXiv: 1702.04815, 2017.

[BL07] Bennett, J.; Lanning, S.: The Netflix Prize, 2007.

[BS08] Blackstock, A.; Spitz, M.: Classifying Movie Scripts by Genre with a MEMM Using NLP-Based Features, 2008.

[Bu87] Burrows, J.: Computation into Criticism: A Study of Jane Austen's Novels and an Experiment in Method. Clarendon Press and Oxford University Press, 1987.

[Ch88] Chen, C.-H.: Signal processing handbook. CRC Press, 1988.

[HG14] Hutto, C. J.; Gilbert, E.: Vader: A parsimonious rule-based model for sentiment analysis of social media text. In: Eighth international AAAI conference on weblogs and social media. 2014.

[Jo44] Johnson, W.: Studies in language behavior: A program of research. Psychological Monographs 56/2, pp. 1–15, 1944.

[MH08] van der Maaten, L.; Hinton, G.: Visualizing Data using t-SNE, 2008.

[MRS08] Manning, C.; Raghavan, P.; Schütze, H.: Introduction to Information Retrieval. Cambridge University Press, 2008.

[NC11] Nessel, J.; Cimpa, B.: The MovieOracle - Content Based Movie Recommendations, 2011.

[Sa01] Sarwar, B.; Karypis, G.; Konstan, J.; Reidl, J.: Item-based collaborative filtering recommendation algorithms. In: Proceedings of the tenth international conference on World Wide Web - WWW '01. the tenth international conference. ACM Press, Hong Kong, Hong Kong, pp. 285–295, 2001.

[Sa04] Santini, M.: A Shallow Approach To Syntactic Feature Extraction For Genre Classification. In: Proceedings of the 7th Annual Colloquium for the UK Special Interest Group for Computational Linguistics (CLUK 2004). 2004.

[Sh48] Shannon, C. E.: A mathematical theory of communication. Bell system technical journal 27/3, pp. 379–423, 1948.

[SK09] Su, X.; Khoshgoftaar, T. M.: A Survey of Collaborative Filtering Techniques. Advances in Artificial Intelligence 2009/1, Jan. 2009.

[TC13] Torruella, J.; Capsada, R.: Lexical Statistics and Tipological Structures: A Measure of Lexical Richness. Procedia - Social and Behavioral Sciences 95/, pp. 447–454, 2013.

[Va85] Van Leeuwen, T.: Rhythmic Structure of the Film Text. In (Dijk, T. v., ed.): Discourse and Communication: New Approaches to the Analysis of Mass Media Discourse and Communication. Walter de Gruyter, pp. 216–232, 1985.

R. Reussner, A. Koziolek, R. Heinrich (Hrsg.): INFORMATIK 2020,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2021 1281

Konzept und Klassifikation literarischer Raumentitäten[1]

Florian Barth[2]

Abstract: Innerhalb der "Computational Narratology" werden literaturwissenschaftlich-textuelle Konzepte formalisiert, um sie anschließend mit algorithmischen und maschinellen Verfahren automatisch identifizieren zu können. Dieser Prozess, der eine enge Verknüpfung theoretischer Grundlagen und computationeller Umsetzung erfordert, wird in diesem Beitrag am Beispiel narratologischer Raumkategorien vorgestellt. Eine Pilotannotation demonstriert konzeptuelle Grundlagen der Kategorien und deren intersubjektives Verständnis anhand des Inter-Annotator-Agreements. Anschließend werden Features zur Erkennung jener Ortsreferenzen entwickelt, an denen die diegetische Handlung eines literarischen Textes angesiedelt ist, und prototypische Machine-Learning-Modelle zur Klassifikation präsentiert.

Keywords: Computational Literary Studies; Natural Language Processing; Machine Learning; Spatial Analysis; Annotation

1 Einordnung

Unter dem Begriff der "Computational Humanities" lassen sich jene Methoden innerhalb der Digital Humanities fassen, die einen stärkeren Fokus auf die Entwicklung numerischer und formaler Modelle legen, beispielsweise durch die Verwendung von Machine Learning, statistischen Verfahren oder algorithmischer Sprachverarbeitung [Bu20; Ro19]. Als spezifischer Unterbegriff hat sich innerhalb der digitalen Literaturwissenschaft in jüngster Zeit das Schlagwort "Computational Literary Studies" durchgesetzt, worunter man die Anwendung obiger Verfahren für genuin literaturwissenschaftliche Fragestellungen versteht. Bei Erzähltexten wird insbesondere die Formalisierung narratologischer Konzepte forciert, etwa im Rahmen von Shared-Tasks [GRW19], sowie eine anschließende maschinelle Erkennung angestrebt, z. B. von narrativen Szenen [Gi19] oder dem diegetischen Status der Erzählinstanz und ihrer Perspektive (point of view; [EF16]).

In diesem Kontext widmet sich der vorliegende Beitrag der Formalisierung und Klassifikation narratologischer Kategorien des Raums nach Dennerlein [De09] und Piatti [Pi08]. In der bisherigen literaturwissenschaftlichen Forschung wurde die Kategorie des Raums insbesondere vor dem Hintergrund des *Spatial Turn* ausführlich besprochen [So89] und innerhalb der Narratologie konzeptualisiert – dies erfolgte jedoch vor allem theoretisch,

[1] Der Beitrag basiert auf der Masterarbeit des Autors, die von Gabriel Viehhauser, Nils Reiter und Roman Klinger unterstützt wurde, und denen hierfür ein besonderer Dank gewidmet ist.

[2] Universität Göttingen, Institut für Informatik / Niedersächsische Staats- und Universitätsbibliothek Göttingen (SUB), Abteilung Forschung & Entwicklung, florian.barth@uni.goettingen.de

doi:10.18420/inf2020_120

ohne eine Anwendung des Konzepts bei längeren Textpassagen oder einer Prüfung auf Formalisierbarkeit. Deshalb wird in diesem Beitrag eine Pilotannotation vorgestellt, welche die Kategorie der literarischen RAUMENTITÄT (SPATIAL ENTITY) anhand einheitlicher Richtlinien erfasst und mittels Inter-Annotator-Agreement evaluiert.

Anschließend erfolgt die Klassifikation der RAUMENTITÄTEN in handlungsrelevante SETTINGS und lediglich ERWÄHNTE RÄUME (MENTIONS) – eine Differenzierung, die einen zentralen Aspekt der narratologischen Auseinandersetzung mit Raum darstellt [De09; Pi08]. Wir bezeichnen diese Unterscheidung im Folgenden auch als *Klassifikation der Handlungsrelevanz*, für deren Automatisierung linguistisch-textuelle Features und prototypische Machine-Learning-Verfahren präsentiert werden. Eine Schwierigkeit dieser Klassifikation liegt in der Einbettung der Kategorie des SETTINGS in übergeordnete Textstrukturen, zu denen insbesondere *narrative Ebenen* gehören, von denen die Klassifikation und bereits die formale Konzeptualisierung von SETTINGS abhängt. Beim vorliegenden SETTINGS-Begriff wurde daher auf eine definitorische Unabhängigkeit von Erzählebenen geachtet und die Textauswahl für die Annotation derart getroffen, dass lokale Features nicht von einer narrativen Ebene beeinflusst werden (siehe Abschnitt 2.2).

2 Konzept und Pilotannotation

2.1 RAUMENTITÄTEN

Literarische RAUMENTITÄTEN (SPATIAL ENTITIES) umfassen i.) Toponyme (konkrete geographische Raumnamen wie "Frankreich" oder "Berlin"), ii.) fiktionale und faktuale Eigennamen ("Eiffelturm", "Schicksalsberg") sowie iii.) Gattungsbezeichnungen ("Problemviertel", "Speisekammer") [De09].[3] Grundsätzlich erfassen wir RAUMENTITÄTEN als ganze Nominalphrase (NP) und, sofern eine Präposition voransteht, als Präpositionalphrase (PP; Bsp. 1.b, 1.d). NPs bzw. PPs, die sich auf unterschiedliche RAUMENTITÄTEN beziehen, können sich zudem überlappen oder umschließen, und werden dann verschachtelt annotiert (1.a).

1. a Passepartout befand sich also, nachdem halb zwölf vorüber war, allein [im Hause [der Savile Row]$_{\text{SPATIAL ENTITY}}$]$_{\text{SPATIAL ENTITY}}$. Sogleich machte er sich daran, [es]$_{\text{SPATIAL ENTITY}}$ [[vom Keller] [bis zum Speicher]$_{\text{SPATIAL ENTITY}}$]$_{\text{SPATIAL ENTITY}}$ zu besichtigen. (Verne: Reise um die Erde in 80 Tagen)

   b Ein junger Mann saß zusammengesunken [neben der Gaslaterne]$_{\text{SPATIAL ENTITY}}$ [auf dem Bordstein]$_{\text{SPATIAL ENTITY}}$. (Dos Passos: Manhattan Transfer)

[3] Im Natural Language Processing werden lediglich Toponyme von bestehenden Named-Entity-Klassifikatoren als Geographical Entity (GEO) oder Geopolitical Entity (GPE) erkannt. Für die englische Sprache gibt es z. B. den Named-Entity-Classifier von Finkel et al. [FGM05] und fürs Deutsche jenen von Faruqui; Padó [FP10].

c Felicie lief [zu Homais]$_{\text{SPATIAL ENTITY}}$, der es aller Welt ausposaunte.[4] (Flaubert: Madame Bovary)

d Als ich [auf dem Wege]$_{\text{SPATIAL ENTITY}}$ hinunter [zum Mittagessen]$_{\text{SPATIAL ENTITY}}$ [an dem Zimmer]$_{\text{SPATIAL ENTITY}}$ vorüberging, sah ich [durch die geöffnete Thür]$_{\text{SPATIAL ENTITY}}$. (Brontë: Jane Eyre)

Als RAUMENTITÄTEN gelten zudem iv.) Objekte, die von Figuren betretbar sind ("Flugzeuge", "Autos", aber auch "Schränke", "Kisten"),[5] v.) Distanzen zwischen zwei RAUMENTITÄTEN (Bsp. 1.a), vi.) Deiktika (die Adverbien "hier", "da" und "dort") sowie vi.) unspezifische Konkreta ("außen"/"innen", "Heimat", "Zuhause") [De09]. Die zuvor genannten Formen lassen sich allein anhand des Nomens innerhalb einer RAUMENTITÄT bzw. des Adverbs (bei Deiktika und unspezifischen Konkreta) bestimmen. Andere Formen sind abhängig vom Kontextverständnis einer Textpassage, etwa bei vi.) Objekten, gegenüber denen sich Figuren verorten (Bsp. 1.b) [De09], vii.) Figurennennungen, die auf einen Ort referieren (Bsp. 1.c), oder bei viii.) Ereignissen, die innerhalb einer NP bzw. PP ausgedrückt werden und gleichzeitig eine Ortsangabe darstellen ("zum Mittagessen" in Bsp. 1.d). Darüber hinaus wurden in der Pilotannotation Pronomen und Konjunktionen als RAUMENTITÄTEN berücksichtigt, wenn sie mit einer zuvor (Katapher) oder danach genannten RAUMENTITÄT (Anapher) koreferent sind ("es" in Bsp. 1.a) [Re13].

2.2 Klassifikation der Handlungsrelevanz

Bei jeder RAUMENTITÄT soll entschieden werden, ob dort die Handlung der Erzählung aktiv stattfindet, wodurch sich die Entität zu einem SETTING qualifiziert. Ein SETTING zeichnet sich aus durch 1.) die Präsenz und aktive Handlung von Figuren verbunden mit einer *Jetzt-Zeitlichkeit* (Bsp. 2.a) [De09; Pi08], oder durch 2.) eine besondere Ereignishaftigkeit bei Vorgängen ohne direkte Figurenbeteiligung (2.b) [De09].[6]

2. a In diesem Augenblicke klopfte es an die Türe [des kleinen Salons, worin sich Phileas Fogg aufhielt.]$_{\text{SETTING}}$.[7] (Verne: Reise)

b Um halb ein Uhr hielt [der Zug]$_{\text{SETTING}}$ [auf der Station Benares]$_{\text{SETTING}}$. (Verne: Reise)

[4] Dennoch wird "Homais" in diesem Beispiel auch als eigenständige Figuren-Entität betrachtet (neben der Funktion als Raumreferenz).

[5] In der Raumnarratologie wird zwischen punktuellen Orten und weiterreichenden Räumen unterschieden [De09], die in unserem Fall beide innerhalb des Konzepts der RAUMENTITÄT angesiedelt sind.

[6] Bestehende informatische oder computerlinguistische Ansätze wie das Konzept des *ISO-Space* [PMV11] erfassen zwar u. a. literaturwissenschaftlich relevante Aspekte der Raumdarstellung sowie die Ereignishaftigkeit eines Raumes im Zusammenhang einer Bewegung [Pu15]. Jedoch wird kein Bezug zur Narration bzw. dem Plot eines Textes hergestellt, und zudem bestehen grundlegende Diskrepanzen zu den literaturwissenschaftlichen Kategorien, z. B. hinsichtlich der Beurteilung von Figuren als räumliche Entitäten.

[7] Relativsätze werden mitannotiert, wenn sie direkt an die NP bzw. PP angeschlossen sind.

Als ERWÄHNTER RAUM (MENTION) werden dagegen RAUMENTITÄTEN im Kontext von Reflexionen oder Beschreibungen (3.a) sowie repetitiv erzählten (3.b) oder zukünftig bevorstehenden Ereignissen (3.c) annotiert.

3. a [Das nicht eben prachtvolle Haus [in Savile Row]$_{\text{MENTION}}$]$_{\text{MENTION}}$ empfahl sich durch größte Bequemlichkeit. (Verne: Reise)
   b Er machte regelmäßig die Fahrten [[von Brindisi]$_{\text{MENTION}}$ [nach Bombay]$_{\text{MENTION}}$]$_{\text{MENTION}}$ [durch den Suez-Canal]$_{\text{MENTION}}$. (Verne: Reise)
   c In einem Monat wollte sie mit ihrem Bruder [mit dem Auto]$_{\text{MENTION}}$ [nach New York]$_{\text{MENTION}}$ fahren. (Kerouac: Unterwegs)

Wie eingangs dargelegt, beeinflusst die jeweilige narrative Ebene die Klassifikation aller darin befindlichen SETTINGS: Auf einer typischen zweiten Erzählebene – einer Geschichte innerhalb einer Geschichte, die z. B. von einer intradiegetischen Figur erzählt wird [De09; Ge88] – entstehen neue SETTINGS, obwohl diese im genannten Beispiel lediglich von einer Figur der ersten Erzählebene "erwähnt" werden. Zum Konzept einer Erzählebene gehören nach Ryan ebenfalls Passagen mit einem abweichenden ontologischen Status, z. B. Traumsequenzen oder eingeschobene Textartefakte wie Briefe [Ry91], die Dennerlein als modale oder mediale Komponente erfasst, und die darin befindlichen RAUMENTITÄTEN als ERWÄHNTE RÄUME klassifiziert [De09]. Im vorliegenden SETTING-Konzept gehen wir hingegen auch bei ontologisch abweichenden Erzählebenen davon aus, dass dort neue SETTINGS erkannt werden können, sodass eine Unabhängigkeit von der Kategorie der narrativen Ebene gewährleistet ist (siehe dazu: [Ba19]).

2.3 Annotation

Die Pilotannotation wurde von zwei Annotatorinnen im Annotationstool *WebAnno* durchgeführt [Yi13],[8] und basiert auf fünf Texten aus dem *Corpus of German Novels* (*DROC*; [Kr17]), in welchem Annotations-Layer für direkte Rede sowie Figuren und deren Koreferenz vorliegen.[9]

Für das Agreement auf Token-Ebene wurde *Cohens Kappa* verwendet, welches sich für den Vergleich von zwei Annotationen eignet [Co60]. Da die Berechnung von Kappa auf dem beobachteten Agreement $P(A)$ sowie dem erwarteten Agreement $P(E)$ basiert, kann jedes Token nur einmalig in den Vergleich einbezogen werden. Bei überlappenden Annotationen

[8] Eine Annotation wurde vom Autor selbst durchgeführt und eine zweite von einer wissenschaftlichen Hilfskraft des Instituts für Literaturwissenschaft (Abteilung Digital Humanities) der Universität Stuttgart. Die Annotationen, die in WebAnno intern mit Apache UIMA realisiert sind [Sc04], wurden ins Format *XML Metadata Interchange* (*XMI*) exportiert. XMI ist ein OMG-Standard und geeignet Objekt-Graphen wie die *Common Analysis Structure* (*CAS*) von UIMA zu repräsentieren [Ap10]. Im Falle der vorliegenden Annotationen eignet sich XMI als Stand-Off-Format, um komplexe, mehrfach überlappende Annotationen abzubilden (vgl. Tabelle 1).

[9] Konkret handelt es sich um Romanauszüge aus Theodor Fontanes *Cécile* sowie *Quitt*, Gustave Flauberts *Madame Bovary*, Lew Tolstois *Anna Karenina* und Émile Zolas *Nana*. Alle Textausschnitte beinhalten lediglich eine narrative Ebene.

	In	einer	Viertelstunde	mußte	der	von	Galveston	nach	dem	Norden	führende	Zug	da sein.
1					M	M	M	M	M	M	M	M	
2						M	M	M	M	M			
3						M	M						
4								M	M	M			

Tab. 1: Mehrfach überlappende Annotationen von ERWÄHNTEN RÄUMEN (MENTIONS)

besteht dadurch die Schwierigkeit, dass für ein Token (z. B. "Galveston" in Tabelle 1) mehrere Annotationen einer Annotatorin bestehen, für Kappa pro Token allerdings nur eine Annotations-Ebene berücksichtigt werden kann. Hier wurde daher der Ansatz gewählt, Überlappungen zu verodern: Aus den maximalen Spans überlappender Annotationen wurde bei jeder Annotatorin zufällig eine Annotation mitsamt aller Tokens herausgegriffen. Insbesondere bei mehrfachen Überlappungen gehen dadurch jedoch Informationen verloren.

Der F_1-Score funktioniert hingegen auf Entitäten-Ebene: Übereinstimmende Entitäten werden als True Positives (TP) erfasst; eine fehlende Übereinstimmung zwischen den Annotatorinnen erzeugt jeweils ein False Positive (FP) und ein False Negative (FN). Wir verwenden einen *Exact-F-Score*, bei dem nur die exakte Übereinstimmung aller Tokens einer RAUMENTITÄT als TP gewertet wird, und einen *Fuzzy-F-Score*, bei dem die Übereinstimmung von mindestens einem Token für ein TP ausreichend ist. Insgesamt kann festgehalten werden, dass eine Entitäten-basierte Agreement-Berechnung mittels F_1-Score bei überlappenden und aus mehreren Token bestehenden RAUMENTITÄTEN ein präziseres Agreement ermöglicht, da eine konkrete Zuordnung der von den Annotatorinnen gemeinten Entitäten vorgenommen und bewertet wird.[10]

	F-Fuzzy	**F-Exact**	**Cohens Kappa**
Erkennung der RAUMENTITÄTEN			
F-Macro / Kappa (Mittelwert aus Einzeltexten)	0.86	0.79	0.79
F-Micro / Kappa (alle Tokens)	0.86	0.72	0.78
Erkennung der RAUMENTITÄTEN & SETTING/MENTION-Klassifikation			
F-Macro / Kappa (Mittelwert aus Einzeltexten)	0.73	0.64	0.72
F-Micro / Kappa (alle Tokens)	0.72	0.61	0.71

Tab. 2: Agreement für RAUMENTITÄTEN sowie RAUMENTITÄTEN inkl. Klassifikation der Handlungsrelevanz

Das Agreement wird zunächst für die Erkennung der RAUMENTITÄTEN berechnet sowie in einem zweiten Schritt für die RAUMENTITÄTEN inklusive der Klassifikation der Handlungsrelevanz. Bei letzterer Variante wird ein TP bzw. eine Übereinstimmung bei Kappa nur akzeptiert, wenn auch die Klassifikation der Handlungsrelevanz für eine RAUMENTITÄT zwischen beiden Annotatorinnen übereinstimmt. Tabelle 2 zeigt die Evaluations-Ergebnisse, bei welchen die Werte als Macro- und Micro-F-Score sowie in zwei Varianten für Kappa angegeben sind. Analog zum Macro-F-Score, bei dem für alle Einzeltexte zunächst

[10] Beim Alignment komplexer überlappender Annotationen wurden zuerst die Entitäten mit dem jeweils niedrigsten Start-Index und der längsten Token-Folge verglichen.

Precision und Recall berechnet und aus den gemittelten Werten (Macro-Precision sowie Macro-Recall) der F-Score gebildet wird [SL09], berechnen wir ein gemitteltes Kappa aus den Kappa-Werten der Einzeltexte. Damit finden unterschiedlich lange Einzeltexte die gleiche Berücksichtigung im Endergebnis. Beim Micro-F-Score, ermittelt aus Precision und Recall über alle TP, FP und FN der Einzeltexte, haben längere Texte einen stärkeren Einfluss auf das Ergebnis – ergänzend dazu wurde ein Kappa für alle Token-Vergleiche in allen Texten ermittelt. Das Agreement beim Macro-F-Score weist mit 0.79 (Exakt) bzw. 0.86 (Fuzzy) eine hohe Übereinstimmung auf, wie auch die gemittelten Kappa-Werte der Einzeltexte mit 0.79, während der Micro-F-Score und der Kappa-Wert basierend auf allen Tokens etwas darunter liegen.[11] Insbesondere der Fuzzy-F-Score und die Kappa-Werte belegen somit ein konstant hohes Agreement für die RAUMENTITÄTEN. Wird zusätzlich die Handlungsrelevanz in das Agreement miteinbezogen, sinken die Werte im Mittel über Micro- und Macro-F-Score um 0.13 und bei Kappa um 0.7. Somit herrscht auch bei der Klassifikation der Handlungsrelevanz ein hohes Agreement.

3 Feature-Engineering für die Klassifikation der Handlungsrelevanz

Als Features für die Klassifikation der Handlungsrelevanz dienen linguistisch-textuelle Marker, die Indikatoren für SETTINGS bzw. ERWÄHNTE RÄUME darstellen, und deren Extraktion sowie Funktionalität im Folgenden dargelegt wird. Wir stellen zunächst Features in gestufter Satzumgebung vor (Verbkategorien, temporale Marker, Zeitformen), bei denen der inhaltliche Kontext von einem oder mehreren Teilsätzen Berücksichtigung findet. Dem schließen sich weitere einstufige Features an, die sich auf die jeweilige RAUMENTITÄT bzw. den kompletten umgebenden Satz beziehen (Direkte Rede, Figurenreferenzen, Präpositionen).

3.1 Features in gestufter Satzumgebung

Zur Funktionalität einiger Features ist eine Eingrenzung auf den lokalen Teilsatz notwendig, der die RAUMENTITÄT umgibt. Beispielsweise können innerhalb eines Satzes verschiedene Klassifikationen getroffen werden (Bsp. 4.), die durch derartige Features anhand der lokalen Umgebung unterscheidbar sind.

4. Endlich hielt er es nicht mehr aus, und da er vermutete, sie sei [nach Rouen]$_{\text{MENTION}}$ gefahren, ging er ihr [auf der Landstraße]$_{\text{SETTING}}$ eine halbe Wegstunde weit entgegen. (Flaubert: Madame Bovary)

Zur Verarbeitung dieser Fälle werden die oben genannten Features (Verbkategorien, temporale Marker, Zeitformen) basierend auf dem Dependenzbaum nach Satzumgebungen

[11] Die niedrigeren Werte für den Micro-F-Score und das Kappa für alle Tokens resultieren aus der stärkeren Gewichtung von Fontanes *Quitt*, welches aufgrund zahlreicher überlappender Entitäten das geringste Agreement aller Einzeltexte aufweist, jedoch die umfangreichste Textprobe darstellt.

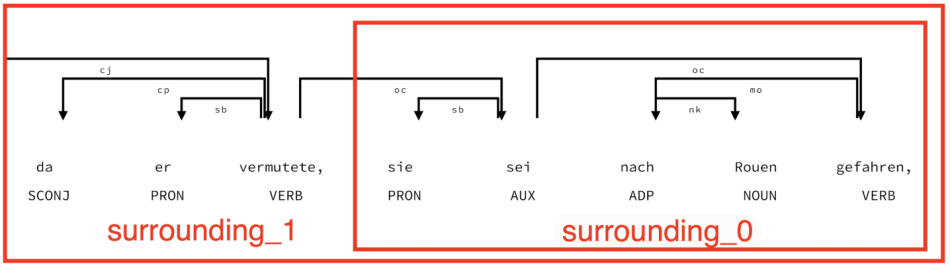


Abb. 1: Satzumgebungen “surrounding_0” und “surrounding_1” für Beispiel 4.

gestaffelt. Für die RAUMENTITÄT “nach Rouen” aus Bsp. 4. wird der unmittelbare Teilsatz extrahiert (“surrounding_0” in Abbildung 1) sowie die im Dependenzbaum nächsthöhere Phrase (“surrounding_1”).[12] Die Klassifikation des ERWÄHNTEN RAUMES ist in diesem Fall maßgeblich durch die Verbkategorien bestimmt: Der dem ERWÄHNTEN RAUM übergeordnete Teilsatz drückt eine Vermutung aus (Verb “vermuten” in surrounding_1, Abbildung 1), an die sich ein Konjunktiv (“sei gefahren”) in der lokalen Satzumgebung der RAUMENTITÄT (“surrounding_0”) anschließt. Wir sprechen von “Umgebungen”, weil bei der im Dependenzbaum nächsthöheren Phrase auch alle untergeordneten Teilsätze abgesehen von den bereits eruierten Umgebungen mit einbezogen werden. So besteht die zweite äußere Umgebung (“surrounding_2”) für den ERWÄHNTEN RAUM in Beispiel 4. nicht nur aus dem direkt übergeordneten Satz zu “surrounding_1” (mit dem Kopf der Phrase “hielt”), sondern auch aus dem angeschlossene Hauptsatz (“surrounding_2”: “Endlich hielt er es nicht mehr aus, und [...] ging er ihr auf der Landstraße eine halbe Wegstunde weit entgegen.”; vgl. auch Tabelle 3). Dies dient der Informationsgewinnung für eine optimale Klassifikation der entsprechenden Entität. Tabelle 3 zeigt für beide RAUMENTITÄTEN aus Beispiel 4. die jeweiligen Umgebungen sowie die extrahierten Verben. Während der ERWÄHNTE RAUM durch eine Vermutung gekennzeichnet ist, steht das SETTING (“auf der Landstraße”) im Kontext einer aktiven Handlung verknüpft mit dem Verb “gehen”.[13]

Zur Vereinheitlichung der Verben dient eine Kategorisierung anhand der *Levin-Klassen* [Le95], die im lexikalisch-semantischen Netz *GermaNet* Verwendung finden und dort ausgelesen wurden [Ge18]. Jedem Verb wird eine von 15 Klassen zugeordnet, z. B. ist für “vermuten” (Bsp. 4.) die Klasse “Kognition” definiert, was als Indikator für einen ERWÄHNTEN RAUM gelten kann.[14]

Temporale Marker können ein eindeutiges Indiz für die Jetzt-Zeitlichkeit bzw. den vergangenen, zukünftigen oder repetitiven Charakter von Ereignissen im Kontext der Raumdarstellung liefern und werden ebenfalls in gestufter Satzumgebung betrachtet. Wir verwenden den

[12] Das Parsing und die Navigation im Dependenz-Baum wurde mit der Python-Library *spaCy* realisiert [HM17].
[13] Bereits Dennerlein [De09] hebt mit Verweis auf Herman [He04] die besondere Bedeutung von Verben bei der Bestimmung von Aktionen bzw. Ereignissen hervor.
[14] Einigen Verben werden in GermaNet mehreren Klassen zugeordnet – in diesem Fall wurde die Mehrheitsklasse als Kategorie gewählt bzw. bei mehreren Klassen eine Kategorie aus den höchstgenannten Klassen gebildet.

Raumentität	surrounding_0	verb_0	surrounding_1	verb_1	surrounding_2	verb_2
nach Rouen (mention)	[sie, sei, nach, Rouen, gefahren]	sein	[da, er, vermutete, ,]	vermuten	[Endlich, hielt, er, es, nicht, mehr, aus, „ und, „ ging, er, ihr, auf, der, Landstraße, eine, halbe, Wegstunde, weit, entgegen, .]	halten
auf der Landstraße (setting)	[„ ging, er, ihr, auf, der, Landstraße, eine, halbe, Wegstunde, weit, entgegen]	gehen	[Endlich, hielt, er, es, nicht, mehr, aus, „ und, da, er, vermutete, „ sie, sei, nach, Rouen, gefahren, .]	halten		

Tab. 3: Gestufte Satzumgebungen und extrahierte Verben für Beispiel 4.

Temponym-Tagger *HeidelTime* [Ku16], der vier verschiedene Typen von Zeitmarkern identifiziert: DURATION (“in einem Monat”, Bsp. 3.c auf S. 4); TIME (“um halb ein Uhr”, Bsp. 2.b); DATE (“September 1828”); SET (“regelmäßig”, Bsp. 3.b).

Das Feature der Zeitform liefert im Idealfall direkte Informationen darüber, ob auf gegenwärtige, vergangene oder zukünftige Ereignisse referiert wird. Die beiden rot dargestellen Teilsätze in Bsp. 5. beinhalten Futur I und referieren 1.) auf die Abfahrt eines Dampfbootes am Datum des 25. und 2.) auf dessen später in der Zukunft liegende Ankunft in Kalkutta. Alle RAUMENTITÄTEN innerhalb dieser Satzteile können anhand der Zeitform klar als ERWÄHNTE RÄUME klassifiziert werden. Nur der grün dargestellte Teilsatz im Präsens beschreibt die Jetzt-Zeitlichkeit. Bei der Integration der Zeitform als Feature greifen wir auf eine Tense-Cluster-Detection von Bögel et al. [BSG14] zurück.

5. » ... Am 25. zu Mittag wird [ein Dampfboot]$_{\text{MENTION}}$ [[von Kalkutta]$_{\text{MENTION}}$ [nach Hongkong]$_{\text{MENTION}}$]$_{\text{MENTION}}$; abgehen. Jetzt haben wir erst den 22., und werden noch zeitig [zu Kalkutta]$_{\text{MENTION}}$ eintreffen.« (Verne: Reise)

3.2 Einstufige Features

Beispiel 5. stellt einen Idealfall dar, weil darin direkte Rede vorliegt – Erzähltexte sind in der Regel jedoch im epischen Präteritum verfasst (Bsp. 1.a–1.d und 2.a–2.b). Daher wird für jede RAUMENTITÄT geprüft, ob sich diese in einer im DROC-Korpus annotierten direkten Rede befindet. Wir verwenden das DROC-Korpus als Goldstandard, anstatt auf automatische Tagger zurückzugreifen (z. B. von Tu et al. [TKB19]), um die Fehlerquellen für die Klassifikation der Handlungsrelevanz zu reduzieren.

Die Präsenz einer literarischen Figur ist ein zentrales Kriterium der SETTING-Definition [De09; Pi08]. Als Feature nutzen wir hierfür die numerische Anzahl von Figurenreferenzen pro Satz aus den annotierten Daten des DROC-Korpus. Darin werden die Figurenreferenzen in “Core” (alle konkreten Figurennamen) und “Non-Core” (auf Figurennamen referierende Personal- und Reflexivpronomen) unterschieden. Für das Feature wurde eine Kombination

aus beiden Gruppen generiert, weil damit eine größere Menge an Figurenreferenzen als bei üblichen Named-Entity-Klassifikatoren einbezogen wird.

Anhand der Präposition ist in einigen Fällen die Klassifikation der Handlungsrelevanz ableitbar, z. B. liefert die Präposition in einem Satz wie "Er ist [in Stuttgart]$_{\text{SETTING}}$" ein starkes Indiz für ein SETTING, ohne dass dafür eine Dependenzrelation zum Subjekt hergestellt werden muss. Eine Modifikation des Verbs sowie der Präposition verändert die Klassifikation dagegen hin zu einem ERWÄHNTEN RAUM ("Er geht [nach Stuttgart]$_{\text{MENTION}}$"). Für das Präpostions-Feature berücksichtigen wir deshalb jede individuelle Präposition als kategoriale Ausprägung, die wir direkt aus den annotierten Präpositionalphrasen extrahieren.

4 Machine-Learning-Modelle zur Klassifikation von SETTINGS und ERWÄHNTEN RÄUMEN

Die Pilotannotation umfasst 510 annotierte RAUMENTITÄTEN (263 SETTINGS und 247 MENTIONS), von denen wir eine *Most Frequent Class Baseline* (Mehrheits-Baseline) mit einer Accuracy von 0.52 ableiten, gegenüber der die folgenden mit der Python-Library *scikit-learn* realisierten Machine-Learning-Modelle verglichen werden [Pe11]. Tabelle 4 zeigt die gemittelten Evaluationsmaße bei einer 10-fold Cross-Validation[15] und zwei genutzten Feature-Sets. Bei allen Features außer der Figurenanzahl pro Satz handelt es sich um kategoriale Variablen, die wir als One-Hot-Vektoren kodieren, um die optimale Funktionalität der Machine-Learning-Algorithmen zu gewährleisten. Feature-Set 1 verwendet alle Merkmale in maximaler Ausprägung (alle Satzumgebungen für Verbkategorien, temporale Marker und Zeitformen; 48 verschiedene Präpositionen; Figurenreferenzen inkl. Personal- und Reflexivpronomen in Core- und Non-Core-Ausprägung; direkte Rede) und Set 2 stellt eine reduzierte Variante mit eingeschränkten Satzumgebungen und Präpositionen dar.[16]

Die höchsten F_1-Score-Werte werden mit einem Random Forest mit unbegrenzter Tiefe und 10 übereinanderliegenden Bäumen erreicht. Bei Entscheidungsbäumen besteht aufgrund der begrenzten Trainingsdaten jedoch das Risiko eines Overfittings, wofür die zusätzlich angegebene Accuracy auf den Trainingsdaten ein Indiz sein kann (0.95 bei unbegrenzter Tiefe im Random Forest bzw. 0.96 beim Decision Tree mit unbegrenzter Tiefe). Um ein Overfitting vorzubeugen, begrenzen wir die Tiefe des Random Forest bzw. der Decision Trees auf auf 4 bzw. 8, woraufhin sich die Accuracy-Werte der Test- und Trainingsdaten deutlich annähern.[17] Entscheidungsbäume bringen insgesamt den Vorteil mit sich, die Klassifikation

[15] Zusätzlich wurden vor Bildung der Folds alle Instanzen randomisiert durchmischt, wodurch einseitige Kategorie-Verteilungen ausgeglichen werden.

[16] Für die Verbkategorien und Zeitformen werden zwei Umgebungen (surrounding 0+1) und für temporale Marker die direkte Umgebung (surrounding_0) verwendet. Statt der vollen 48 Präpositionen werden nur die Vektoren für "nach" und "zu" genutzt. Während Feature-Set 1 nach Konvertierung aller kategorialen Features in One-Hot-Vektoren insgesamt 122 Einzel-Features enthält, sind es bei Feature-Set 2 nur noch 48 Feature-Vektoren.

[17] Für die Entscheidungsbäume werden nur Ergebnisse aus dem reduzierten Feature-Set 2 angegeben, weil bei vollem Feature-Umfang die Diskrepanz zwischen Test- und Trainings-Accuracy noch größer ist.

	Feature-Set	Modell-spezifische Parameter	Accuracy (Acc.-Train)	F_1-micro	F_1-macro	F_1-weighted-macro
Decision Tree	**2**	Tiefe: unbegrenzt	0.76 (0.96)	0.76	0.76	0.76
		Tiefe: 4	0.74 (0.78)	0.74	0.73	0.73
Random Forest	**2**	**Tiefe: unbegrenzt; Bäume: 10**	**0.78 (0.95)**	**0.78**	**0.78**	**0.78**
		Tiefe: 8; Bäume: 10	**0.77 (0.84)**	**0.77**	**0.76**	**0.76**
Naïve Bayes (Multinominal)	**1**	-	0.74 (0.79)	0.74	0.73	0.74
Naïve Bayes (Bernoulli)	**1x**	-	0.75 (0.8)	0.75	0.75	0.75
	2x	-	**0.76 (0.77)**	**0.76**	**0.75**	**0.76**
KNN	**1**	neighbors: 5	0.73 (0.8)	0.73	0.73	0.73
	2	neighbors: 10	0.7 (0.75)	0.7	0.7	0.7
Support Vector Machine	**1**	C=0.1	0.74 (0.81)	0.74	0.73	0.73
	2	C=0.1	0.7 (0.77)	0.7	0.69	0.7
Logistic Regression	**1**	**C=0.1**	**0.75 (0.81)**	**0.75**	**0.74**	**0.75**
	2	**C=0.1**	**0.75 (0.77)**	**0.75**	**0.74**	**0.74**

Tab. 4: Machine-Learning-Modelle im Vergleich (gemittelte Evaluationsergebnisse bei einer 10-fold Cross-Validation; Feature-Set 1x und 2x ohne Figurenreferenzen)

direkt anhand der einzelnen Features nachvollziehen zu können:[18] Beispielsweise wird der erste, auf dem Gini-Gain basierende Split der Daten anhand des Features der direkten Rede durchgeführt, da in Passagen ohne direkte Rede deutlich mehr ERWÄHNTE RÄUME auftreten.

Lineare Klassifikations-Algorithmen wie die Support Vector Machine oder die logistische Regression liefern ebenfalls solide Ergebnisse, bei denen durch eine stärkere Regularisierung (C=0.1) die Gefahr eines Overfittings zusätzlich reduziert wurde.[19] Die logistische Regression sowie auch das Bernoulli-Modell des Naïve Bayes erzielen im reduzierten Feature-Set 2 Ergebnisse, die fast an die Werte des Random Forest heranreichen,[20] wobei sich die Modelle weniger stark an die Trainingsdaten anpassen. Für eine genauere Beurteilung der Modellqualität wäre jedoch ein größeres Test-Datenset wünschenswert. Die F_1-Score-Werte (Micro, Macro, Weighted-Macro) bewegen sich bei allen Modellen im jeweiligen Optimum zwischen 0.73 und 0.78, womit die Mehrheits-Baseline deutlich geschlagen wird. Beim F_1-Weighted-Macro wird zusätzlich die Verteilung der SETTINGS/MENTIONS einbezogen, sodass kürzere Texte weniger Gewicht erhalten, und der F_1-Weighted-Macro oft mit dem F_1-Micro übereinstimmt. Insgesamt demonstrieren die gezeigten Machine-Learning-Verfahren

[18] Bei den Entscheidungsbäumen kommt eine optimierte Version des CART-Algorithmus (Classification and Regression Trees) zum Einsatz, bei der One-Hot-Vektoren vorausgesetzt sind. Beim Random Forest wird für das Sampling der Daten die Bagging-Methode (bootstrap aggregating) verwendet.

[19] Für KNN, Support Vector Machine und die logistische Regression wurden die Features auf die Einheit der Standardabweichung vom Mittelwert normiert.

[20] Für die Verwendung des Bernoulli-Modells des Naïve Bayes, welches binäre Daten erfordert, wurde bei den Feature-Sets 1x und 2x das numerische Feature der Figurenreferenzen entfernt, sodass nur noch One-Hot-Vektoren verarbeitet werden.

die Funktionalität der automatischen Klassifikation der Handlungsrelevanz sowie der dafür verwendeten Features.

5 Fazit

Anhand der vorliegenden Evaluation der Annotation von RAUMENTITÄTEN sowie ihrer Klassifikation wurde die Formalisierbarkeit der raumnarratologischen Kategorien demonstriert. Dabei liefert das Agreement der RAUMENTITÄTEN die Basis zur Entwicklung eines Entity-Classifiers, der den spezifischen Ansprüchen literaturwissenschaftlicher Raumkonfiguration gerecht werden kann. In einer zweiten Annotationsrunde sollte das Agreement insbesondere bei Objekten, gegenüber denen sich Figuren verorten (vi. auf S. 3), oder bei RAUMENTITÄTEN, die ein konkretes Ereignis auf Phrasenebene beinhalten (viii.), optimiert werden.

Im Hinblick auf die Unterscheidung von SETTINGS und ERWÄHNTEN RÄUMEN zeigt das Feature-Engineering, wie eine inhaltlich komplexe Klassifikations-Aufgabe anhand lokaler (teil-)satzbasierter Merkmale erfolgreich automatisierbar ist, wenn genügend Wissen über die entsprechende Domäne eingebracht wird. Bei einer Weiterentwicklung des Klassifikations-Modells ist zunächst eine umfassendere Annotation zur Gewinnung hinreichender Trainingsdaten erstrebenswert. Für ein unabhängiges Modell sollten zudem jene Features, die bisher auf Gold-Annotationen des DROC-Korpus beruhen (direkte Rede, Named-Entities), durch bestehende oder eigene Classifier ersetzt werden. Ferner gilt es neben syntaktischen und satzübergreifenden Merkmalen (z. B. Koreferenzketten) insbesondere die Abhängigkeit und Wechselwirkung mit übergeordneten textuell-narratologischen Kategorien (Erzählebenen, narrative Szenen) in die weitere Konzeption des Feature-Designs einzubinden. Angesichts der komplexen, auf fachwissenschaftlichen Konzepten basierenden Features wäre dann auch zu prüfen, wie ein derartig trainierter Classifier gegenüber Deep-Learning-Modellen abschneidet, die weitgehend ohne spezifisches Feature-Engineering auskommen.

Literatur

[Ap10] Apache: Development Community,"UIMA Tutorial and Developers' Guides", The Apache Software Foundation, 2010, Stand: 23. 06. 2019.

[Ba19] Barth, F.: Annotation Guideline No. 5: Annotation Guidelines for Narrative Levels and Narrative Acts. Journal of Cultural Analytics/, S. 11201, 2019.

[BSG14] Bögel, T.; Strötgen, J.; Gertz, M.: Computational Narratology: Extracting Tense Clusters from Narrative Texts. In: LREC. Bd. 14, S. 950–955, 2014.

[Bu20] Burghardt, M.: Theorie und Digital Humanities – Eine Bestandsaufnahme. In: Digital Humanities Theorie. 2020.

[Co60] Cohen, J.: A coefficient of agreement for nominal scales. Educational and psychological measurement 20/1, S. 37–46, 1960.

[De09] Dennerlein, K.: Narratologie des Raumes. De Gruyter, 2009.

[EF16] Eisenberg, J.; Finlayson, M.: Automatic Identification of Narrative Diegesis and Point of View. In: Proceedings of the 2nd Workshop on Computing News Storylines (CNS 2016). S. 36–46, 2016.

[FGM05] Finkel, J. R.; Grenager, T.; Manning, C.: Incorporating non-local information into information extraction systems by gibbs sampling. In: Proceedings of the 43rd annual meeting on association for computational linguistics. Association for Computational Linguistics, S. 363–370, 2005.

[FP10] Faruqui, M.; Padó, S.: Training and Evaluating a German Named Entity Recognizer with Semantic Generalization. In: KONVENS. S. 129–133, 2010.

[Ge18] GermaNet: GermaNet – An Introduction, 2018, URL: http://www.sfs.uni-tuebingen.de/GermaNet/, Stand: 20. 03. 2018.

[Ge88] Genette, G.: Narrative Discourse Revisited, translated by Jane E. Lewin, 1988.

[Gi19] Gius, E.; Jannidis, F.; Krug, M.; Zehe, A.; Hotho, A.; Puppe, F.; Krebs, J.; Reiter, N.; Wiedmer, N.; Konle, L.: Detection of Scenes in Fiction. In: DH 2019 Book of Abstracts. 2019.

[GRW19] Gius, E.; Reiter, N.; Willand, M., Hrsg.: Cultural Analytics. A Shared Task for the Digital Humanities, Nov. 2019.

[He04] Herman, D.: Story logic: Problems and possibilities of narrative. U of Nebraska Press, 2004.

[HM17] Honnibal, M.; Montani, I.: spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing./, 2017.

[Kr17] Krug, M.: DROC-Release, 2017, URL: https://gitlab2.informatik.uni-wuerzburg.de/kallimachos/DROC-Release, Stand: 26. 02. 2018.

[Ku16] Kuzey, E.; Strötgen, J.; Setty, V.; Weikum, G.: Temponym tagging: Temporal scopes for textual phrases. In: Proceedings of the 25th International Conference Companion on World Wide Web. S. 841–842, 2016.

[Le95] Levin, B.: English verb classes and alternations. A preliminary Investigation 1/, 1995.

[Pe11] Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V. et al.: Scikit-learn: Machine learning in Python. the Journal of machine Learning research 12/, S. 2825–2830, 2011.

[Pi08] Piatti, B.: Die Geographie der Literatur: Schauplätze, Handlungsräume, Raumphantasien. Wallstein, Göttingen, 2008.

[PMV11] Pustejovsky, J.; Moszkowicz, J. L.; Verhagen, M.: Using ISO-Space for Annotating Spatial Information. In: COSIT 2011: 10th International Conference on Spatial Information Theory. 2011.

[Pu15] Pustejovsky, J.; Kordjamshidi, P.; Moens, M.-F.; Levine, A.; Dworman, S.; Yocum, Z.: SemEval-2015 Task 8: SpaceEval. In: Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015). Association for Computational Linguistics, S. 884–894, 2015.

[Re13] Reznicek, M.: Linguistische Annotation von Nichtstandardvarietäten – Guidelines und "Best Practices". Annotation Koreferenz. 2013.

[Ro19] Roth, C.: Digital, digitized, and numerical humanities. Digital Scholarship in the Humanities 34/3, S. 616–632, 2019.

[Ry91] Ryan, M.-L.: Possible worlds, artificial intelligence, and narrative theory. Indiana University Press, 1991.

[Sc04] Schor, M.: An Effective, Java-Friendly Interface for the Unstructured Management Architecture (UIMA) Common Analysis System, Techn. Ber. IBM RC23176, IBM T. J. Watson Research Center, 2004.

[SL09] Sokolova, M.; Lapalme, G.: A systematic analysis of performance measures for classification tasks. Information processing & management 45/4, S. 427–437, 2009.

[So89] Soja, E. W.: Postmodern geographies: The reassertion of space in critical social theory. Verso, 1989.

[TKB19] Tu, N. D. T.; Krug, M.; Brunner, A.: Automatic recognition of direct speech without quotation marks. A rule-based approach. In (Sahle, P., Hrsg.). Digital Humanities: multimedial & multimodal., Konferenzabstracts, Frankfurt am Main, S. 87–89, 2019.

[Yi13] Yimam, S. M.; Gurevych, I.; de Castilho, R. E.; Biemann, C.: WebAnno: A flexible, web-based and visually supported system for distributed annotations. In: Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics: System Demonstrations. S. 1–6, 2013.

R. Reussner, A. Koziolek, R. Heinrich (Hrsg.): INFORMATIK 2020,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2021 1295

Understanding Perceptual Bias in Machine Vision Systems

Visual Analytics as a (Digital) Humanities Challenge

Fabian Offert[1], Peter Bell[2]

Abstract: Machine vision systems based on deep convolutional neural networks are increasingly utilized in digital humanities projects, particularly in the context of art-historical and audiovisual data. As research has shown, such systems are highly susceptible to bias. We propose that this is not only due to their reliance on biased datasets but also because their perceptual topology, their specific way of representing the visual world, gives rise to a new class of bias that we call perceptual bias. Perceptual bias, we argue, affects almost all currently available "off-the-shelf" machine vision systems, and is thus especially relevant for digital humanities applications, which often rely on such systems for hypothesis building. We evaluate the nature and scope of perceptual bias by means of a close reading of a visual analytics technique called "feature visualization" and propose to understand the development of critical visual analytics techniques as an important (digital) humanities challenge, situated at the interface of computer science and visual studies.

Keywords: machine learning; visual analytics; computer vision; bias; interpretability; digital art history

1 Introduction

The susceptibility of machine learning systems to bias has recently become a prominent field of study in many disciplines, most visibly at the intersection of computer science [Fr18] and science and technology studies [Se19], but also in disciplines such as African American studies [Be19]. As part of this development, machine vision has moved into the spotlight of critique as well, particularly where it is used for socially charged applications like facial recognition [BG18; Ga16]. In many critical investigations of machine vision, however, the focus lies almost exclusively on dataset bias [CP19], and on fixing datasets by introducing more, or more diverse sets of images [Me19]. In the following, we argue that this focus on dataset bias in critical investigations of machine vision paints an incomplete picture, metaphorically and literally. In the worst case, it increases trust in quick technological fixes that fix (almost) nothing, while systemic failures continue to reproduce.

We propose that machine vision systems are often inherently biased not only because they rely on biased datasets (which they do) but also because their perceptual topology, their

[1] University of California, Santa Barbara / Friedrich-Alexander-Universität Erlangen-Nürnberg, offert@ucsb.edu
[2] Friedrich-Alexander-Universität Erlangen-Nürnberg, peter.bell@fau.de

doi:10.18420/inf2020_121

specific way of representing the visual world, gives rise to a new class of bias that we call perceptual bias. Concretely, we define perceptual topology as the set of those inductive biases in machine vision systems that determine its capability to represent the visual world. Perceptual bias, then, describes the difference between the assumed "ways of seeing" of a machine vision system, our reasonable expectations regarding its way of representing the visual world, and its actual perceptual topology. Research in computer science has shown that the perceptual topologies of many commonly used machine vision systems are surprisingly non-intuitive, and that their perceptual bias is thus surprisingly large.

We show how perceptual bias affects the interpretability of machine vision systems in particular, by means of a close reading of a visual analytics technique called "feature visualization" [Er09]. Feature visualization can be used to visualize the image objects that specific parts of a machine vision system are "looking for". While, on the surface, such visualizations do make machine vision systems more interpretable, we show that the more legible a feature visualization image is, the less it actually represents the perceptual topology of a specific machine vision system. While feature visualizations thus indeed mitigate the opacity of machine vision systems, they also conceal, and thus potentially perpetuate, their inherent perceptual bias. Feature visualizations and other visual analytics techniques, we argue, should thus not be understood so much as direct "traces" or "reproductions" of the perceptual topology of machine vision systems (analog to the technical images of photography) but more as indirect "illustrations", as "visualizations" in the literal sense of forcibly making-visual (and thus making visible and subsequently making interpretable) the non-visual. They should be understood as technical metapictures in the sense of W.J.T. Mitchell [Mi95], as images about (machine) seeing. The development of critical visual analytics techniques, then, becomes an important (digital) humanities challenge, situated at the interface of computer science and visual studies.

2 Building Blocks of Perceptual Bias

2.1 Deep convolutional neural networks

Our investigation looks at machine vision systems based on deep convolutional neural networks (CNNs), one of the most successful machine learning techniques within the larger artificial intelligence revolution we are witnessing today [Kr12]. CNNs have significantly changed the state of the art for many computer vision applications: object recognition, object detection, human pose estimation, and many other computer vision tasks are powered by CNNs today, superseding "traditional" feature engineering processes. For the purpose of this investigation, we will describe CNNs from a topological perspective rather than a mathematical perspective. In other words, we propose to understand CNNs as spatial structures. From the topological perspective, we can describe CNNs as layered systems. In the simplest version of a (non-convolutional) neural network, individual layers consist of neurons, atomic units that take in values from neurons in the previous layer and return

some weighted sum of these values. Deep convolutional neural networks, then, introduce new classes of neurons, which perform more complex functions like convolution. Common CNN architectures can have millions of neurons and even more interconnections between these neurons. It is thus close to impossible to infer from looking at the source code, data, weights, or any other aspect of a CNN, either alone or in conjunction, what it does, or what it has learned. [SB18] have suggested calling this opacity "inscrutability".

Inscrutability, however, is not the only reason for the notorious opacity of CNNs. As [SB18] argue, CNNs are also non-intuitive. The internal "reasoning" of neural networks does not necessarily correspond to intuitive methods of inference, as hidden correlations often play an essential role. [SB18] have argued that the non-intuitiveness of CNNs could be described as an "inability to weave a sensible story to account for the statistical relationships in the model. Although the statistical relationship that serves as the basis for decision-making might be readily identifiable, that relationship may defy intuitive expectations about the relevance of certain criteria to the decision."

2.2 Interpretable machine learning

This problem has been widely recognized in the technical disciplines as the problem of building interpretable machine learning systems, also referred to as explainable artificial intelligence systems. Such systems would, either by design or with the help of external tools, provide human-understandable explanations for their decisions, self-mitigating both their inscrutability and non-intuitiveness. In the past three to five years, research in interpretable machine learning has matured into a proper subfield of computer science [Li16; DK17; Gi18] and a plethora of statistical tricks has been developed to ensure the interpretability of simpler models like linear regression. Beyond these technical results, however, a larger conceptual discussion has emerged in the technical disciplines as well that "infringes" on the terrain of the humanities [Of17]. It is centered around attempts to find quantitative definitions for concepts that naturally emerge from the problem at hand, such as "interpretation" and "representation", with the help of methods and concepts from disciplines as diverse as psychology, philosophy, and sociology, building a "rigorous science of interpretable machine learning", as [DK17] write. We propose that, for machine vision systems, this inherent transdisciplinarity implies linking technical concepts and concepts from visual studies. In particular, it suggests understanding the interpretation of machine vision systems as an act of image-making, both literally and metaphorically. This is why, in the following, we will look at feature visualization.

2.3 Feature visualization

Feature visualization belongs to a range of techniques for the visual analysis of machine learning systems called visual analytics [Ho18]. Originally developed by [Er09] and

continuously improved since, feature visualization has been shown to produce remarkable results [Ol17; Ol18]. Importantly, we choose to investigate feature visualization not for its specific relevance to the digital humanities context – in fact, attribution methods [Se17; Ch19] are more commonly employed in the analysis of cultural data [BO20] – but because it best demonstrates the issue of perceptual bias that affects all visual analytics methods. Technically, feature visualization is a straightforward optimization process. To visualize what a neuron in a deep convolutional neural network has learned, a random noise image is passed through the layers of the network up until the hidden layer that contains the neuron of interest. Normally, during the training or prediction stages, the image would be passed further on to the output layer. For the purpose of visualization, however, we are not interested in a prediction but in the "activation" of a single neuron, its individual response to a specific input image when it reaches the neuron's layer. Hence, instead of utilizing the original loss function of the network, this response is now interpreted as its loss function. In other words, it is now the response of a single neuron that drives the "learning" process. The important difference is that this new loss flows back through the network beyond the input layer and is used to change the raw pixel values of the input image. The input image is thus altered, while the network's internal interconnections remain untouched. The altered image is then being used again as the input image during the next iteration, and so on. After a couple iterations, the result is an image that highly activates one specific neuron.

3 Manifestations of Perceptual Bias

3.1 Syntactic bias

This process, however, is called "naïve" feature visualization for a reason. In almost all cases, images obtained with it will exclusively contain very high frequencies and will thus be "illegible" in both the syntactic and semantic sense: there will be no visible structure, and no recognizable content (fig. 1). The images may very well be the best possible images with regard to a specific neuron and may very well be the closest possible visualizations of what this neuron has learned. To the human observer, however, they contain no information. They are adversarial examples [Sz13; Go14b] – images that highly activate specific neurons or classes in a fully trained deep convolutional neural network, despite being utterly uninterpretable. Naïve feature visualization, then, shows us a first glimpse of the peculiar perceptual topology of CNNs. Perceptual bias, here, takes the form of syntactic bias. This syntactic bias, in turn, manifests as texture bias [Ge19], an inductive bias in CNNs that "naturally" appears in all common CNN architectures. Inductive biases are "general", prior assumptions that a learning system uses to deal with new, previously unseen data.

At this point, it is important to note that we will not consider modifying the inductive biases of the CNN itself as a solution to the problem of perceptual bias, as, for instance, [Ge19] suggest. More precisely, for the purpose of our investigation, we are interested in interpretable machine learning as a narrow set of post-hoc methods to produce explanations.

Thus, we will also not take the field of representation learning [Be13] into consideration, which is concerned with the development of mechanisms that enforce the learning of "better" representations. This restriction to the scope of our investigation has three main reasons. The first reason is the post-hoc nature of the bias problem. While efforts to build resistance to bias into machine learning models exist, there is, at the moment, no clear incentive for industry practitioners to do so, except for marketing purposes. It can thus be assumed that, in real-world scenarios, the detection and mitigation of bias will be mostly a post-hoc effort. The second reason is a simple historical reason. Thousands of machine learning models based on the exact perceptual topologies under investigation here have already been deployed in the real world, and the digital humanities in particular rely on these "off-the-shelf" models. Thus, it is of vital importance to understand, and be able to critique, such models and their perceptual biases. Finally, while impressive progress has been made in other areas of machine learning [Cr20], in machine vision, controlling and harnessing inductive biases can still be considered an open problem. Recent research suggests that at least one established principle of gestalt theory (the law of closure) does emerge in CNNs [Ki19; Ri17; FL18] as an inductive bias. Overall, however, the inductive biases of CNNs are still unclear [CS17] and thus unmanageable.

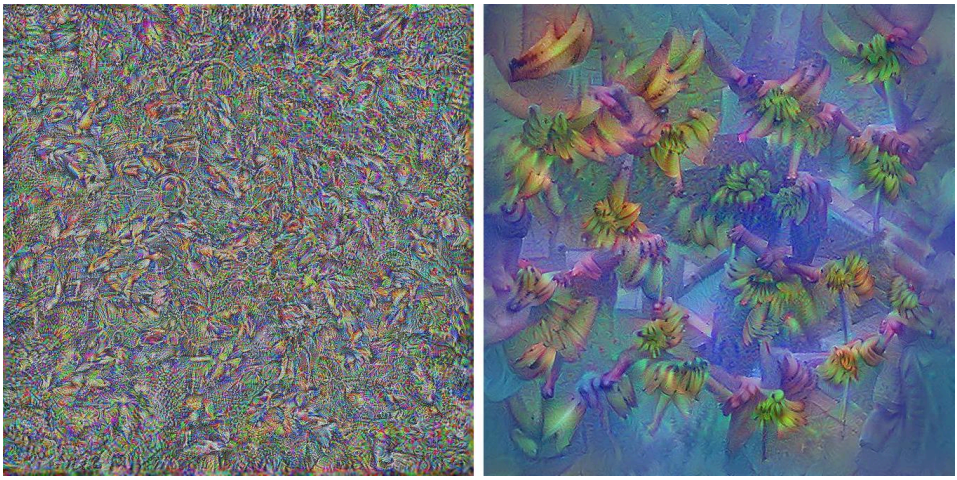


Fig. 1: Left: Unregularized feature visualization of the "banana" class of an InceptionV3 CNN [Sz16] trained on the 1,000 ImageNet classes in the ILSVRC2012 ImageNet subset [Ru15]. Right: regularized visualization of the same class.

Given these restrictions, the only option to mitigate this specific textural aspect of perceptual bias is to not change the model, but to change our image of it. In the case of feature visualization, it means adding back representational capacity to these images. It means introducing constraints – in other words, different biases – that allow the production of images that are images of something, instead of "just" images. Importantly, any such constraint, however, automatically moves the image further away from showing the actual perceptual topology of a CNN. It becomes less of a visualization, and more of a reconstruction. This

trade-off is the core problem of perceptual bias: it can only be overcome by shifting towards different, "better" biases, i.e. biases that shape our perception of the visual world. One strategy to "add back" the representational capacity to feature visualization is regularization (fig. 1). Regularization, here, simply means adding additional constraints to the optimization process. This can be achieved either by adapting the loss function – for instance, by using a quadratic loss function instead of just taking the mean of some values – or by applying transformations to the input image in regular intervals, for instance every few iterations in the optimization process. [Er09] introduced the concept of activation maximization, the core idea of iteratively optimizing an image to highly activate a selected neuron. From there, more and more elaborate regularization techniques started to appear, each introducing concrete suggestions for signal processing operations on the input image between iterations, on top of more common regularization techniques introduced through the loss function, like L2 regularization. Among these are jitter, blur total variation filters, bilinear filters, stochastic clipping and Laplacian pyramids. What all these techniques have in common is some kind of frequency penalization, i.e. the active avoidance of input images evolving into adversarial examples, either through optimizing for transformation robustness or through direct filtering.

3.2 Semantic bias

Despite all regularization efforts, however, feature visualizations often still present "strange mixtures of ideas" [Ol18]. Visualizing higher-level neurons in particular produces ambiguous results, images that might, or might not, show proper "objects" (fig. 2). To learn more about the logic of representation in CNNs, we thus have to ask: what is the relation between technical and semantic units, between artificial neurons and meaningful concepts, in CNNs? Trivially, at least for higher level neurons, individual feature visualization images must always have a degree of ambiguity that is directly correlated to the diversity of the training set. After all, the network has to be able to successfully classify a range of instances of an object with very different visual properties. In that sense, reality is "distributed", and it is no surprise that feature visualization images will reflect different manifestations of, and perspectives on, an object, akin to Cubist paintings.

But, the entanglement of concepts in the internal representations of a CNN goes beyond this "natural" ambiguity. Generally, we can state that, in all predictions of a CNN, all neurons play "a" role. Even if their role is just to stop the information flow, i.e. to pass on zero values to the next layer, these one-way streets are in no way less relevant to the classification accuracy of the whole system than all other neurons. In a way, concepts are thus "dissolved", or "entangled", when they are learned, and represented, by a CNN. Early work [Sz13] suggests that this entanglement is inevitable and absolute. Later work [Ba17] shows that some neural network architectures are less "naturally entangled" then others. Generally, however, significant supervision or, again, artificial inductive biases [Lo19] are required to

"disentangle" CNNs and arrive at a meaningful correspondence of technical and semantic units.

Perceptual bias, here, thus takes the form of semantic bias. Other than in the case of adversarial examples/texture bias, where perceptual bias affects the formal aspects of the visualization, here, it concerns aspects of meaning. Objects, for us, are necessarily spatially cohesive. If they are represented by CNNs, however, they lose this spatial coherence, different aspects of an object are attached to different neurons, which, in turn, get re-used in the detection of other objects. This missing coherence does not interfere with the CNN's ability to detect or classify spatially coherent objects in images but enables it. For feature visualization, which visualizes CNNs in their "natural", entangled state, reaching semantic interpretability thus implies the introduction of even more constraints. These additional constraints are so called natural image priors. Just as regularization is a syntactic constraint, biasing the visualization towards a more natural frequency distribution, so called natural image priors are a semantic constraint, biasing the visualization towards separable image objects.

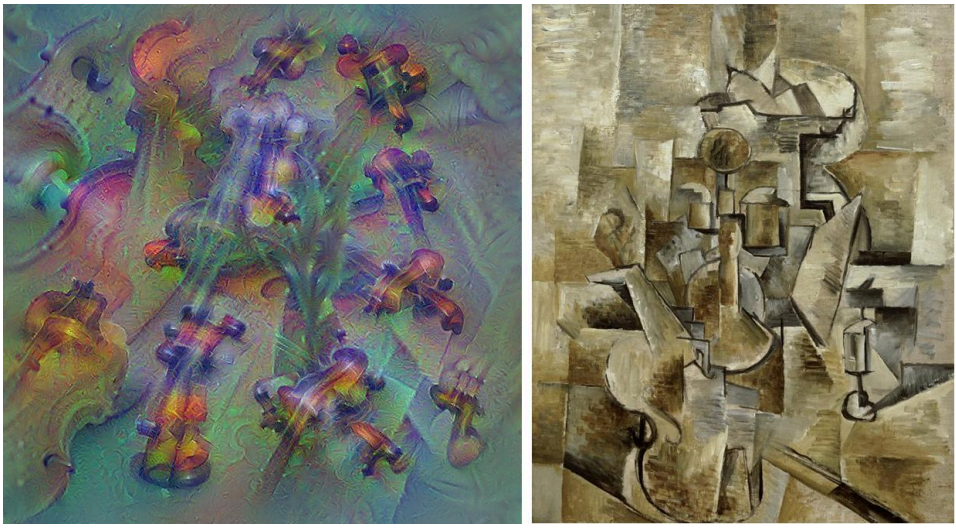


Fig. 2: Left: regularized feature visualization of the "violin" class of an InceptionV3 CNN [Sz16] trained on the 1,000 ImageNet classes in the ILSVRC2012 ImageNet subset [Ru15]. Right: George Braque, Violin and Candlestick (1910).

To produce natural image priors, [DB16; Ng16] propose to use a generative adversarial network (GAN)[3]. This implies, however, that the images that can be produced with this feature visualization method are entirely confined to the latent space of the specific GAN employed. Where regularization constrains the space of possible images to those with a "natural" frequency distribution, natural image priors constrain the space of possible images

[3] Unfortunately, we cannot explain GANs in detail here, and instead refer the reader to [Go14].

to the distribution of a GAN generator. In both cases, interpretable images are the result. These interpretable images, however, do not reflect the perceptual topology of the analyzed CNN. On the contrary: they intentionally get rid of the non-humanness that defines this topology, translating it into a human mode of perception that, in this form, simply does not exist in the CNN. To be images of something, feature visualizations have to be freed from the very mode of perception they are supposed to illustrate.

4 Technical Metapictures

As we have seen, the perceptual topology of machine vision systems, based on CNNs, is not "naturally interpretable". It is biased towards a distributed, entangled, deeply non-human way of representing the world. Mitigating this perceptual bias thus requires a forced "making legible". Feature visualization, as we have seen, is one possibility to achieve this forced legibility. However, feature visualization also exemplifies an essential dilemma: the representational capacity of feature visualization images is inverse proportional to their legibility. Feature visualizations that show "something" are further removed from the actual perceptual topology of the machine vision system than feature visualizations that show "nothing" (i.e. illegible noise). There is thus an irreconcilable difference between the human and machine perspective. As Thomas Nagel reminds us, there is a "subjective character of experience" [Na74], a surplus generated by each specific perceptual approach to the world that can never be "translated". Even if an external observer would be able to attain all the facts about such an inherently alien experience (analyze it in terms of "functional states"), they would still not be able to reconstruct said experience from these facts. Feature visualizations, then, should not be understood so much as direct "traces" or "reproductions" of the perceptual topology of machine vision systems (analog to the technical images of photography) but more as indirect "illustrations", as "visualizations" in the literal sense of forcibly making-visual (and thus making visible and subsequently making interpretable) the non-visual.

We thus propose to understand these images as technical metapictures, a term we adapt from W.J.T. Mitchell's picture theory [Mi95]. For Mitchell, metapictures are pictures that are "deeper" than "regular" pictures, as they incorporate a form of recursion: they are representations of representation "pictures about pictures" [Mi95, 36]. Mitchell identifies certain abilities of these pictures. "The metapicture [...] is the place where pictures reveal and 'know' themselves, where they reflect on the intersections of visuality, language, and similitude, where they engage in speculation and theorizing on their own nature and history" [Mi95, 82]. They are not only self-reflective but reflective on imagery and perception. "The metapicture is a piece of moveable cultural apparatus, one which may serve a marginal role as illustrative device or central role as a kind of summary image, what I have called a 'hypericon' that encapsulates an entire episteme, a theory of knowledge" [Mi95, 49]. The technical metapictures that feature visualization produces realize exactly this idea of a "summary image." They promise not a theory of images but a theory of seeing. More

precisely, their promise is exactly that of interpretable machine learning: to provide an intuitive visual theory of the non-intuitive perceptual topology of neural networks. In a sense, technical metapictures, and their use in interpretable machine learning, are thus an operationalization of the notion of metapicture itself.

For Mitchell, this epistemological power of metapictures, then, equips them with a sort of agency. Metapictures "don't just illustrate theories of picturing and vision: they show us what vision is, and picture theory" [Mi95, 57]. This agency, however, is actualized only if and when it comes into contact with a viewer. To make sense, to actually provide the reflection on images and vision that they promise, metapictures require a viewer. In the case of feature visualization, this interpretation has to happen not only on the level of the viewer but also on the technical level, where a significant effort has to be made to translate the anti-intuitive perceptual topology of a machine vision system into human-interpretable images in the first place. This includes adding information from the outside, for instance in the form of natural image priors. In other words, technical metapictures manifest an implicit, technical notion of interpretation, that is inseparable from the explicit interpretation that they also require. Interpretations based on feature visualization images thus become (human) interpretations of (technical) interpretations [Of18].

In conclusion: analyzing and understanding perceptual bias in machine vision systems requires reframing it as a problem of interpretation and representation, for which we have adapted W.J.T. Mitchells notion of the metapicture. Technical metapictures, we have argued, mirror the act of interpretation in the technical realm: regularization and natural image priors make feature visualization images legible before any interpretation can take place. Paradoxically, however, as the representational capacity of feature visualization images is inverse proportional to their legibility, this pre-interpretation presents itself as a massive technical intervention as well, that disconnects the visualization from the visualized. All of this suggests that visual analytics is an essentially humanist endeavor that calls for additional transdisciplinary investigations at the interface of computer science and visual studies.

Bibliography

[Ba17] Bau, D. et al.: Network dissection: Quantifying interpretability of deep visual representations. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2017).

[BO20] Bell, P., Offert, F.: Reflections on connoisseurship and computer vision. Journal of Art Historiography 23 (2020).

[Be13] Bengio, Y. et al.: Representation learning: A review and new perspectives. IEEE Transactions on Pattern Analysis and Machine Intelligence. 35, 8, 1798–1828 (2013).

[Be19] Benjamin, R.: Race After Technology: Abolitionist Tools for the New Jim Code. John Wiley & Sons (2019).

[BG18] Buolamwini, J., Gebru, T.: Gender shades: Intersectional accuracy disparities in commercial gender classification. In: Conference on Fairness, Accountability and Transparency. (2018).

[Ch19] Chen C. et. al.: This looks like that. Deep learning for interpretable image recognition. In: Advances in Neural Information Processing Systems. pp 8930–8941 (2019).

[CS17] Cohen, N., Shashua, A.: Inductive bias of deep convolutional networks through pooling geometry. arXiv preprint arXiv:1605.06743. (2017).

[Cr20] Cranmer, M. et al.: Discovering symbolic models from deep learning with inductive biases. arXiv preprint arXiv: 2006.11287. (2020).

[CP19] Crawford, K., Paglen, T.: Excavating AI: The politics of images in machine learning training sets. (2019).

[DK17] Doshi-Velez, F., Kim, B.: Towards a rigorous science of interpretable machine learning. arXiv preprint arXiv:1702.08608. (2017).

[DB16] Dosovitskiy, A., Brox, T.: Generating images with perceptual similarity metrics based on deep networks. In: Advances in Neural Information Processing Systems. pp. 658–666 (2016).

[Er09] Erhan, D. et al.: Visualizing higher-layer features of a deep network. Université de Montréal (2009).

[FL18] Feinman, R., Lake, B.M.: Learning inductive biases with simple neural networks. arXiv preprint arXiv:1802.02745. (2018).

[Fr19] Friedler, S.A. et al.: A comparative study of fairness-enhancing interventions in machine learning. In: ACM Conference on Fairness, Accountability, and Transparency (FAT*). (2019).

[Ga16] Garvie, C. et al.: The perpetual line-up: Unregulated police face-recognition in America. Georgetown Law, Center on Privacy & Technology (2016).

[Ge19] Geirhos, R. et al.: ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness. arXiv preprint arXiv:1811.12231. (2019).

[Gi18] Gilpin, L.H. et al.: Explaining explanations: An overview of interpretability of machine learning. In: 2018 IEEE 5th International Conference on Data Science and Advanced Analytics. pp. 80–89 IEEE (2018).

[Go14] Goodfellow, I. et al.: Generative adversarial nets. In: Advances in Neural Information Processing Systems. pp. 2672–2680 (2014).

[Go14b] Goodfellow, I.J. et al.: Explaining and harnessing adversarial examples. arXiv preprint arXiv:1412.6572. (2014).

[Ho18] Hohman, F.M. et al.: Visual Analytics in deep learning: An interrogative survey for the next frontiers. IEEE Transactions on Visualization and Computer Graphics. (2018).

[Ki19] Kim, B. et al.: Do neural networks show gestalt phenomena? An exploration of the law of closure. arXiv preprint arXiv:1903.01069. (2019).

[Kr12] Krizhevsky, A. et al.: ImageNet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems. pp. 1097–1105 (2012).

[Li16] Lipton, Z.C.: The mythos of model interpretability. In: 2016 ICML Workshop on Human Interpretability in Machine Learning (WHI 2016), New York, NY. (2016).

[Lo19] Locatello, F. et al.: Challenging common assumptions in the unsupervised learning of disentangled representations. arXiv preprint arXiv:1811.12359. (2019).

[Me19] Merler, M. et al.: Diversity in faces. arXiv preprint arXiv:1901.10436. (2019).

[Mi95] Mitchell, W.J.T.: Picture Theory: Essays on Verbal and Visual Representation. University of Chicago Press (1995).

[Mi19] Mittelstadt, B. et al.: Explaining Explanations in AI. In: ACM Conference on Fairness, Accountability, and Transparency (FAT*). (2019).

[Na74] Nagel, T.: What is it like to be a bat? The Philosophical Review. 83, 4, 435–450 (1974).

[Ng16] Nguyen, A. et al.: Synthesizing the preferred inputs for neurons in neural networks via deep generator networks. In: Advances in Neural Information Processing Systems. pp. 3387–3395 (2016).

[Of17] Offert, F.: "I know it when I see it". Visualization and intuitive interpretability. arXiv preprint arXiv:1711.08042. (2017).

[Of18] Offert, F.: Images of image machines. Visual interpretability in computer vision for art. In: European Conference on Computer Vision. pp. 710–715 Springer (2018).

[Ol17] Olah, C. et al.: Feature visualization. Distill. (2017).

[Ol18] Olah, C. et al.: The building blocks of interpretability. Distill. (2018).

[Ri17] Ritter, S. et al.: Cognitive psychology for deep neural networks: A shape bias case study. arXiv preprint arXiv:1706.08606. (2017).

[Ru15] Russakovsky, O. et al.: ImageNet large scale visual recognition challenge. International Journal of Computer Vision. 115, 3, 211–252 (2015).

[Se19] Selbst, A.D. et al.: Fairness and abstraction in sociotechnical systems. In: ACM Conference on Fairness, Accountability, and Transparency (FAT*). (2019).

[SB18] Selbst, A.D., Barocas, S.: The intuitive appeal of explainable machines. Fordham Law Review. 87, (2018).

[Se17] Selvaraju R.R. et. al.: Grad-CAM. Visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 618–626 (2017).

[Sz13] Szegedy, C. et al.: Intriguing properties of neural networks. arXiv preprint arXiv:1312.6199. (2013).

[Sz16] Szegedy, C. et al.: Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2818–2826 (2016).

R. Reussner, A. Koziolek, R. Heinrich (Hrsg.): INFORMATIK 2020,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2021 1307

Proof of Concept: Automatic Type Recognition

Vincent Christlein[1] Nikolaus Weichselbaumer[2] Saskia Limbach[3] Mathias Seuret[1]

Abstract: The type used to print an early modern book can give scholars valuable information about the time and place of its production as well as its producer. Recognizing such type is currently done manually using both the character shapes of 'M' or 'Qu' and the size of the total type to look it up in a large reference work. This is a reliable method, but it is also slow and requires specific skills. We investigate the performance of type classification and type retrieval using a newly created dataset consisting of easy and difficult types used in early printed books. For type classification, we rely on a deep Convolutional Neural Network (CNN) originally used for font-group classification while we use a common writer identification method for the retrieval case. We show that in both scenarios, easy types can be classified/retrieved with a high accuracy while difficult cases are indeed difficult.

Keywords: type recognition; type classification; type retrieval; deep learning

1 Introduction

Type recognition is one of the central methods of analytical bibliography [Sc18, p. 42–70]. It is used to date printed books and identify both the printer and the publication place. Type recognition is traditionally done manually by using a combination of the type size (measured over 20 lines) and the characteristic shape of the letters 'M' or 'Qu' to identify the correct type in the Typenrepertorium der Wiegendrucke (TW), a reference work for all known incunabula types [Ha05].[4] This method is slow and requires specific skills. Most importantly, it relies on the existence of a reference work that is difficult and time-consuming to compile, making it next to impossible to use type recognition for material beyond the incunabula period. For this approach, pattern recognition methods are extremely helpful as they increase the speed and ease of type recognition and thereby widen the scope of material that it can be applied to. For book historians, using type recognition for books printed in the incunabula period and beyond helps to answer important questions. In the early modern period, many books appeared without any indication about when and where the book was printed. This was often done when authors and printers feared political prosecution, but not limited to such cases. Type recognition would enable us to identify the producers of

[1] Friedrich-Alexander-Universität Erlangen-Nürnberg, Pattern Recognition Lab, Martensstr. 3, 91058 Erlangen, Germany, vincent.christlein@fau.de, mathias.seuret@fau.de
[2] Johannes Gutenberg-Universität Mainz, Gutenberg-Institut für Weltliteratur und schriftorientierte Medien, Abt. Buchwissenschaft, Jakob-Welder-Weg 18, 55128 Mainz, Germany, weichsel@uni-mainz.de
[3] Università degli Studi di Milano, Department of Economics, Management and Quantitative Methods, Via Conservatorio 7, 20122 Milano, Italy, dr.saskia.limbach@gmail.com
[4] The Typenrepertorium der Wiegendrucke is available online: https://tw.staatsbibliothek-berlin.de

doi:10.18420/inf2020_122

these books. On top of that, we would gain a better understanding of the material used in a given print shop, which could tell us more about the economic background of the printer. This paper explores the effectiveness of existing pattern recognition methods from writer identification and font group recognition.

2 Related Work

We want to investigate type recognition in two different ways: classification and retrieval.

Classification An early work for font classification [Wa15] – working on modern computer fonts – built a large dataset of real and mostly synthetic images of about 5000 classes and 2 mio. images. The proposed classifier achieves about 80 % accuracy on the test set. Closer related are the competitions and datasets for cursive script type classification [Cl16; Cl17], where the best method can differentiate between 12 script classes with an accuracy of about 85 % [Ch19b; Cl17]. In contrast, font group classification [Se19], seems to be an easier task achieving accuracies of about 98 % [We20]. Note that we strive to classify types, which are much more challenging than font groups since the differences are often much smaller.

Retrieval Type retrieval is closely related to other image retrieval tasks, such as writer retrieval for historical data. The current writer identification performance is well represented in the last image retrieval competitions [Ch19a; Fi17]. These competitions involved large datasets containing 3600 [Fi17] and 20 000 test images [Ch19a]. The accuracies vary widely (74 % to 97 %) depending on the data source and image quality. The current state-of-the-art approach for historical writer identification is given by Lai et al. [LZJ20]. They propose "pathlet" features, which they combine with SIFT [Lo04] and encode it in a novel bagged version of VLAD encoding [Jé12]. They achieve about 90 % Top-1 accuracy and outperform the previous unsupervised deep learning-based approach by Christlein et al. [Ch17b]. In this study, we evaluate the performance of a baseline writer identification method [Ch18] based on SIFT descriptors.

3 Dataset

TW lists about 6000 different types used between 1450-1500. Yet, it is impossible to use all of these types in this dataset. The sheer size of TW already presents a challenge, but it is mainly its design and original purpose which prevents us from using all of it. TW only lists the type used in a given book, but gives no indication if this type is the only type used in the book and – more importantly in this context – on which pages the type is used. Most books were printed with more than one type. Additional types were used in order to emphasize words, highlight headlines, etc. As of now the only way to select training data is to manually

label every image. This made us limit our dataset to 8 examples which were selected to illustrate how easy and how difficult it can be to differentiate between various types.

The easy examples, cf. Fig. 1a, consist of types with very distinctive shapes. TW ma00131 is a Rotunda used by the Augsburg printer Anton Sorg between 1475 and 1478 in at least 19 editions of which we have a total of 21 digital copies. The type is very recognizable because of its unique decorative upper case characters. Note, the red highlights were later added by a contemporary hand and not printed.

TW ma00967 is a Textura used by an unknown printer from Salamanca from 1481 to 1490. While there is not as much surviving material for this type as for others (only 8 editions and 4 digital copies), the type is still easy to recognize for a human expert. The type is of rather poor quality and has a characteristic jagged look.

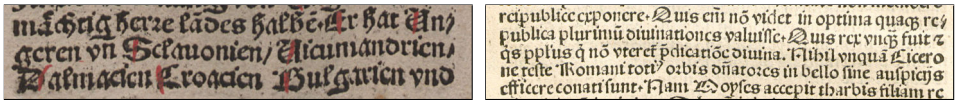
TW ma02771 is a Bastarda used by Jean du Pré in Lyon between 1489 and 1491. Relatively little material survives – TW lists 8 editions and only 2 digital copies –, but the type is rather large at 119 mm over 20 lines and with its looped ascenders and flourished upper case characters it uses very complex shapes that are often easier to attribute correctly than the regular shapes of e. g. a well-cut Textura.

TW ma04614 is a Textura with some unusual letter shapes used by Arnold ter Hoernen in Cologne from 1474 to 1482. It survives in 11 editions of which 5 copies are available as scans. The type is unusual in that most of its letter shapes follow the model of a Textura, but ‘f’ and ‘f’ have descenders and the lower-case ‘a’ is of the cc-type, not the uncial type that is more common for Textura types.

To represent the other end of the spectrum we selected two pairs of types that we assumed to be particularly challenging. TW ma07487 and TW ma07488 were both used by Bartholomäus Kistler in Straßburg, ma07487 from 1498 to 1499 (2 editions, 2 digital copies) and ma07488 from 1499 to 1501 (6 editions, 7 digital copies). The types share identical upper case Rotunda characters, but combine them with Bastarda lower case (ma07487) and Rotunda lower case (ma07488). Presumably, ma07487 was a temporary fix as the production of the lower case was not yet finished. This kind of combination appears fairly often and poses particular challenges as parts of the types are not similar, but completely identical.

The second pair of difficult types, TW ma07721 and TW ma07718 were successively used by Johann Schäffler in Ulm from 1492 to 1494 (5 editions, 2 digital copies) and from 1496 to 1500 (22 editions, 11 digital copies). Both types are Upper Rhine Bastardas of very similar size and design. They do however not share any perfectly identical characters. TW ma07718 only replaced TW ma07721 after Schäfflers attempt to establish a print shop in Freising. It was apparently made to replace the older type [Am79, p. 370–371]. These types can be told apart by a human expert, but only by methodical comparison. At a cursory examination they can easily be mistaken for one another.

(a) Easy cases – TW ma00131 (top left), TW ma00967 (top right), TW ma02771 (bottom left), TW ma04614 (bottom right).



(b) Difficult case 1 – TW ma07487 (left) and TW ma07488 (right).

(c) Difficult case 2 – TW ma07721 (left) and TW ma07718 (right).

Fig. 1: Type dataset examples.

To create training data we used TW and its sister catalogue the Gesamtkatalog der Wiegendrucke (GW) [Ge5–]. For a given type we looked up all editions that are known to contain this type in TW. Via GW we searched for and downloaded all accessible scans of these editions. In the next step, we selected pages that contained only the respective type and in some cases cropped images in order to delete headlines, woodcuts and other irrelevant material. The dataset consists of 9083 labeled images and is publicly available.[5]

4 Classification

As a first step to approach this problem, we train a classifier to recognize the selected types. We use the same methodology as in [Se19]. The baseline method of this work consists of a Convolutional Neural Network (CNN) to classify the text of document images as belonging to different font groups, such as Antiqua, Fraktur or Textura.

4.1 Methodology

We use a DenseNet-121 [Hu17], a deep CNN with densely-connected blocks, to classify overlapping patches over the whole surface of the input image, and average the results. We

[5] https://doi.org/10.5281/zenodo.3923638

Tab. 1: Confusion matrix containing the classification results. To enhance readability, the diagonal is bold. Rows correspond to types, and columns to classification results. The first four rows depict the easy cases while the last four rows show the difficult cases.

prediction →	ma00131	ma00967	ma02771	ma04614	ma07487	ma07488	ma07718	ma07721
ma00131	**160**	0	0	0	0	0	0	0
ma00967	0	**115**	0	0	0	0	0	0
ma02771	0	0	**711**	0	0	0	0	0
ma04614	2	0	0	**225**	0	0	0	0
ma07487	1	0	0	0	**0**	6	225	0
ma07488	0	0	0	0	0	**124**	11	0
ma07718	0	0	0	0	0	0	**2**	0
ma07721	0	0	0	0	0	6	1012	**0**

took the pre-trained one provided by Seuret *et al.*[6] [Se19], and replaced its last layer by a new one with eight outputs. As this network has been trained for font groups classification, we expect it to have learned features useful for type classification. Note that we train only the last layer, leaving the other ones frozen. In earlier tests, which we conducted, we realized that fine-tuning the whole CNN makes it over-fit quickly.

The training is done as follows: First, we create a training set consisting of 5000 patches of 300×300 pixels for each type, uniformly distributed over the available training images. Then, during the training of the CNN, we apply data augmentation on these patches and extract 224×224 pixels center crops. We used the following data augmentation strategies: random rotations between $]-15, 15[$ degrees, shearing of an angle between $]-3, 3[$, re-scaling by a factor in the range of $]0.9, 1.1[$, color jittering (PyTorch settings: 0.7, 0.7, 0.3, 0.03). Additionally, we add JPG artifacts with quality factors in the range of $[2, 100[$, and a low binarization probability with Otsu (5 %) or Sauvola (2.5 %).

The network is trained for 10 epochs, with an initial learning rate of 0.0005, a weight decay of 0.00001, and a momentum of 0.9. After each epoch, the learning rate is decreased by 5 %. Due to the small amount of images for some types, no validation set is used.

4.2 Results

For the evaluation, we split the dataset into a training and a test set. The training set consists of 6472 samples while the test set contains 2600 document images. We made sure that there is no document overlap in the subsets to guarantee document-independent testing.

A confusion matrix presenting the classification results is shown in Tab. 1. We can see that the system reaches an overall classification accuracy of 51.4 % and an average accuracy of

[6] https://github.com/seuretm/ocrd_typegroups_classifier

73.9 %, which already indicates that some classes are well recognized in contrast to others. The overall accuracy is significantly lower than the accuracy obtained on the training data (over 80 % of the patches), which implies that we are running into over-fitting despite the rather aggressive augmentation approach.[7]

While the classification of the easy types was successful, several documents of the difficult types could not be detected properly. What makes the training for type ma07721 especially difficult is the fact that there is only one single training sample available, even when cropping a sufficient number of patches, the script variance might be too low for a reliable training. This suggests the use of a retrieval scenarios where the learning of a good embedding is in focus.

5 Retrieval

In addition to classification, we test the retrieval scenario. That means, we want to retrieve the most similar types given a query image.

5.1 Methodology

We make use of the general writer identification framework by Christlein [Ch18]. It consists of a sampling step, where we evaluate two strategies: SIFT keypoints [Lo04] computed at (a) the original images, (b) contours extracted by means of the well-known Canny edge detector [Ca86]. For the latter approach, we set the two hysteresis thresholds automatically [WMB15].

Afterwards, SIFT descriptors [Lo04] are computed at the keypoint locations. They are Dirichlet-normalized and PCA-whitened following Christlein [Ch18]. Afterwards, the local descriptors are encoded using VLAD [Jé12] encoding using 100 clusters for the codebook. For improving the VLAD embedding [CM18], we employ Generalized Max-Pooling (GMP) [Mu16] with $\lambda = 1000$ in combination with power normalization (power of 0.5) and ℓ^2-normalization, i. e. normalizing the global descriptor such that its norm equals one. This process is repeated five times, the resulting global descriptors are concatenated and jointly PCA-whitened and dimensionality reduced to 6400 components and ℓ^2-normalized again.

Finally, an Exemplar-SVM (ESVM) is computed, which has shown to improve the writer identification results [Ch17a]. We use the ESVM as a feature transformation [Ch17b], i. e. the ℓ^2-normalized coefficients of the ESVM are used as new feature descriptor. These descriptors are then compared using the cosine distance, which equals a dot product of the ℓ^2-normalized descriptors.

[7] Note that a major constraint is that the augmentation should not make a type look like another one.

Tab. 2: Retrieval results for (a) easy testset using the difficult for training and (b) difficult testset using the easy one for training. The first row denotes a document-dependent scenario, where all samples of the test set are used for each query sample. For the other experiments, retrieved samples from the same document are ignored.

		(a) Easy			(b) Difficult		
	Sampling	Top-1	Top-10	mAP	Top-1	Top-10	mAP
1vsAll	Keypoint+SIFT	98.9	99.9	62.7	99.8	99.9	96.3
1vsOtherDocs	Keypoint+SIFT	88.9	94.2	54.6	50.0	50.2	46.6
	Keypoint+SIFT+ESVM	93.3	95.3	56.8	50.0	50,3	47.1
	Contour+SIFT	72.8	80.8	47.1	49.9	65.4	57.3
	Contour+SIFT+ESVM	76.5	80.2	49.2	49.9	65.4	57.3

5.2 Results

For the evaluation, we split the dataset into a type-independent training and test set. For simplicity, we choose all images of the easy types (#samples: 7029) as one subset and all the images of the difficult types (#samples: 2043) as the other subset. We then evaluate the following two configurations: (1) trained with the difficult subset and tested with the easy one and (2) the other way around, i. e. trained with the easy subset and tested with the difficult one. We made sure that there is no type overlap in the subsets to guarantee type independent testing. Note that this is different from the classification scenario, where we know the classes in advance.

We report typical retrieval measures, such as Top-1 accuracy as well as mean average precision (mAP), which is a measure of the overall ranking of the relevant documents in respect to the query sample. Additionally, we give the Top-10 accuracy, i. e. the chance of finding at least one sample of the query type among the first ten ranked results.

First, we compute the typical leave-one-image-out scenario, i. e. every test sample is used as query and *all* remaining ones are ranked according to their similarity with the query. Tab. 2 (first row) shows that this works astonishingly well with rates beyond 99 %. This is only natural, since each query sample comes from a specific document, and the remaining document images are among the other samples of the test set. In other words, the algorithm most probably retrieves images from the same document.

For the remaining experiments, we evaluate the retrieval performance in a document-independent way. Therefore, we ignore images from the same document during the metric computation. This results in a drop in performance but is a much more realistic scenario. When comparing the different sampling methods (keypoint vs. contour), we see that keypoint-based sampling is superior to contour-based sampling when trained on the difficult types samples and tested on the easy ones, see Tab. 2a. The reverse behavior is shown for the difficult cases where contour-based sampling is in favor, cf. Tab. 2b. This might be

related to the quite imbalanced test document sizes; one of the difficult types has only one image in one document, thus the metrics are biased when retrieving this specific one. This might also be the reason for the bad results for the difficult subset. ESVMs show also to be beneficial for type retrieval, at least for the easy cases.

6 Conclusion

In this work, we analysed the possibilities of type recognition in the two scenarios classification and retrieval. Therefore, we adopted and evaluated two baseline systems, originally developed for font-group classification and writer identification. While the classification network achieves an overall low accuracy, this can be attributed to the class imbalance of the training set. In the case of type retrieval, a careful document-independent evaluation reveals that very similar-looking types are problematic for a common retrieval pipeline. For future work, we would like to investigate networks, trained by deep metric-learning methods, e. g. with the use of contrastive or triplet loss, which might enable to differentiate also very similar-looking types.

References

[Am79] Amelung, P.: Der Frühdruck im deutschen Südwesten. 1473–1500. Bd. 1, Ulm. Hiersemann, Stuttgart, 1979, ISBN: 3777279293.

[Ca86] Canny, J.: A Computational Approach to Edge Detection. IEEE Transactions on Pattern Analysis and Machine Intelligence 8/6, pp. 679–698, Nov. 1986.

[Ch17a] Christlein, V.; Bernecker, D.; Hönig, F.; Maier, A.; Angelopoulou, E.: Writer Identification Using GMM Supervectors and Exemplar-SVMs. Pattern Recognition 63/, pp. 258–267, 2017.

[Ch17b] Christlein, V.; Gropp, M.; Fiel, S.; Maier, A.: Unsupervised feature learning for writer identification and writer retrieval. In: 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR). Vol. 1, IEEE, Kyoto, pp. 991–997, Nov. 2017.

[Ch18] Christlein, V.: Handwriting Analysis with Focus on Writer Identification and Writer Retrieval, PhD thesis, FAU Erlangen-Nürnberg, 2018.

[Ch19a] Christlein, V.; Nicolaou, A.; Seuret, M.; Stutzmann, D.; Maier, A.: ICDAR 2019 Competition on Image Retrieval for Historical Handwritten Documents. In: 2019 International Conference on Document Analysis and Recognition (ICDAR). Pp. 1505–1509, 2019.

[Ch19b] Christlein, V.; Spranger, L.; Seuret, M.; Nicolaou, A.; Král, P.; Maier, A.: Deep Generalized Max Pooling. In: 2019 International Conference on Document Analysis and Recognition (ICDAR). Pp. 1090–1096, Sept. 2019.

[Cl16] Cloppet, F.; Eglin, V.; Kieu, V. C.; Stutzmann, D.; Vincent, N.: ICFHR2016 Competition on the Classification of Medieval Handwritings in Latin Script. In: 2016 15th International Conference on Frontiers in Handwriting Recognition (ICFHR). Pp. 590–595, 2016.

[Cl17] Cloppet, F.; Eglin, V.; Helias-Baron, M.; Kieu, C.; Vincent, N.; Stutzmann, D.: ICDAR2017 Competition on the Classification of Medieval Handwritings in Latin Script. In: 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR). Kyoto, pp. 1371–1376, Nov. 2017.

[CM18] Christlein, V.; Maier, A.: Encoding CNN Activations for Writer Recognition. In: 13th IAPR International Workshop on Document Analysis Systems. Vienna, pp. 169–174, Apr. 2018.

[Fi17] Fiel, S.; Kleber, F.; Diem, M.; Christlein, V.; Louloudis, G.; Stamatopoulos, N.; Gatos, B.: ICDAR2017 Competition on Historical Document Writer Identification (Historical-WI). In: 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR). Vol. 01, Kyoto, pp. 1377–1382, Nov. 2017.

[Ge5–] für den Gesamtkatalog der Wiegendrucke, K., ed.: Gesamtkatalog der Wiegendrucke. Hiersemann [u.a.], Leipzig; [later:] Stuttgart, 1925–.

[Ha05] Haebler, K.: Einführung. In (Haebler, K., ed.): Typenrepertorium der Wiegendrucke. Abt. I. Deutschland und seine Nachbarländer. Haupt, pp. IX–XXVIII, 1905.

[Hu17] Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K. Q.: Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. Pp. 4700–4708, 2017.

[Jé12] Jégou, H.; Perronnin, F.; Douze, M.; Sánchez, J.; Pérez, P.; Schmid, C.: Aggregating Local Image Descriptors into Compact Codes. Pattern Analysis and Machine Intelligence, IEEE Transactions on 34/9, pp. 1704–1716, Sept. 2012.

[Lo04] Lowe, D. G.: Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision 60/2, pp. 91–110, Nov. 2004.

[LZJ20] Lai, S.; Zhu, Y.; Jin, L.: Encoding Pathlet and SIFT Features With Bagged VLAD for Historical Writer Identification. IEEE Transactions on Information Forensics and Security 15/, pp. 3553–3566, 2020.

[Mu16] Murray, N.; Jegou, H.; Perronnin, F.; Zisserman, A.: Interferences in Match Kernels. IEEE Transactions on Pattern Analysis and Machine Intelligence 39/9, pp. 1797–1810, Oct. 2016, arXiv: 1611.08194.

[Sc18] Schmitz, W.: Grundriss der Inkunabelkunde: das gedruckte Buch im Zeitalter des Medienwechsels. Hiersemann, Stuttgart, 2018, ISBN: 9783777218007.

[Se19] Seuret, M.; Limbach, S.; Weichselbaumer, N.; Maier, A.; Christlein, V.: Dataset of Pages from Early Printed Books with Multiple Font Groups. In: Proceedings of the 5th International Workshop on Historical Document Imaging and Processing. Association for Computing Machinery, Sydney, pp. 1–6, 2019.

[Wa15] Wang, Z.; Yang, J.; Jin, H.; Shechtman, E.; Agarwala, A.; Brandt, J.; Huang, T. S.: DeepFont: Identify Your Font from An Image. In: Proceedings of the 23rd ACM International Conference on Multimedia. MM '15, Association for Computing Machinery, Brisbane, Australia, pp. 451–459, 2015, ISBN: 9781450334594.

[We20] Weichselbaumer, N.; Seuret, M.; Limbach, S.; Hinrichsen, L.; Christlein, V.: The rapid rise of Fraktur. In (Schöch, C., ed.): DHd 2020, Spielräume, Paderborn Mar. 2, 2020–Mar. 6, 2019. Zenodo, pp. 229–232, 2020.

[WMB15] Wahlberg, F.; Mårtensson, L.; Brun, A.: Large Scale Style Based Dating of Medieval Manuscripts. In: 3rd International Workshop on Historical Document Imaging and Processing (HIP'15). ACM, Nancy, pp. 107–114, Aug. 2015.

R. Reussner, A. Koziolek, R. Heinrich (Hrsg.): INFORMATIK 2020,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2021 1317

Modelling Medieval Vagueness

Towards a Methodology of Visualising Geographical Uncertainty in Historical Texts

Mateusz Fafinski [1], Michael Piotrowski [2]

Abstract: The project *An Agile Approach Towards Computational Modeling of Historiographical Uncertainty* is building a taxonomy of historiographical uncertainty. We are focusing on early medieval texts as our case studies, because they are characterised by a high degree of "high stakes" uncertainty and a varied historiography characterised by a vivid debate. The additional factor of the manuscript text-transmission ensues that also the material aspect of the textual study will be covered in our attempt to build an adaptable taxonomy of historiographical uncertainty. Computational humanities need a robust methodological platform, that can be applied to a wide variety of projects. Uncertainty in general and geographical uncertainty in particular stand as the crucial aspects of this platform. We investigate a methodology of visualising geographical locales in historical texts and their historiographies that explicitly models uncertainty in.

Keywords: uncertainty; mapping; historiography; medieval history

1 The Problem of Uncertainty in Historical Methodologies

The problem of uncertainty and vagueness in history and historiography is deeply embedded in historiographical practice. While vagueness and uncertainty are impossible to fully separate, they can nevertheless be modelled on a spectrum where *vagueness* is a category rooted on the source side and *uncertainty* on the side of the historiographical interpretation. While each of them is anchored at opposing sides of a gradient, they are both always present, and trying to fully separate them is counterproductive – as Edgington [Ed92, p. 203] remarked, "vagueness and uncertainty can interact."

For the early narrative historians like Thucydides [Th98] uncertainty was more or less a question of believability of sources. Uncertainty was the absence of reliable information and not necessarily a presence of ambiguity. Indeed, the citing practice of "it is said," a distancing technique, allowed for a binary understanding of uncertainty between hearsay and "perfect" knowledge [Gr11].

[1] Faculty of Arts, Department of Language and Information Sciences, University of Lausanne, bâtiment Anthropole, 1015 Lausanne, Switzerland, mateusz.fafinski@unil.ch, https://orcid.org/0000-0003-1637-8174
[2] Faculty of Arts, Department of Language and Information Sciences, University of Lausanne, bâtiment Anthropole, 1015 Lausanne, Switzerland, michael.piotrowski@unil.ch, https://orcid.org/0000-0003-3307-5386

doi:10.18420/inf2020_123

This is a feature, not a bug, of early historiographies, as uncertainty becomes essentially a narrative technique to make a point, a claim. This method underlines the early attempts to tackle uncertainty, but they can be summarised under the equation of uncertainty with unreliability. This process was of extreme importance for later methodology of history, as it put source criticism and narrative techniques in the very centre of strategies to deal with historical uncertainty.

This strategy of choosing between variants, especially in ancient writers like Herodotus or Xenophon, has been deemed "narrative uncertainty" [Ma97, p. 281]. As a strategy (not a model) it allowed the early narrative historians to choose among the variants in their sources in order to shape their stories. Narrative uncertainty permeates all the levels and types of vagueness present in those texts. Moreover, scholarly editions and digital facsimiles introduce another layer between us and the source and thus another level of uncertainty. Imaging (or creation of digital facsimiles) is in this respect no different to any other form of processing of historical sources [Pr08].

The focus to date in many disciplines of historical research has often been on reducing uncertainty [see, e.g., Bl11]. Even when acknowledged, uncertainty was to be modelled in order to be factored *out* rather than factored *in*. In this method the vagueness of the sources should be analysed to the point of the lowest possible uncertainty in their interpretation. This reductive approach is caused by the deep unease with fuzziness in some methodologies of history, seen as responsible for potentially false outcomes. The goal of the historian was in those approaches to reconstruct the one-dimensional facts of the past, "to extract the facts in such a way as to arrive at the truth" [Sk97, p. 306]. Nevertheless, among the researchers of the historical method the need to model and factor uncertainty in has been recognised, including the importance it can play at the interface between history and informatics [To84, pp. 510–513]. In this spirit, there is today a growing, although still mostly ad hoc, understanding in digital scholarship that this "spurious exactitude" [Ta11] and attempts to force uncertainty out at every cost is detrimental to our ability to actually research the past. More and more projects are thus explicitly factoring in uncertainty in their individual methodologies [see, e.g., Bi14].

2 Factoring Uncertainty In

As opposed to the minimising approach, we want to focus on the explicit modelling of uncertainty in order for it to become an integral part of computational humanities methodology, as we have already advocated elsewhere [Pi19]. As our case study we have chosen the work of Gregory of Tours, a 6th-century historian concerned mainly with the events, locales, and persons in the territory of modern France, Germany, Italy, and Spain [Gr74]. We recognise the rich historiographical tradition on Gregory and the fact that his work is in itself a historiography, in which vagueness and uncertainty are not a simple matter of a lack of knowledge but are conscious tools for creating community [Re13], presenting a particular vision of the past [He94], and which have generated rich reflection already in

the early medieval period [Re15]. Our attempt is based on a three-pronged approach to visualising geographical vagueness in early medieval texts. First, we are concerned with the uncertainty concerning the manuscripts that transmit the texts, crucial to the creation of what we call today *Historia Francorum* – a very much interpretative creation on its own [Go89]. Their age and place of production are crucial for the editorial choices undertaken when producing the editions and translations of those texts and the introduction of the "editorial narrative" [Ra16, p. 152]. Second, we are concerned with the distribution of the vagueness and uncertainty within the text: its typology and ontology, an issue already flagged as crucial for knowledge retrieval from texts [KC15]. Third, we are concerned with the actual mapping of the locations within the text: how the vagueness and uncertainty of the text of Gregory is projected onto a two-dimensional map.

We can see that when it comes to modelling uncertainty there is a high degree of interrelatedness between those different types. Because the aim of our project is to work towards a historiographical methodology of uncertainty we also try to identify not only its level but also the historiographical stakes involved. The level of uncertainty is established based on how much information the text delivers about a particular category. The historiographical stakes are determined based on how much this particular type of uncertainty influences the historiographical interpretation of the text itself. And so, we identify different forms of uncertainty in our case study and categorise them according to those two factors (uncertainty level/historiographical stakes of that form of uncertainty):

1. In-source uncertainty:
    - sources of Gregory (high/high)
    - trustworthiness of his text (high/high)
    - language of Gregory (to what extent the texts that we have in later copies, reflect the language of Gregory himself); his orthography, matters of transition from late Latin to Romance (high/low)
    - locations, dates, persons – the content uncertainty, the area where the most historiographical debates happen (low/high)
2. Supra-source uncertainty:
    - the manuscript transmission, which models also the extent to which the text that we have is actually the text of Gregory (low/low)
    - the texts for which Gregory is a source (low/high)
    - the historiographical uncertainty, i.e., the historiographical models and narratives built on the basis of particular interpretations of the in-source uncertainties (high/high)

In this paper we focus on the geographical uncertainty in both domains: in the text and outside of it.

3 Visualizing Geographical Uncertainty

Vagueness is inherent in the descriptions of locales mentioned in Gregory's writings. We recognise that these texts are imbued with a degree of vagueness and background noise – in effect every location is to a certain extent uncertain and so is its approximation on a two-dimensional map. In this respect as a work of history it shows striking similarities to literary texts – being in effect both – and requires similar attention to modelling its uncertain geodata [see RPH13]. In geographical information systems (GIS), uncertainty is often defined as "a measure of the user's understanding of the difference between the contents of a dataset and the real phenomena that the data are believed to represent" [Lo05, p. 128], i.e., the difference between the geographical position of a locale and the author's understanding of that position. In our case, there are two additional levels. One is the semantic uncertainty: differing meanings that are assigned to the linguistic markers representing these locales [BGP12]. The second one are the uncertainties of translation [He16]. It features prominently in translation theory [see, e.g., HM91] and directly influences historiographies in various languages. In other words, our author operates on a high initial degree of vagueness (the difference between his understanding of the locales and their actual geographical positions is large); his understanding of the semantic quantifications of areas is uncertain (e.g., defining kingdoms as areas of influence of particular rulers); those locales are originally described in Latin, but are in modern historiographies translated into different languages.

Geographical locations in historical texts might be referred to through terms, phrases, and concepts that have nothing – or very little – to do with geographical terminology. This renders any attempt to automate their extraction and visualisation without a robust uncertainty schema almost futile. Inclusion of uncertainty modelling remains in this respect a crucial aspect. While in GIS a strong focus is laid on the uncertainty of geospatial data, [Go20] when it comes to modelling uncertainty in historical and historiographical texts additional layers appear and we are confronted with a much richer structure of uncertainty.

Visualising this vagueness requires the application of different degrees of uncertainty. Even points on a map (e.g., "Roma") can be recognised as being in essence fuzzy approximations of (a) Gregory's understanding of where "Roma" is; (b) our understanding of what area Gregory means by "Roma"; (c) our understanding of what "Rom," "Rome," "Rzym," etc., represent on a map. Visualising historical sources without acknowledging and factoring in uncertainty is then in effect a visualisation of no more than a historiographical narrative – an interpretation of those sources. Oftentimes digital humanities projects leave the explicit acknowledgment of this narrative out in order to factor the uncertainty out, but in reality, by failing to make this narrative explicit, they are, simply speaking, mapping the wrong thing [Fa20]. Our understanding of the geographical space is also different from the understanding of the authors of our sources. This understanding has been progressively translated through various historiographical interpretations and created a new geography to be mapped: a *subjective structure* [To97], an additional layer of interpretative geography created by historians. Thus vagueness and uncertainty make numerous (but nevertheless limited) historiographical narratives possible and lead to sometimes risky but high-stakes

statements [Ko77]. In historiography this creation of interpretative layers is a long-recognised phenomenon [see, e.g., Wh73]. But with the advent of digital and computational humanities it remained an intuitive and implicit element of the methodology of those new branches. It can help us, for example, recognise the geographical horizon of the author of a source through computational methods. The measure of the area which can be assigned as characterised by a low level of geographical uncertainty corresponds with the expression of the geographical horizon of the author in a particular text. But this method will only work if we recognise, model, and factor in the historiographical uncertainty associated with a particular source.

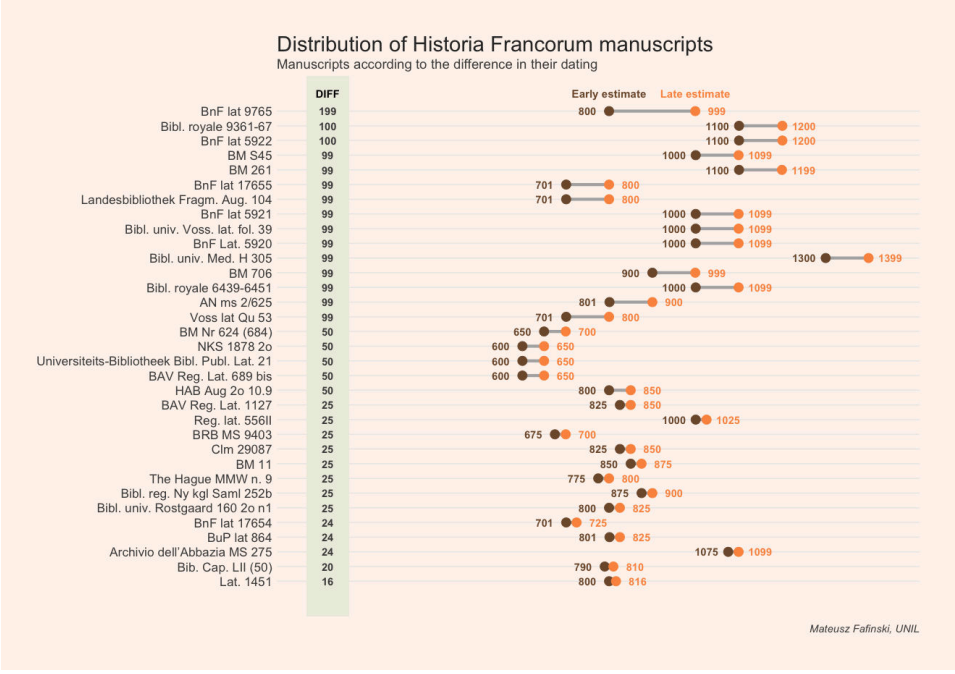


Fig. 1: Surviving manuscripts of *Historia Francorum* and their dating

We recognise this conundrum and see the need to assign different methods of mapping to different types of uncertainty in historical texts. And thus, while individual locales of low uncertainty can be assigned points, those of a higher degree need to be presented through polygons and those exhibiting a high degree of uncertainty across the three domains (1. Uncertainty about a primary source author's knowledge of a locale position; 2. Uncertainty about a scholar's understanding of a primary source's reference to a locale's position; 3. Uncertainty how much a single point can stand for the area(s) represented by a locale name) need to be visualised using fuzzy methods. Those problems are visible not only in case of the in-text data but also outside of it, as exemplified by the manuscript transmission of Gregory of Tours's main work, *Historia Francorum* (fig. 1).

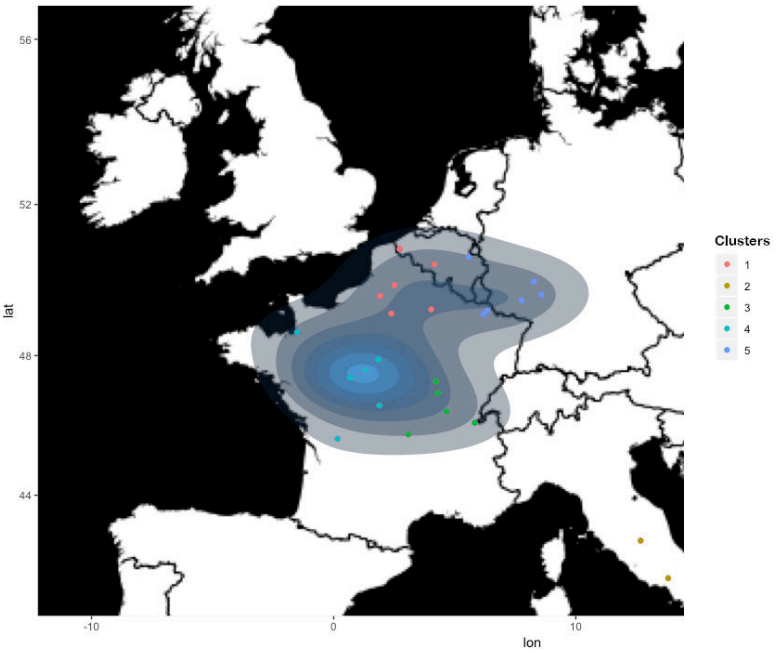


Fig. 2: Map of production locations of surviving manuscripts of *Historia Francorum*

The dating of various manuscripts as well their assignment to a particular space reflects a historiographical tradition that is characterised by a very high degree of uncertainty. While palaeographical dating and localising remains the basic method of work with those manuscripts, and is characterised by taking into account a high degree of uncertainty, both the current predominance of digital facsimiles [Te10] and the inherent lack of ability to accommodate fuzzy dating in catalog metadata [Da19] make it difficult to include this uncertainty in current digital projects. Moreover, a lack of precise uncertainty taxonomy makes comparisons between those projects difficult, if not misleading: the understanding, for example, what degree of correspondence between terms like "Northern France" and "Northern Gaul" exists and what is their level of uncertainty is almost entirely lacking. We propose therefore, as a form of stop-gap solution and a stepping stone in modelling this particular form of historiographical uncertainty, to map the distribution of those manuscripts through k-means clustering and kernel density estimation. This method, based on the idea of dividing observations into clusters with the nearest mean as a centroid [Ma67] and the smoothing of data based on the bandwidth [Pa62], showcases one possible example of computationally representing uncertainty of historiographical and chronological data on a two-dimensional map (see fig. 2). It should be also noted that the use of fuzzy clustering (c-means) did not produce significant differences at this scale and with this bandwidth.

This map (fig. 2) is not so much a map of the provenance and dating of the manuscripts of Gregory of Tours's *Historia Francorum* (although one might interpret it as such) as it is a map of the historiographical uncertainty about their localisation and dates of production: a map of uncertainty, if you will. This is even more visible through the nature of bandwidth in kernel density estimation: the choice of value of this parameter is in itself laden with uncertainty. This observation is crucial in order to use such visualisations at all. Providing the correct context is an important step to make such maps usable. It has been noted by Drucker [Dr14] that while the methods underpinning the algorithms we use often lack contextualisation, it is the very goal of humanities to provide such context. We see the recognition of such visualisations as *visualisations of uncertainty* as an important step forward in this respect.

4 Moving Forward with Uncertainty

There are tangible gains from including uncertainty in our models. As we strive to go beyond the narrow application inside a singular case-study, we want to highlight how modelling uncertainty and operating within a theoretically-based taxonomy might prove to be one of the crucial contributions of *theoretical digital humanities* [Pi18] to computational humanities and to the historian's toolbox alike. In order for computational humanities to function as a self-defined and independent field it requires a robust theoretical and methodological framework of its own. When it comes to uncertainty, a robust taxonomy will allow for a creation of project-independent methodology. When it comes to mapping historical sources it will finally allow not only for a basis of comparison between projects but also for a distinction between mapping sources and mapping historiography, thus bringing the methodologies of computational humanities on the same page as the methodologies of history. Using a taxonomy of uncertainty might also help to fine-tune geotagging of historical sources. By modelling vagueness in and assigning the correct level of uncertainty, the most appropriate method of visualisation can be assigned to a locale. This method can supplement models based on fuzzy representation of spatial data in texts [BGP12].

5 Conclusions

A robust methodology for uncertainty is a necessity for computational humanities to advance as a field. Through factoring vagueness in and modelling it for our visualisations we can finally achieve a more stable common ground between various, currently methodologically disjoint, projects that constitute the field of computational humanities.

Acknowledgments

This work is supported by a Spark grant from the Swiss National Science Foundation (no. 190306) awarded to M.P.

Bibliography

[BGP12] Bordogna, G.; Ghisalberti, G.; Psaila, G.: Geographic Information Retrieval: Modeling Uncertainty of User's Context. Fuzzy Sets and Systems 196/, pp. 105–124, June 1, 2012, DOI: 10.1016/j.fss.2011.04.005.

[Bi14] Binder, F.; Entrup, B.; Schiller, I.; Lobin, H.: Uncertain about Uncertainty, Different Ways of Processing Fuzziness in Digital Humanities Data. In: Digital Humanities 2014 Conference Abstracts, Lausanne July 8–12, 2014. Alliance of Digital Humanities Organizations, pp. 95–98, 2014, URL: http://nbn-resolving.de/urn:nbn:de:bsz:mh39-76383.

[Bl11] Blau, A.: Uncertainity and the History of Ideas. History and Theory 50/3, pp. 358–372, 2011, DOI: 10.1111/j.1468-2303.2011.00590.x.

[Da19] Davis, L. F.: Manuscript Road Trip: Linked Data, Library Science, and Medieval Manuscripts, Dec. 2, 2019, URL: https://manuscriptroadtrip.wordpress.com/2019/12/02/manuscript-road-trip-linked-data-library-science-and-medieval-manuscripts/, visited on: 06/22/2020.

[Dr14] Drucker, J.: Graphesis: Visual Forms of Knowledge Production. Harvard University Press, 2014.

[Ed92] Edgington, D.: Validity, uncertainty and vagueness. Analysis 52/4, pp. 193–204, Oct. 1992, DOI: 10.1093/analys/52.4.193.

[Fa20] Fafinski, M.: Facsimile Narratives: Researching the Past in the Age of Digital Reproduction. Digital Scholarship in the Humanities submitted/, 2020.

[Go20] Goodchild, M. F.: How Well Do We Really Know the World? Uncertainty in GIScience. Journal of Spatial Information Science 2020/20, pp. 97–102, 2020, DOI: 10.5311/JOSIS.2020.20.664.

[Go89] Goffart, W.: From Historiae to Historia Francorum and Back Again: Aspects of the Textual History of Gregory of Tours. In: Rome's Fall and After. Hambledon, London, pp. 255–274, 1989.

[Gr11] Gray, V.: Thucydides' Source Citations: "It Is Said". The Classical Quarterly 61/1, pp. 75–90, 2011, DOI: 10.1017/S0009838810000418.

[Gr74] Gregory of Tours: The History of the Franks. Penguin, Harmondsworth, 1974.

[He16] Hewson, L.: Les incertitudes du traduire. French, Meta 61/1, pp. 12–28, 2016, DOI: 10.7202/1036980ar.

[He94] Heinzelmann, M.: Gregor von Tours (538–594): "zehn Bücher Geschichte", Historiographie und Gesellschaftskonzept im 6. Jahrhundert. Wissenschaftliche Buchgesellschaft, Darmstadt, 1994.

[HM91] Hewson, L.; Martin, J.: Redefining Translation: The Variational Approach. Routledge, 1991.

[KC15] Kerdjoudj, F.; Curé, O.: Evaluating Uncertainty in Textual Document. In: Uncertainty Reasoning for the Semantic Web. 11th International Workshop on Uncertainty Reasoning for the Semantic Web (URSW 2015), co-located with the 14th International Semantic Web Conference (ISWC 2015), Bethlehem, PA Oct. 12, 2015. 2015, URL: http://ceur-ws.org/Vol-1479/paper1.pdf, visited on: 08/05/2020.

[Ko77] Koselleck, R.: Standortbindung und Zeitlichkeit. Ein Beitrag zur historiographischen Erschließung der geschichtlichen Welt. In (Koselleck, R.; Mommsen, W. J.; Rüsen, J., eds.): Objektivität und Parteilichkeit in der Geschichtswissenschaft. dtv, München, pp. 17–46, 1977.

[Lo05] Longley, P. A.; Goodchild, M. F.; Maguire, D. J.; Rhind, D. W.: Geographic Information Systems and Science. Wiley, 2005.

[Ma67] MacQueen, J.: Some Methods for Classification and Analysis of Multivariate Observations. In. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics. The Regents of the University of California, 1967, URL: https://projecteuclid.org/euclid.bsmsp/1200512992, visited on: 08/20/2020.

[Ma97] Marincola, J.: Authority and Tradition in Ancient Historiography. Cambridge University Press, 1997.

[Pa62] Parzen, E.: On Estimation of a Probability Density Function and Mode. Annals of Mathematical Statistics 33/, pp. 1065–1076, Sept. 1962, DOI: 10.1214/aoms/1177704472, URL: https://projecteuclid.org/euclid.aoms/1177704472, visited on: 08/21/2020.

[Pi18] Piotrowski, M.: Digital Humanities: An Explication. In (Burghardt, M.; Müller-Birn, C., eds.): Proceedings of INF-DH 2018, Berlin Sept. 25, 2018. Gesellschaft für Informatik, 2018, DOI: 10.18420/infdh2018-07.

[Pi19] Piotrowski, M.: Accepting and Modeling Uncertainty. Zeitschrift für digitale Geisteswissenschaften/Sonderband 4 *Die Modellierung des Zweifels, Schlüsselideen und -konzepte zur graphbasierten Modellierung von Unsicherheiten*, ed. by Kuczera, A.; Wübbena, T.; Kollatz, T., 2019, DOI: 10.17175/sb004_006a, URL: http://www.zfdg.de/sb004_006.

[Pr08] Prescott, A.: The Imaging of Historical Documents. In (Greengrass, M.; Hughes, L. M., eds.): The Virtual Representation of the Past. Digital Research in the Arts and Humanities, Ashgate, Aldershot, pp. 7–22, 2008.

[Ra16] Ralle, I. H.: Maschinenlesbar – Menschenlesbar, Über die grundlegende Ausrichtung der Edition. Editio 30/1, pp. 144–156, 2016, DOI: 10.1515/editio-2016-0009.

[Re13] Reimitz, H.: Cultural Brokers of a Common Past, History, Identity and Ethnicity in Gregory of Tours and Chronicles of Fredegar. In (Pohl, W.; Heydemann, G., eds.): Strategies of Identification: Ethnicity and Religion in Early Medieval Europe. Cultural Encounters in Late Antiquity and the Middle Ages (CELAMA) 13, Brepols, Turnhout, pp. 257–301, 2013.

[Re15] Reimitz, H.: History, Frankish Identity and the Framing of Western Ethnicity, 550–850. Cambridge University Press, Cambridge, 2015.

[RPH13] Reuschel, A.-K.; Piatti, B.; Hurni, L.: Modelling Uncertain Geodata for the Literary Atlas of Europe. In (Kriz, K.; Cartwright, W.; Kinberger, M., eds.): Understanding Different Geographies. Lecture Notes in Geoinformation and Cartography, Springer, Berlin, Heidelberg, pp. 135–157, 2013, DOI: 10.1007/978-3-642-29770-0_11.

[Sk97] Skinner, Q.: Sir Geoffrey Elton and the Practice of History. Transactions of the Royal Historical Society 7/, pp. 301–316, 1997, DOI: 10.2307/3679282.

[Ta11] Tarte, S. M.: Digitizing the Act of Papyrological Interpretation, Negotiating Spurious Exactitude and Genuine Uncertainty. Literary and Linguistic Computing 26/3, pp. 349–358, 2011, DOI: 10.1093/llc/fqr015.

[Te10] Terras, M. M.: Artefacts and Errors, Acknowledging Issues of Representation in the Digital Imagining of Ancient Texts. In (Fischer, F.; Fritze, C.; Vogeler, G., eds.): Codicology and Palaeography in the Digital Age 2. BoD, Norderstedt, pp. 43–61, 2010, URL: http://kups.ub.uni-koeln.de/4337/, visited on: 08/05/2020.

[Th98] Thucydides: The Peloponnesian War. Hackett, Indianapolis, IN, 1998.

[To84] Topolski, J.: Metodologia historii. Państwowe Wydawnictwo Naukowe, Warszawa, 1984.

[To97] Topolski, J.: Narrare la storia. Nuovi principi di metodologia storica. Mondadori, Milano, 1997.

[Wh73] White, H.: Interpretation in History. New Literary History 4/2, pp. 281–314, 1973, DOI: 10.2307/468478.

R. Reussner, A. Koziolek, R. Heinrich (Hrsg.): INFORMATIK 2020,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2021 1327

Exploring the Use of the Pronoun *I* in German Academic Texts with Machine Learning

Melanie Andresen[1], Dagmar Knorr[2]

Abstract: The use of the pronoun *ich* ('I') in academic language is a source of constant debate and a frequent cause of insecurity for students. We explore manually annotated instances of *I* from a German learner corpus. Using machine learning techniques, we investigate to what extent it is possible to automatically distinguish between different types of *I* usage (author *I* vs. narrator *I*). We additionally inspect which context words are good indicators of one type or the other. The results show that an automatic classification is not straightforward, but the distinctive features are in line with previous research. The results of the automatic classification are not perfect, but would greatly facilitate manual annotation. The distinctive words are in line with previous research and indicate that the author *I* is a more homogeneous class.

Keywords: annotation; academic language; German; machine learning; classification

1 Introduction

We present an exploratory study about the use of *ich* ('I')[3] in German academic texts by students. The main focus is on a quantitative exploration of different types of *I* using machine learning techniques, namely principal component analysis (PCA) and classification with a support vector machine (SVM). The task can be roughly understood as a case of word-sense disambiguation, even though the types of *I* differ functionally rather than semantically. Our aim is not to achieve full automation, but to deepen the understanding of the use of *I* from a humanities point of view by modeling its uses quantitatively.

The use of references to the author in academic language, most often realized by the pronoun *I*, is a source of constant debate and a frequent cause of insecurity for students. Authorial identity and self-reference in academic writing have therefore been a popular research topic ([Hy05], [Äd06], [Kr12]). We want to highlight two typologies of first person references that are suitable for empirical application: [TJ99] examine English academic essays written by students and suggest six types of first person references that form a continuum from low to high authorial power: The first two types, a) *I* as the representative (of the general public or the discourse community), and b) *I* as the guide through the essay, are low in

[1] Universität Stuttgart, Institut für Maschinelle Sprachverarbeitung, Pfaffenwaldring 5b, 70569 Stuttgart, Germany, melanie.andresen@ims.uni-stuttgart.de
[2] Leuphana Universität Lüneburg, Schreibzentrum/Writing Center, Universitätsallee 1, 21335 Lüneburg, Germany, dagmar.knorr@leuphana.de
[3] For brevity, we will subsequently refer to the target word as *I*, even though we always mean the German *ich*.

doi:10.18420/inf2020_124

authorial power and mostly realized by *we* rather than *I*. The more powerful types include c) *I* as the architect of the essay, d) *I* as the recounter of the research process, e) *I* as the opinion-holder, and f) *I* as the originator.

Steinhoff [St07] explores the use of *I* in a corpus of German research articles and distinguishes between 1) the author *I* that comments on the text and guides the reader, corresponding to type c) by [TJ99] (*In the following chapter, I will present my results about ducks*), and 2) the researcher *I* refers to the research object, terminology, or claims by other researchers (*I use the term 'duck' as referring to the waterbird*), roughly summarizing d), e), and f) by [TJ99]. He further includes 3) the narrator *I* that gives subjective, often auto-biographic information (*I originally wanted to write about birds, but then I learned there are so many different kinds of birds*). While the author *I* and the researcher *I* are considered acceptable in academic writing, the narrator *I* is mostly deemed inappropriate.We decided to apply the simpler typology by Steinhoff to our data. To the best of our knowledge, there has been no empirical application of the model beyond the original study.

2 Data

Corpus. The texts we use for our experiments are taken from the learner corpus *KoLaS* ('Kommentiertes Lernendenkorpus akademisches Schreiben', [AK17]). *KoLaS* comprises of academic texts written by students that visited the *Writing Center Multilingualism* at the Universität Hamburg between 2011 and 2016. The corpus is very diverse with respect to text types, disciplines, language skills, and the progression in the writing process. For many texts, several versions and comments by writing tutors are available. For this study, we reduced the corpus to the first version of every text and excluded text types with a focus on personal reflection. This resulted in a corpus size of 330 texts. We use a learner corpus, because we expect to find a substantial number of the narrator *I* not common in published academic texts. Doing so, however, can lead to the possibility of there being language errors and it affecting the results.

Annotation. Four annotators classified the instances of *I* in the corpus. The annotators were students trained as writing tutors. Before partaking in the annotation, they participated in a workshop about Steinhoff's *I* types where they reflected on textual indicators for these categories (see [AK17] for a description of the workshop concept). In order to cover all instances of *I* in the corpus, we extended the tagset with categories for *I* in example sentences, *I* referring to the general public, and *I* in comments by writing tutors.

Data Extraction. Since the annotation project was originally aimed at a qualitative analysis, the annotation was carried out with the support of the *MAXQDA* tool[4]. Unfortunately, this

[4] VERBI Software, `https://www.maxqda.de/`.

tool does not offer a direct export of annotated spans with their context. While a large part of the data could be extracted from the database format *MAXQDA* offers, not all annotations could be extracted for technical reasons.[5] In total, 2784 instances of *I* could be extracted. For this analysis, we were only interested in the three types of *I* by Steinhoff. We therefore filtered the data set for those instances, where at least two annotators agreed on one of those three categories. The resulting data set (n=360) comprises of 213 instances of the author *I* and 122 instances of the narrator *I*. As there were only 25 instances of the researcher *I*, we decided to exclude this type from the analysis. The data set is publicly available at Zenodo.[6]

Inter-Annotator Agreement. The inter-annotator agreement was calculated using Krippendorff's alpha [Kr80]. The agreement on the full data set (n=2784) that could be extracted from the annotation files is 0.76, which is a substantial agreement, following the (rather generous) scale by [LK77]. However, the three types by Steinhoff are among the more difficult categories. The agreement on the data set filtered for these categories (n=360) drops to 0.56 (moderate agreement). This indicates that further refinement of the guidelines could be beneficial. However, we can reasonably assume that the phenomenon as such shows ambiguities that cannot always be resolved. Under these conditions, we cannot expect excellent classification results.

Feature Extraction. In our experiments, the frequency of words in the immediate context before and after the *I* served as features. We did experiments based on words in a context window of three, five, and seven words left and right of the *I*. The smallest context of three words turned out to be the most helpful. Larger context windows led to an increased impact of idiosyncrasies of individual texts. To further reduce this effect, we exclude all word forms that occur only once. This results in a vocabulary of 183 words, whose frequencies serve as our features.

3 Unsupervised Experiment: PCA

As an initial exploration of the data, we use an unsupervised experiment[7] to learn about patterns emerging from the data without enforcing our ideas about the typology. For this purpose, we use a principal components analysis (PCA). The PCA takes the original data set with many variables or dimensions (one for each context word type in our data) as input and transforms them into new dimensions that capture as much variation in the data as possible.

Figure 1 shows the distribution of all instances of the author *I* and the narrator *I* across the first two dimensions of the PCA. While the samples of the two classes overlap, there

[5] More specifically, it was not always possible to unambiguously identify the position of the annotated span in the full text.

[6] https://doi.org/10.5281/zenodo.3999304

[7] For the implementation, we use *Python 3* with *pandas* [Th20], *scikit-learn* [Pe11], and *matplotlib* [Hu07].

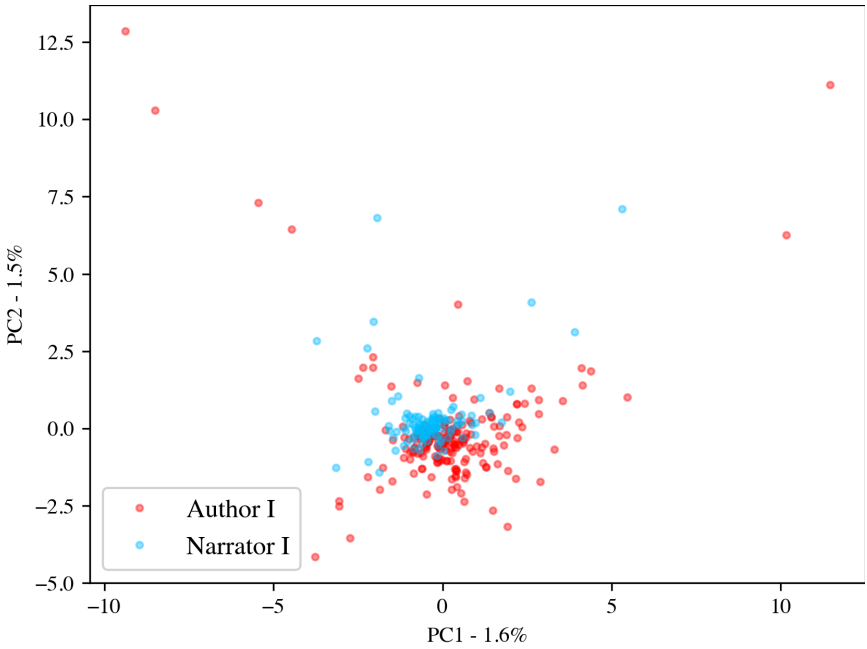


Fig. 1: Distribution of the author *I* and the narrator *I* in the first two dimensions of the PCA

is a tendency for the narrator *I* to score higher than the author *I* in the second dimension of the PCA. Notably, the variance in the data set that is explained by this dimension is rather low (1.5%). We conclude that a clear distinction between the types will not be possible. Nevertheless, the word frequency data analyzed using the PCA clearly contain some information that is related to the distinction of *I* types. The next section examines this information in greater detail.

4 Supervised Experiment: SVM

In addition to the PCA, we present a supervised experiment using a linear support vector machine. This type of classifier allows us to inspect the contribution of the individual features to the classification. These are important for our interpretation beyond classification scores.

Classification Success. Table 1 shows the results of a 5-fold crossvalidation. We include two baselines: One for a classifier that makes random choices and one that always votes for the majority class, the author *I*. We report precision, recall and f1-scores for the two classes together with a macro average and a weighted average that takes the number of instances

		Precision	Recall	F1-Score
SVM	Narrator *I*	0.72	0.77	0.74
	Author *I*	0.86	0.83	0.84
	Mean (macro)	**0.79**	**0.80**	**0.79**
	Mean (weighted)	**0.81**	**0.81**	**0.81**
Baseline (random)	Narrator *I*	0.50	0.50	0.50
	Author *I*	0.50	0.50	0.50
	Mean (both)	**0.50**	**0.50**	**0.50**
Baseline (majority class)	Narrator *I*	0.00	0.00	0.00
	Author *I*	0.62	1.00	0.77
	Mean (macro)	**0.31**	**0.50**	**0.39**
	Mean (weighted)	**0.39**	**0.62**	**0.48**

Tab. 1: Classification results for the SVM and two baseline models (random choice and majority class)

per class into account. The weighted means are always (equal or) higher as they give more weight to the majority class which also scores higher.

Our classifier clearly outperforms both baseline models in all metrics—with the obvious exception of recall for the majority class in the majority class baseline. For the author *I*, we achieve very good scores with a precision of 0.86 and a recall of 0.83. Both metrics are lower for the narrator *I*. This might be due to the fact that we have less examples of narrator *I*s in our data. To sum all scores up, our classifier achieves a mean f1-score of 0.79. This does not allow for full automation but is a very good result that could, for instance, serve as a useful pre-analysis to facilitate manual classification.

Features. In Figure 2, we can see the features that contributed most to the distinction between the two classes based on the coefficients of the SVM. Words on the left are indicators of the narrator *I* and words on the right are indicators of the author *I*. The best indicators for the author *I* have higher coefficients,[8] i. e. they are more helpful for the classifier. The words can be clearly interpreted with respect to the function of the author *I*: the verbs *werde* ('will'), *möchte* ('want to'), and *kann* ('can') are used to cataphorically announce what follows in the text. The word *Arbeit* ('work') is frequently used to refer to the text (analogous to *In this paper, I will. . .*). The indicators for the narrator *I*, on the other hand, have slightly lower coefficients and their interpretation is less straightforward. This can be due to the smaller sample for this type. Another explanation is that while the author *I* is limited to a rather fixed set of possible expressions, the narrator *I* has more freedom in wording as well as topic.

[8] Scores can be read from the side of the words that faces the center of the plot.

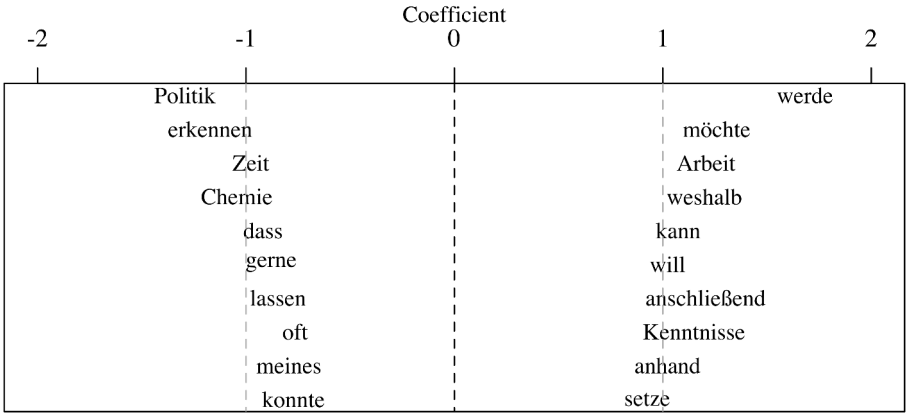


Fig. 2: Distinctive features for narrator *I* (left) and author *I* (right) based on SVM coefficients

5 Discussion and Future Work

Our exploration of *I* types in German academic language with machine learning shows promising tendencies, though the classification results are far from perfect. We consider the results highly beneficial for the understanding of the phenomenon from a humanities perspective. In particular, the features of the SVM allow for interpretations that build on existing research. One important result is that the author *I* appears to be a more homogeneous class than the narrator *I*. This is suggested by the higher scores in the classification, the higher SVM coefficients for features indicating the author *I*, and a more straightforward interpretability of these features. This result indicates that the author *I* type is restricted to a very specific function and a limited number of possible realizations at the text surface.

In order to obtain more stable and generalizable results, more data would be beneficial. This might also allow for the inclusion of the third type of *I*, the researcher *I*. A potential problem of the utilized data set is that it includes more than one instance of *I* per text. Consequently, these data points are not independent and their shared topic can have distorting effects. A larger data set would reduce this effect or allow us to include only one *I* per text. However, we have to take into account that the categories are, to some extend, ambiguous, as reflected in the inter-annotator agreement. Refinement of the guidelines might result in some improvement, but there is an upper limit as to what can be meaningfully disambiguated.

In the future, we intend to broaden our database by annotating more academic data. The inclusion of reviewed and published academic texts could be helpful in identifying the effects of limited language skills in our learner corpus. The narrator *I* indicates that a comparison to literary data could be beneficial: Do the narrator *I* type in academic texts and the narrator *I* type in literary texts have commonalities? In terms of features, we would like to refine our approach by using linguistic annotations and including, for instance, verbal morphology, which is known to be a good indicator for narration.

6 Acknowledgements

Melanie Andresen's work on this paper was funded by the *Landesforschungsförderung Hamburg* in the context of the project *hermA* [Ga17] (LFF-FV 35) at Universität Hamburg.

References

[Äd06] Ädel, A.: Metadiscourse in L1 and L2 English. Benjamins, Amsterdam, 2006.

[AK17] Andresen, M.; Knorr, D.: KoLaS – Ein Lernendenkorpus in der Schreibberatungsausbildung einsetzen. Zeitschrift Schreiben/, pp. 10–17, 2017, URL: https://zeitschrift-schreiben.ch/2017/#andresen.

[Ga17] Gaidys, U.; Gius, E.; Jarchow, M.; Koch, G.; Menzel, W.; Orth, D.; Zinsmeister, H.: hermA: Automated Modelling of Hermeneutic Processes. Hamburger Journal für Kulturanthropologie/7, pp. 119–123, 2017.

[Hu07] Hunter, J. D.: Matplotlib: A 2D Graphics Environment. Computing in Science & Engineering 9/3, pp. 90–95, 2007, DOI: 10.1109/MCSE.2007.55.

[Hy05] Hyland, K.: Metadiscourse: Exploring Interaction in Writing. Continuum, London, 2005.

[Kr12] Kruse, O.: Wissenschaftliches Schreiben mehrsprachig unterrichten: Was ist möglich, was ist nötig? ÖDaF-Mitteilungen/2, pp. 9–25, 2012.

[Kr80] Krippendorff, K.: Content Analysis: An Introduction to Its Methodology. Sage, Beverly Hills, California, 1980.

[LK77] Landis, J. R.; Koch, G. G.: The Measurement of Observer Agreement for Categorical Data. Biometrics 33/1, pp. 159–174, 1977.

[Pe11] Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, E.: Scikit-Learn: Machine Learning in Python. Journal of Machine Learning Research 12/, pp. 2825–2830, 2011.

[St07] Steinhoff, T.: Zum ich-Gebrauch in Wissenschaftstexten. Zeitschrift für germanistische Linguistik 35/1-2, pp. 1–26, 2007.

[Th20] The pandas development team: Pandas 1.0.3, Mar. 18, 2020, DOI: 10.5281/zenodo.3715232.

[TJ99] Tang, R.; John, S.: The 'I' in Identity: Exploring Writer Identity in Student Academic Writing through the First Person Pronoun. English for Specific Purposes 18, Supplement 1/, S23–S39, 1999, DOI: 10.1016/S0889-4906(99)00009-5.

R. Reussner, A. Koziolek, R. Heinrich (Hrsg.): INFORMATIK 2020,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2021 1335

Exploring the content composition of online book reviews

Kristin Kutzner[1] Thorsten Schoormann[2] Ralf Knackstedt[3]

Abstract: Today, anyone can perform an opinion-expressing form of literary criticism by writing online book reviews. Sellers and publishers recognised the strategic potential of such reviews, for example, to increase sales. However, despite the popularity and recognised importance of book reviews, only little is known about the actual content in detail. Drawing on a category system and manually annotated reviews, this study explores the content composition of book reviews. We disclose frequently used content-related book review components and perform a cluster analysis, exploring which components often occur together. Our results support literary scholars in investigating the digital phenomenon of literary criticism and the study illustrates a sample Computational Humanities project which can be transferred to other research endeavours.

Keywords: Literary Criticism; Culture; Content Analysis; Archetypes

1 Introduction

«*If literary criticism has a future, it is on the web*» (translation of Wolfram Schütte, publicist, 2015). Especially the booming digitalisation has enormous effects on creative industries and offers novel forms of collaboration and participation in cultural practices. As an example, there is a growing interest in online platforms such as LovelyBooks which foster a sociable and collaborative mentality, enabling new formats for anyone to discuss and share opinions on books in a community [KPK19]. Writing an online book review can be described as an opinion-expressing form of literary criticism [LHM13], and is therefore related to the tradition of literary criticism as a professional journalistic form of reviewing and discussing new publications [St97]. In consequence, the formerly clear separation of laypersons and professionals is blurring, and, as suggested by Wolfram Schütte, a user-generated form of literary criticism takes place online [KM17]. Consumers generally trust online reviews as a source of brand information [Ni12] and use reviews for purchase decisions [Dr13]. To gain useful information, the consumers care about the review content characteristics [Wi11]. For example, they are interested in book recommendations or opinions regarding the authors' language style, to consider whether the book might be suitable for themselves. Likewise, sellers and publishers recognized book reviews as strategic instruments to improve book

[1] Universität Hildesheim, Informationssysteme und Unternehmensmodellierung, Universitätsplatz 1, 31141 Hildesheim, kristin.kutzner@uni-hildesheim.de
[2] Universität Hildesheim, Informationssysteme und Unternehmensmodellierung, Universitätsplatz 1, 31141 Hildesheim, thorsten.schoormann@uni-hildesheim.de
[3] Universität Hildesheim, Informationssysteme und Unternehmensmodellierung, Universitätsplatz 1, 31141 Hildesheim, ralf.knackstedt@uni-hildesheim.de

doi:10.18420/inf2020_125

visibility, recommendations and sales [Ch08]. To enhance future book production and sales strategies based on the consumer needs, they need to know, for instance, whether their consumers are completely satisfied or whether they express deficiencies like a boring story line, annoying protagonist or cheap thin paper. Although online book reviews are of high importance for various stakeholders, only little is known about the content of such reviews in detail [KM17]. This is problematic as it hinders, for example, literary scholars in redefining the identity and functions of literary criticism in the digital age [KM17], and sellers in understanding the people's opinions on books which might help improve their business [KCR18]. As prior studies focused on general review characteristics like numerical ratings [CM06, Su11] or review sentiments [Sr18], there is a need for understanding the content composition of book reviews. By seeking to answer the following research questions (RQ), we are, to the best of our knowledge, one of the first gaining detailed insights into this:

RQ1: *What kind of content-related components are expressed in online book reviews?*

RQ2: *What archetypes of online book reviews can be identified?*

To answer these questions, based on a large data set of online book reviews, we use a combination of a manual and an automated approach in order to get knowledge from the text [LZH13]: We combine manual text analysis and computer-aided analysis to (1) examine frequently used components of book reviews and to (2) perform a cluster analysis to identify review archetypes. This content-driven analysis supports, for example, literary scholars in investigating the digital phenomenon of literary criticism and in building an enhanced understanding of digital culture and society. For online platform users, our results indicate a variety of common ideas that can be addressed in further book reviews which might contribute to diverse book discussions. Sellers and publishers get deep insights into discussed book components and review archetypes which might help to improve future book productions and sales. From a technical perspective, we provide an alternative to counteract the lack of proper text analysis tools for analysing detailed review content [CM06] by using a combined method approach. Overall, this study illustrates a sample case in which researchers from Computer Science (information systems, computational linguistics) and Humanities (literary studies, cultural politics) are cooperating to investigate a digital cultural phenomenon, which could be transferred to further Computational Humanities projects.

2 Research background

The term *book review* can be defined as an opinion-expressing form of literary criticism [LHM13], and it means asking for terms of arts, their functions and their origin [Ra07]. Characteristic components of reviews include descriptions, explanations, interpretations, recommendation/dissuasion and/or evaluations of cultural artefacts [St97]. Thus, a review is characterised by plenty of different text segments, the so-called *components*. Analysing book reviews, researchers investigated directly observable determinants such as star ratings [CM06, Su11] and review length [CM06, Ku15], and suggested a contribution to review

helpfulness and sales. For instance, the results of analysing Amazon book reviews indicated that the average rating tends to be positive [CM06, HPZ09]. To extract whether book review texts are positive or negative, sentiment analysis [Sr18] has been studied, again, suggesting a predominance of positive sentiments. Referring to literary criticism, through manually analysing book reviews, [KM17] examined categories of everyday communication and professional literary criticism, [Ba15] analysed informative and evaluative statements in laypersons book reviews, and [St15] explicated typical review characteristics. Nonetheless, previous studies are restricted as they focus on limited review characteristics (e.g., star rating, sentiment, evaluative statements) and as they select either a computer-aided or manual text analysis approach. However, to derive meaningful knowledge from text data, the combination of both approaches is suggested [LZH13]. We contribute to this research by investigating a multitude of book review characteristics (i.e., components) and by combining manual text and computer-aided analysis as it seems worthwhile for gaining detailed insights into content-related features of book reviews, based on a large data set.

3 Research design

We conducted a three-staged research design (Fig. 1). To obtain frequently used review components, we iteratively developed a category system (Stage 1) and applied it for annotating book reviews (Stage 2). Based on this, we performed a cluster analysis to explore which of these components often occur together (i.e., archetypes) (Stage 3).

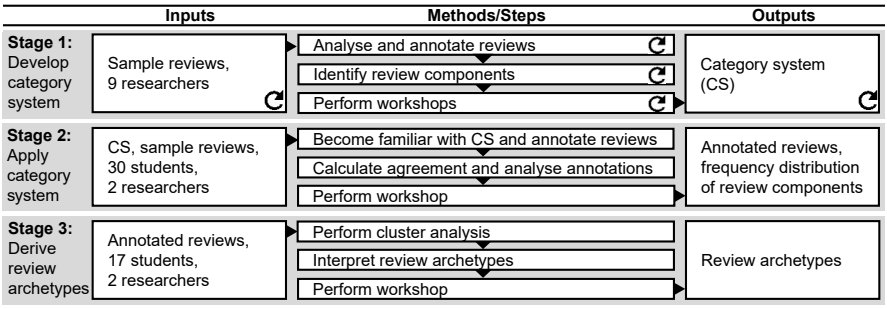


Fig. 1: Research design

Stage 1. First, we selected ten reviews from different platform types (e.g., social media, other rating and exchange platforms), addressing artistic artefacts and books (as our overriding goal is to analyse reviews of cultural artefacts in general across platforms). Nine researchers from Computer Science and Humanities independently analysed the reviews, named text segments with short labels that characterise the review components, and consolidated their components in a workshop. Second, the researchers independently structured and reassembled the identified components, and built a category system by consolidating their results in a workshop. Third, we selected further reviews and the researchers independently applied the system, annotating review text segments with the components. They compared

their experiences while annotating reviews and modified the system. These steps were repeated several times until the saturation of the system was perceived by the researchers [for more details see Ku18].

Stage 2. After becoming familiar with the system, 30 German students annotated review text segments with the help of our system. The completion of the task was rewarded with bonus points in a course. We randomly selected 430 German Amazon book reviews (e.g., novel, textbook) [HM16, Mc15]. Due to availability for scientific research, the large database, and the variety of products (different books, authors) and reviews, we decided to use Amazon reviews. To measure the nominal scale agreement (Fleiss kappa) among different annotators [Fl71], each review has been assigned to three students. However, not all reviews have been annotated threefold because only 24 of 30 expected students completed the task. Besides, sometimes, not all text segments have been fully annotated with components. Therefore, we got rather low kappa values. To ensure the inter-annotator agreement, we passed back the annotated reviews to small student groups who discussed and, if deemed necessary by the majority, adapted their annotations. To support focused discussions, we provided a subset of reviews and suggested, if not present, additional annotations. Finally, we got 282 agreed and annotated reviews.

Stage 3. To derive archetypes, we performed a cluster analysis which is an established analytical tool for investigating correlations in datasets [BMI11]. We utilised 282 reviews as objects and the annotated components of the category system as clustering variables, and we applied a two-step approach [PS83, Re16], using the python module sckit-learn [Pe11]. (1) To identify the number of clusters, we performed the *ward's method* which forms hierarchical clusters of subsets on the foundation of their similarity. Initially, the method combines two closest subsets into one cluster. This step is repeated until all subsets are in one cluster [Wa63]. The number of identical components of the category system determined the similarity between the two subsets. To follow the sequence in which the subsets have been united in relation to the distances, we plotted a *dendrogram*. Regarding the significant jumps in the distance of the joint clusters, we identified four or eight clusters as useful. (2) We applied the *K-means algorithm*, one of the most common clustering methods [El03], for the four and the eight cluster solution. The method divides data into clusters, minimising the within-cluster sum of squares [HW79]. We used k-means++ to select initial cluster centres, the algorithm iterated 300 times, and within each iteration it ran with ten different centroid seeds to get the best results. Thereafter, we manually evaluated the resulting clusters for their explanatory power and chose eight clusters, to get deep insights into the nature of the review archetypes. Next, five groups of students and two researchers independently interpreted the clusters. They analysed the most common components of each cluster and compared the clusters to find characteristic differences and similarities. Besides, they randomly traced back assigned reviews of each cluster, to check and, if necessary, adapt their interpretations. Finally, the results were consolidated in a workshop, resulting in eight interpreted archetypes of book reviews.

4 Content of online book reviews

Annotation example. 24 students annotated review text segments (text segments can be multiple sentences, a sequence of terms or a single term) with the components of the category system. Text segments can be assigned to more than one component. To illustrate how components are attached to review text segments, we present an annotation example (Fig. 2). Reading the sample review "I can't wait for the last book to come. The book is highly recommended.", we can identify and annotate components of our category system. First, the reviewer shows his/her feelings (relation to own emotions) and addresses another book of the same author (addressing other artefacts of the same author), writing that he/she can't wait for the last book to come. At the same time, the reviewer positively assesses the artefact. Second, the reviewer recommends the book in general (view of the artefact as a whole), without detailed explanations.

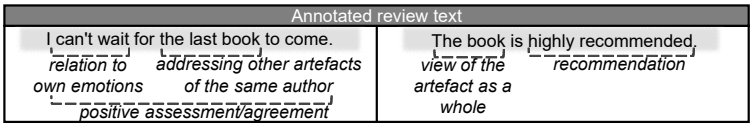


Fig. 2: Annotation example of a book review (translated from German)

Frequency analysis. The students analysed the reviews, using 65 components of the category system. In sum, in 282 reviews 5381 text segments have been manually identified and annotated with components. Some of them have been more frequently used than others. For reasons of space limitations, we present the ten most common and the ten least recognised components (Fig. 3).

	Review components	abs. freq.	rel. freq.		Review components	abs. freq.	rel. freq.
1	positive assessment/agreement	905	16,82%	56	relation to other reviews	6	0,11%
2	view of the artefact as a whole	519	9,65%	57	history of the artefact	5	0,09%
3	content – fictional character	296	5,50%	58	detailed relation to other artefacts	5	0,09%
4	content – story line	293	5,45%	59	history of publisher	4	0,07%
5	summary	290	5,39%	60	literary-historical epoch	2	0,04%
6	relation to own emotions	221	4,11%	61	goal/task of review(ing process)	2	0,04%
7	negative assessment/disagreement	193	3,59%	62	structure of the review	1	0,02%
8	language style	187	3,48%	63	introduction of your own person	1	0,02%
9	classification/interpretation	174	3,23%	64	ISBN information	0	0,00%
10	mention without assessment	158	2,94%	65	technical comments	0	0,00%
...	...	...	...%		∑	5381	100%

Fig. 3: Distribution of components of the category system for book reviews

Regarding the frequency distribution, *very often*, reviewers express their positive assessment/agreement, discuss the artefact (i.e., the book) as a whole without quoting detailed information (the second sentence of the annotation example illustrates such an observation, Fig. 2) and also address individual content-related aspects of the artefact (i.e., addressing the characters or story line of the book). Besides, reviewers summarise contents, express their own emotions, assess negatively and discuss the language style of the book. Moreover, interpretations and mention of the content without assessment are recognised. *Very rarely*, reviewers relate to other reviews (e.g., responding to other reviews) or address the history

of the artefact (i.e., the context of book emergence or publication). In addition, books are rarely discussed on a detailed level in terms of other artefacts (i.e., discussion with a concrete reference to certain book contents). Besides, the publisher or its backgrounds and literary-historical epochs are seldom quoted in reviews. Moreover, reviewers rarely address the general goal or task of writing a review (i.e., reflections on the review text itself) and the review structure (e.g., use of headlines or paragraph). Very rarely, reviewers introduce themselves, address ISBN information or leave technical comments related to the platform (e.g., ease of use, format requirements).

Archetypes. We identified eight clusters of online book reviews, each comprises between 17 and 73 reviews and has a different focal point along the components of the category system. As the components of each cluster are collectively exhaustive, the results can be read as percentages, for example: 18,39% of the reviews of Cluster 1 contain positive assessments and 5,84% address the character of the artefact (Fig. 4). The darker the colour of a cell, and the percentage of a component, the more it is shaping a cluster. For reasons of presentation, we illustrate the topmost shaping components of the clusters. Thus, we consolidate 46 components that are not characteristic for one of the clusters and therefore have low percentages ("other 46 (of 65) components").

		Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8
Number of reviews per cluster		73	17	56	29	33	31	24	19
Components of the category system	positive assessment/agreement	18,39%	18,12%	19,62%	12,71%	17,70%	16,06%	14,77%	13,28%
	view of the artefact as a whole	10,95%	8,09%	7,53%	11,30%	14,40%	8,72%	8,44%	6,94%
	content - fictional character	5,84%	2,59%	3,63%	5,78%	1,65%	3,07%	7,17%	9,02%
	content - story line	4,00%	1,29%	4,57%	5,78%	2,06%	3,77%	7,17%	9,91%
	summary	4,24%	2,59%	3,76%	3,59%	3,70%	5,75%	6,75%	8,92%
	relation to own emotions	3,60%	4,53%	5,11%	3,08%	2,47%	5,15%	2,53%	3,57%
	negative assessment/disagreement	3,20%	2,59%	4,17%	3,47%	4,94%	1,78%	4,22%	4,66%
	language style	3,76%	2,91%	3,76%	2,95%	1,65%	3,07%	1,69%	4,06%
	classification/interpretation	3,28%	5,83%	3,76%	3,72%	3,29%	2,68%	3,38%	1,49%
	mention without assessment	1,44%	3,56%	1,61%	3,72%	2,06%	4,86%	1,69%	2,97%
	content in general	1,68%	3,24%	2,82%	2,05%	2,88%	4,76%	1,27%	3,07%
	physical properties of the artefact	1,44%	2,27%	1,48%	0,77%	6,17%	4,56%	3,38%	1,09%
	author	2,88%	0,97%	2,82%	2,44%	0,00%	1,49%	2,11%	1,88%
	recommendation in general	2,08%	0,97%	1,08%	3,59%	2,06%	1,78%	2,53%	0,59%
	outer appearance	1,28%	4,21%	1,61%	0,77%	2,47%	1,98%	5,49%	0,89%
	representation of your own conviction	1,60%	1,62%	0,94%	1,80%	1,23%	2,08%	1,27%	1,49%
	history of provision	1,52%	2,27%	2,28%	0,51%	4,94%	0,79%	5,06%	0,30%
	recommendation for certain target group	0,80%	0,97%	1,08%	1,80%	4,12%	2,38%	2,11%	0,59%
	citation	0,72%	4,21%	1,75%	0,64%	0,00%	1,09%	0,00%	0,69%
	other 46 (of 65) components	27,34%	27,18%	26,61%	29,53%	22,22%	24,18%	18,99%	24,58%
	∑	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%

Caption: the darker the colour of a cell, the higher the percentage within a cluster

Fig. 4: Results of the cluster analysis

In the following, we present the interpreted clusters, utilising illustrative review cut-outs (translated from German to English) and highlighting the most typical components of each cluster: **Cluster 1**—*content summary and positive assessment of the artefact*. "I haven't read such a well thought-out, perfidious book for a long time [. . .] Awesome! [. . .] But first, let's talk about the writing style of Gillian Flynn. She changes between the perspective of Nick and Amy. Nick is exposed with all feelings, sensations and actions. Amy, on the

other hand, can only be seen from a diary perspective [. . .].” This review snippet illustrates that the reviewers of Cluster 1 positively assess the artefact as a whole, summarize detailed content information (i.e., characters and story line), address the language style of the book, and offer insights into their own emotions. **Cluster 2**—*emotional, positive interpretation of the artefact*. “The combination of human everyday life and fairy tales makes this series so fascinating. Where would Snow White work if she lived in our world? [. . .] This first volume mainly deals with the clarification of the murder of Rose Red [. . .].” Reviews of this cluster are characterised by emotional narrations and positive assessments, interpreting and scrutinising general content-related aspects of the artefact. Sometimes, reviewers not only refer to the content in general, but also to certain text passages or quotations. Besides, outer appearances of the artefact are discussed. **Cluster 3**—*emotional, critical interpretation of the artefact*. “Partly, the story line had some hanging parts, however, I still devoured the book within a few days. The author manages the development of the main character [. . .] well. This fascinated me.” As Cluster 2, the reviews of Cluster 3 are characterized by emotional narrations related to the artefact as a whole. Interpreting detailed content-related aspects (story line), the reviewers express themselves more critically, indicating both positive and negative assessments. Moreover, reviews of this cluster contain a content summary and address language style. **Cluster 4**—*recommendation of the artefact*. “[. . .] Hanna, the main protagonist, can sometimes be a little annoying [. . .]. But she balances it out with her refreshing personality. She just doesn’t leave everything behind as soon as she meets a man (even a man like Drew who is already a real treat). A great book, I can only recommend it [. . .].” Positively assessing the artefact as a whole, reviewers of this cluster recommend the artefact in general, without addressing a certain target group. Doing so, they summarise detailed content-related aspects (character, story line) as well as interpret and/or only mention them. **Cluster 5**—*critical assessment of physical properties and history of provision*. “I was one of the first pre-orders and I was really happy when the book arrived. After opening the package, I was really disappointed: slanting printing, broken binding, cheap thin paper. Of course, I immediately complained about it and five weeks later, I got a new book. Thank God, the binding and printing were ok.” This review is exemplary for Cluster 5, addressing physical properties of the artefact (e.g., quality of printing or paper) and explaining the history of provision of the artefact (e.g., delivery by mail). In general, the reviewers assess these aspects both positively and negatively. Moreover, reviews frequently contain recommendations for a certain target group. **Cluster 6**—*superficial summary and positive assessment*. “I love this book as much as I love the movie. The book gives nice insights into the work of an animation studio.” In contrast to Cluster 1, where reviews are characterised by detailed content summaries, reviews of Cluster 6 contain more superficial summaries or mentions of the artefact. As the sample review illustrates, the artefact as a whole is positively assessed, addressing the content in general. Furthermore, reviewers show their own emotions and sometimes address the physical properties of the artefact. **Cluster 7**—*critical assessment of outer and inner appearances*. “I ordered this book in order to improve my drawing skills after it had been recommended from various sources. First, I can say that I would recommend this book to anyone interested in drawing. In multiple chapters, Andrew Loomis gives hints for more credible characters. [. . .] The only criticism

is minor printing errors that sometimes leave black spots on the pages [. . .]." Regarding the characteristic components of Cluster 7, there are overlapping characteristics to Cluster 1 and 5. Reviews of Cluster 7 address both detailed content-related aspects and the history of the provision. Besides, they are characterised by addressing the artefact as a whole and its outer appearance. In general, these aspects are both positively and negatively assessed. **Cluster 8**—*content summary and criticism*. "I have to say that I really like the protagonist Quentin and I liked the first third of the book very much. Margo and Quentin drive around the city at night and Quentin helps Margo playing jokes. [. . .] After that, unfortunately, the story was dragged into the long run. Not much really happened [. . .]." Like Cluster 1, the reviews of Cluster 8 are also characterised by a detailed content summary. However, in contrast to Cluster 1, the reviewers are more critical of content-related aspects by assessing it both positively and negatively.

5 Discussion, implications and conclusion

We believe this study to be an important step in investigating online book reviews as a user-generated form of literary criticism. Applying the category system for Amazon book reviews, we identified that the ten most commonly annotated components represent about 60% of all 65 considered components. Consequently, only a small subset of components is usually addressed in our sample, which might indicate a strong focus on certain aspects in a review. Exploring review archetypes, we found that despite some kind of closeness between some clusters (e.g., positive assessment/agreement and view of the artefact as a whole are present in all clusters) there can be identified nuanced and distinct differences which provide deep insights into the nature of review archetypes.

Based on these archetypes, literary scholars are supported in investigating the digital, user-generated form of literary criticism. Hence, they can compare literary criticism of users (i.e., laypersons) and professionals, to identify the differences and commonalities of criticism in consequence of digitization. In addition, knowing which book characteristics are discussed by reviewers, platform providers can develop strategies enhancing customers' review participation behaviour. The formulation of templates for writing reviews, for instance, by requesting the reviewers to express their opinions on the language style of the book, might support platform providers in managing the review content to improve the desired review quality.

To conclude, this study determines frequently used content-related review components and derives eight archetypes of online book reviews. Overall, we hope that this work provides interesting insights into a Computational Humanities project and that it raises new discussions on the literary criticism field in a progressively digitalised world.

This research was conducted as part of the research project "Rez@Kultur" (01JKD1703) which is funded by the Bundesministerium für Bildung und Forschung (BMBF).

Bibliography

[Ba15] Bachmann-Stein, A.: Zur Praxis des Bewertens in Laienrezensionen. In (Kaulen, H., Gansel, C. eds.): Literaturkritik heute. Tendenzen – Traditionen – Vermittlung. V&R unipress, Göttingen, pp. 77–91, 2015.

[BMI11] Balijepally, V.G.; Mangalaraj, G.; Iyengar, K.: Are we Wielding this Hammer Correctly? A Reflective Review of the Application of Cluster Analysis in Information Systems Research. Journal of the Association for Information Systems 12(5)/11, pp. 375–413, 2011.

[Ch08] Chen, Y.-F.: Herd Behavior in Purchasing Books Online. Computers in Human Behavior 24/08, pp. 1977–1992, 2008.

[CM06] Chevalier, J.A.; Mayzlin, D.: The Effect of Word of Mouth on Sales: Online Book Reviews. Journal of Marketing Research 43(3)/06, pp. 345–354, 2006.

[Dr13] Drewnicki, N.: Survey: 90% Say Positive Reviews Impact Purchase Decisions. https://www.reviewpro.com/blog/survey-zendesk-mashable-dimensional-research-90-say-positive-reviews-impact-purchase-decisions/, accessed: 24/07/2019.

[El03] Elkan, C.: Using the Triangle Inequality to Accelerate k-Means. Proc. 20th Int. Conf. on Machine Learning, Washington DC, 2003.

[Fl71] Fleiss, J.L.: Measuring Nominal Scale Agreement among many Raters. Psychological Bulletin 76(5)/71, pp. 378–382, 1971.

[HM16] He, R.; McAuley, J.: Ups and downs: Modeling the Visual Evolution of Fashion Trends with one-class Collaborative Filtering. Proc. 25th Int. Conf. on world wide web International World Wide Web Conferences Steering Committee, pp. 507–517, 2016.

[HPZ09] Hu, N.; Pavlou, P.A.; Zhang, J.: Overcoming the J-shaped Distribution of Product Reviews. Communications of the ACM 52(10)/09, pp. 144–147, 2009.

[HW79] Hartigan, J.A.; Wong, M.A.: Algorithm AS 136: A k-Means Clustering Algorithm. Journal of the Royal Statistical Society. Series C 28(1)/79, pp. 101–108, 1979.

[KCR18] Kwark, Y.; Chen, Y.; Raghunathan, S.: User-Generated Content and Competing Firms' Product Design. Management Science 64(10)/18, pp. 4608–4628, 2018.

[KM17] Kellermann, H.; Mehling, G.: Laienrezensionen auf amazon.de im Spannungsfeld zwischen Alltagskommunikation und professioneller Literaturkritik. In (Bartl, A., Behmer, M. eds.): Die Rezension: aktuelle Tendenzen der Literaturkritik. Königshausen & Neumann, Würzburg, pp. 173–202, 2017.

[KPK19] Kutzner, K.; Petzold, K.; Knackstedt, R.: Characterising Social Reading Platforms – A Taxonomy-Based Approach to Structure the Field. Proc. 14th Int. Conf. on Wirtschaftsinformatik, Siegen 2019, pp. 676–690, 2019.

[Ku15] Kuan, K.K.Y.; Hui, K.-L.; Prasarnphanich, P.; Lai, H.-Y.: What Makes a Review Voted? An Empirical Investigation of Review Voting in Online Review Systems. Journal of the Association for Information Systems 16(1)/15, pp. 48–71, 2015.

[Ku18] Kutzner, K.; Moskvina, A.; Petzold, K.; Roßkopf, C.; Heid, U.; Knackstedt, R.: Reviews of Cultural Artefacts: Towards a Schema for their Annotation. Proc. Workshop on Annotation in Digital Humanities (annDH 2018) co-located with ESSLLI 2018, Sofia 2018. CEUR-WS, pp. 17–23, 2018.

[LHM13] La Roche, W.; Hooffacker, G.; Meier, K.: Einführung in den praktischen Journalismus: Mit genauer Beschreibung aller Ausbildungswege Deutschland Österreich Schweiz. Springer VS, Wiesbaden, 2013.

[LZH13] Lewis, S.C.; Zamith, R.; Hermida, A.: Content Analysis in an Era of Big Data: A Hybrid Approach to Computational Manual Methods. Journal of Broadcasting & Electronic Media 57(1)/13, pp. 34–52, 2013.

[Mc15] McAuley, J.; Targett, C.; Shi, Q.; Van Den Hengel, A.: Image-based Recommendations on Styles and Substitutes. Proc. 38th Int. ACM SIGIR Conf. on Research and Development in Information Retrieval. ACM, pp. 45–52, 2015.

[Ni12] Nielsen: Nielsen's latest Global Trust in Advertising Report. https://retelur.files.wordpress.com/2007/10/global-trust-in-advertising-2012.pdf, accessed: 30/04/2018.

[Pe11] Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; et. al.: Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research 12/11, pp. 2825–2830, 2011.

[PS83] Punj, G.; Steward, D.W.: Cluster Analysis in Marketing Research: Review and Suggestions for Application. Journal of Marketing Research, pp. 134–148, 1983.

[Ra07] Rauterberg, H.: Und das ist Kunst?!: Eine Qualitätsprüfung. S. Fischer, Frankfurt a. M., 2007.

[Re16] Remane, G.; Nickerson, R.C.; Hanelt, A.; Tesch, J.F.; Kolbe, L.M.: A Taxonomy of Carsharing Business Models. Proc. 37th Int. Conf. on Information Systems, Dublin, 2016.

[Sr18] Srujan, K.S.; Nikhil, S.S.; Raghav Rao, H.; Karthik, K.; Harish, B.S.; Keerthi Kumar, H.M.: Classification of Amazon Book Reviews Based on Sentiment Analysis. Information Systems Design and Intelligent Applications 672/18, pp. 401–411, 2018.

[St97] Stegert, G.: Die Rezension: Zur Beschreibung einer komplexen Textsorte. Beiträge zur Fremdsprachenvermittlung 31/97, pp. 89–110, 1997.

[St15] Stein, S.: Laienliteraturkritik – Charakteristika und Funktionen von Laienrezensionen im Literaturbetrieb. In (Kaulen, H., Gansel, C. eds.): Literaturkritik heute. Tendenzen – Traditionen – Vermittlung. V&R unipress, Göttingen, pp. 59–76, 2015.

[Su11] Sun, M.: How Does the Variance of Product Ratings Matter? Management Science 58(4)/11, pp. 696–707, 2011.

[Wa63] Ward, J.H. Jr.: Hierarchical Grouping to Optimize an Objective Function. Journal of the American Statistical Association 58(301)/63, pp. 236–244, 1963.

[Wi11] Willemsen, L.M.; Neijens, P.C.; Bronner, F.; Ridder, J.A.: „Highly Recommended!" The Content Characteristics and Perceived Usefulness of Online Consumer Reviews. Journal of Computer-Mediated Communication 17/11, pp. 19–38, 2011.

R. Reussner, A. Koziolek, R. Heinrich (Hrsg.): INFORMATIK 2020,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2021 1345

Towards End-to-End Deep Learning-based Writer Identification

Zhenghua Wang[1] Andreas Maier[1] Vincent Christlein[1]

Abstract: Writer identification is an important task to gain knowledge about life in the past, which is commonly solved by paleographic experts. In this work, we investigate an automatic writer identification procedure based on deep learning. So far, the most approaches are based on two or more different pipeline steps and only few of them can be trained in an end-to-end manner. In this paper, we propose a fully end-to-end deep learning-based model, which consists of a U-Net for binarization, a ResNet-50 for feature extraction, and an optimized learnable residual encoding layer to obtain global descriptors. We evaluate the proposed end-to-end model on the ICDAR17 competition dataset on historical document writer identification (Historical-WI) dataset. Moreover, we investigate the performance of our optimized encoding layer on three texture datasets. While the optimized encoding layer does not work well in the task of writer identification, it provides better performance on the texture datasets. Furthermore, we show that a pre-trained U-Net can improve the performance for writer identification.

Keywords: writer identification; writer retrieval; deep learning; end-to-end

1 Introduction

Writer identification aims to retrieve the writer of a query document image in a dataset. It is playing a more and more important role for history sciences and especially for paleography, where it can help to search through a large dataset. The typical scenario is to obtain a short list, e. g. 20 samples that are most similar to the query sample of the whole dataset. In this way, the respective scientist only needs to check this short list, because with a high probability, it contains the correct writer.

Typical retrieval methods make use of the penultimate layer of a trained Convolutional Neural Network (CNN). However, script has not an object-structure, thus common writer identification methods [Ch15; Ch17; KFS18] rely on local descriptors, which are nowadays often learned by a CNN. Overall, these methods consist of three main stages: (1) pre-processing, (2) local feature extraction to obtain local descriptors, and (3) global descriptor computation to obtain a global representation. The pre-processing stage aims at segmenting the text from image patches. Afterwards, local features for each image patch are extracted. Finally, in the encoding stage, a global feature descriptor is computed from all local descriptors of the document image. Until now, each stage of the proposed method is

[1] Friedrich-Alexander-Universität Erlangen-Nürnberg, Pattern Recognition Lab, Martensstr. 3, 91058 Erlangen, Germany, firstname.lastname@fau.de

doi:10.18420/inf2020_126

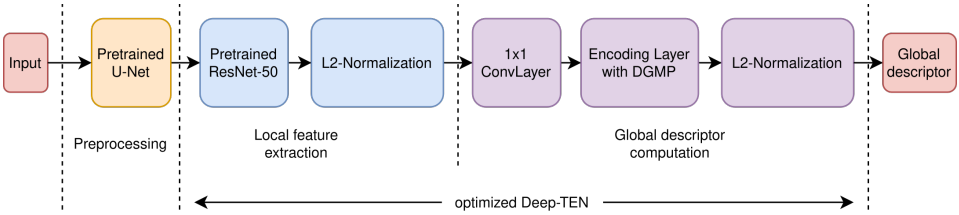


Fig. 1: Architecture of proposed end-to-end deep learning-based writer identification network.

optimized individually. In this paper, we aim to offer and evaluate a fully end-to-end deep learning-based model for writer identification.

In detail, our contributions are as follows: (1) We pre-train a U-Net [RFB15] on the Document Image Binarization Competition (DIBCO) datasets and transfer it to our end-to-end model and evaluate if a fine-tuning is beneficial. (2) We propose the use of the Deep-Ten method [ZXD17] as an encoding layer to form global descriptors. Additionally, we integrate Deep Generalized Max Pooling (DGMP) [Ch19], and evaluate it on both a writer identification dataset and texture datasets. (3) An end-to-end deep learning-based writer identification model is proposed and evaluated on the Historical-WI dataset. The end-to-end model is constructed using a pre-trained U-Net, residual network (ResNet) [He16] with 50 layers and an optimized encoding layer, see Fig. 1.

The rest of the paper is organized as follows. Sect. 2 gives an overview of the related work in encoding techniques and writer identification. The methodology of our end-to-end model is explained in Sect. 3. In Sect. 4, the evaluation protocol and results are shown. Finally, the conclusions are in Sect. 5.

2 Related Work

In this work, we focus on the group of codebook-based methods. Codebook-based methods create a global descriptor for each image by encoding local feature descriptors. Wu et.al [WTB14] apply SIFT [Lo99] to extract descriptors for word regions. In contrast, Christlein et al. [CBA15] calculate Contour-Zernike moments as local descriptors before aggregation using Vector of Locally Aggregated Descriptors (VLAD) [Jé12].

Nowadays, deep learning-based approaches are proposed for computing local feature descriptors. Fiel et al. [FS15] first propose a writer identification method based on document line segmentation and CNN activation features. In a concurrent work, Christlein et al. [Ch15] suggest to generate a global descriptor by computing Gaussian mixture model (GMM) supervectors to encode CNN activations. Christlein et al. [Ch17] also propose an unsupervised writer identification method. Keglevic et al. [KFS18] propose to train a DenseNet with triplet loss function [SKP15] to learn a similarity measurement between writers.

However, the pre-processing and global descriptor computation (encoding) stages of all above models are non-deep learning-based and optimized individually. In our work, we incorporate the binarization step into an end-to-end trainable network, by using a U-Net. A similar method was also proposed by Tensmeyer et al. [TM17], who adopt a fully convolutional network (FCN) [LSD15] for the binarization of handwriting images. As an extension of FCN, U-Net provides better segmentation performance with a small training dataset [RFB15]. For the encoding stage, Arandjelovic et al. [Ar16] introduce a new generalized VLAD layer, which is trainable on any CNN network. Zhang et al. [ZXD17] improves upon this approach by proposing a general residual encoding layer integrating the dictionary learning and residual encoding into a single learnable layer, called Deep Texture Encoding Network (Deep-TEN) model. Moreover, Christlein et al. [Ch19] suggest to apply Deep Generalized Max Pooling (DGMP) instead of global max pooling [SF16] or global average pooling [He16] when aggregating local embeddings. DGMP balances the activations of specific locations to address over-represented activations. In our work, we optimize the encoding layer of Deep-TEN model by fusing the idea of DGMP and add it on top of the local feature extraction layer.

3 Methodology

Our end-to-end model consists of three deep learning-based parts, a pre-trained U-Net [RFB15] for document images binarization, ResNet-50 [He16] architecture for local descriptors extraction and an optimized encoding layer for global descriptors computation.

3.1 U-Net Pre-training

While layout can give also clues about a writer, we want to rely solely on the script for writer identification. We suggest to apply the common deep learning-based segmentation network *U-Net* before the local feature extraction. Since the Historical-WI dataset does not provide the ground truth for training such a binarization network, we pre-train a U-Net on the DIBCO datasets,[2] which are document image datasets with ground truth for the training of segmentation networks. Afterwards, the U-Net can either be hold fixed or fine-tuned further. Instead of fine-tuning, we could also train the network from scratch but early experiments showed that this was not beneficial.

A huge number of parameters in the transferred network might cause overfitting when it is fine-tuned [Yo14]. Especially in our end-to-end model, the following ResNet-50 already contains a large number of parameters. Therefore, we shrink the size of the transferred U-Net by reducing the number of channels in the first convolutional layer of standard U-Net and maintaining its ratio in the contracting and expansion path [MS18]. Through experimental

[2] We used all available DIBCO datasets below `https://vc.ee.duth.gr/dibco2019/` (excluding year 2019)

validation, we reduce the number of channels in the first convolutional layer to 16 without deteriorating the network performance.

3.2 Optimized Learnable Residual Encoding Layer

After the local feature extraction, a global descriptor for each document image is created using robust residual encoders. In our work, we employ Deep-TEN [ZXD17], which can be seen as a learnable version of VLAD [Jé12], to build our end-to-end writer identification system. Additionally, we integrate DGMP [Ch19] to weigh each local embedding.

The computation of the global representation can be split into two stages: an *embeddig* and an *aggregation* phase [Mu16]. The embedding function ϕ maps each local descriptor to a high-dimensional space, while the aggregation function ψ computes a single global descriptor from the local embedded descriptors, typically by means of sum-pooling. Assuming we have N local descriptors $\boldsymbol{X} = \left\{\boldsymbol{x}_i \in \mathbb{R}^D, i = 1, ..., N\right\}$, the global descriptor is defined as:

$$\boldsymbol{\xi} = \psi\left(\phi\left(\boldsymbol{X}\right)\right) \ . \tag{1}$$

Specifically, if we employ the residual encoding model for the embedding phase, and given a learned codebook $\mathbf{C} = \left\{\boldsymbol{c}_k \in \mathbb{R}^D, k = 1, ..., K\right\}$ with K codewords, we obtain:

$$\phi\left(\boldsymbol{x}_i\right) = a_{ik}\boldsymbol{r}_{ik} \ , \tag{2}$$

where $\boldsymbol{r}_{ik} = \boldsymbol{x}_i - \boldsymbol{c}_k$ represents the residual vector for each local descriptor, a_{ik} is the weight for assigning the local descriptor to the codewords.

Deep-TEN: Zhang et al. [ZXD17] suggests that the codewords are learnable parameters and the weights a_{ik} are calculated by extending the soft-assignment with a learnable smoothing factor s_k for each codeword:

$$a_{ik} = \frac{\exp\left(-s_k \left\|\boldsymbol{r}_{ik}\right\|^2\right)}{\sum_{j=1}^{K} \exp\left(-s_j \left\|\boldsymbol{r}_{ij}\right\|^2\right)} \ . \tag{3}$$

Deep-TEN + DGMP: In our work, instead of using global sum pooling for the aggregation step as Zhang et al. [ZXD17] suggests, we adopt a DGMP [Ch19] layer and integrate it in the encoding layer. Therefore, we introduce a learnable weight $\boldsymbol{\beta} \in \mathbb{R}^{1\times N}$:

$$\boldsymbol{\beta} = \left(\mathbf{K} + \lambda \mathbf{I}_N\right)^{-1} \mathbf{1}_N \ , \tag{4}$$

where $\mathbf{K}$ is the Gram matrix of embeddings, λ is a regularization parameter, $\mathbf{I}_N$ and $\mathbf{1}_N$ are the $N \times N$-dimensional identity matrix and the N-dimensional vector with each element being 1, respectively. Finally, the global descriptor becomes $\boldsymbol{\xi} = \left(\boldsymbol{\xi}_1^\top, ..., \boldsymbol{\xi}_K^\top\right)^\top$, where

$$\boldsymbol{\xi}_k = \psi\left(\phi\left(\boldsymbol{X}\right)\right) = \sum_{i=1}^{N} \beta_i a_{ik} \boldsymbol{r}_{ik} \ . \tag{5}$$

4 Evaluation

In this paper, we first investigate how much accuracy gain can we obtain when applying the optimized encoding layer (DeepTEN+DGMP) on texture datasets compared to the original encoding layer proposed in [ZXD17]. Afterwards, we evaluate our end-to-end model on the Historical-WI dataset [Fi17].

4.1 Data

Texture datasets: (1) The MIT-Indoor [QT09] dataset consists of 67 categories and works for indoor scene recognition. Similar to the experiments mentioned in [ZXD17], we use the standard splittings, i. e. contains 80 images of each category for training and 20 for testing. (2) The outdoor dataset, Flickr Material Dataset (FMD) [Sh13], contains 10 different material categories. For each category, there are 90 images in the training set and 10 in the test set. (3) The publicly available MINC dataset is a large-scale outdoor material dataset with 23 categories. For each category, there are 2500 images. We train the models with 2250 images per category and evaluate the models with 250 images per category.

The Historical-WI dataset [Fi17]: It consists of 4782 handwriting document images. In this dataset, there are 1182 pages provided by 394 writers (each contributes 3 pages) for training, 3600 pages by 720 writers (each contributes 5 pages) for testing. Note that there is no writer providing pages for both training and test set, i. e. the two data splits are disjoint. Next to the color version, the dataset also provides automatically computed binarized images by means of the method of Su et al. [SLT10]. Thus, the segmentation is not always exact, especially in the case of ink artifacts and bleed-through artifacts.

4.2 Experiments

All experiments in this subsection are repeated three times using different random seeds. The applied ResNet50 networks are pre-trained on the ImageNet dataset. First, we evaluate the optimized encoding layer on texture datasets. Then the performance of the end-to-end model on the Historical-WI dataset is evaluated in the second part.

Evaluation of optimized encoding layer on texture datasets: We re-implemented the Deep-TEN model [ZXD17] (Deep-TEN (ours)) and evaluated them on three texture datasets. Afterwards, we evaluate the optimized Deep-TEN (Deep-TEN + DGMP) model by replacing the original encoding layer while keeping the rest unchanged.

Tab. 1: Comparison of the optimized Deep-TEN model with the original Deep-TEN model and our re-implementation on three runs.

Model	MIT-Indoor				FMD				MINC-2500			
	1st	2nd	3rd	Avg.	1st	2nd	3rd	Avg.	1st	2nd	3rd	Avg.
Deep-TEN (orig) [ZXD17]	-	-	-	71.3	-	-	-	80.2	-	-	-	80.6
Deep-TEN (ours)	69.5	69.5	69.6	69.6	75.4	74.5	75.1	75.0	77.2	77.2	77.0	77.1
Deep-TEN + DGMP	71.4	71.6	71.3	71.4	78.8	77.8	77.4	78.0	78.5	78.6	78.4	78.5

Our implementation follows the practice in paper [ZXD17]. For data augmentation, all images are normalized and resized to 400×400. The images in the training set are randomly flipped horizontally (50 % probability) and randomly cropped between 9 % and 100 % of the image areas. The aspect ratio is kept between 3/4 and 4/3. The standard color augmentation [KSH12] is applied. Instead of using stochastic gradient descent (SGD), we apply Adam [KB14] optimizer. The learning rate is initialized to 10^{-4} and multiplied 0.1 when the classification accuracy plateaus. A uniform distribution in range $\left[-\frac{1}{\sqrt{K}}, \frac{1}{\sqrt{K}}\right]$ is used as random initialization for the codewords and smoothing factor. The hyperparameter λ for the optimized model equals to $\lambda = 10^3$. We report Top-1 accuracy for the experiment in this part.

Tab. 1 illustrates the average and standard deviation of Top-1 accuracy of original and optimized Deep-TEN models on the test sets. First, we compare our Deep-TEN implementation with the original paper [ZXD17]. While we tried to be as close as possible to the settings of the paper, we believe that there is still a difference in the evaluation protocol hindering a reproduction of the results. However, we can observe that the performance of the Deep-TEN model with an optimized encoding layer (DeepTEN + DGMP) is always better than our Deep-TEN model using sum-pooling. The biggest gain appears on the FMD dataset, where the optimized Deep-TEN improves by about 3 % on average. For the other two datasets, our optimized encoding layer performs slightly better (> 1 %).

Evaluation of the end-to-end model on Historical-WI dataset: Similar to the implementation in [Ch19], we subdivide the original images into 400×400 patches with a stride of 256. We apply a canny edge detector with a threshold (2000) to judge whether these binarized patches contain sufficient amount of text. However, for color image patches, we set the value of the threshold to 1500 since the edges are less strong. For data augmentation, all subdivided patches are randomly cropped with a size of 300×300. We didn't apply rotation or aspect ratios changes to the patches. For the color image patches, we normalize them to have zero-mean and unit standard deviation. We report both Top-1 accuracy and mean Average Precision (mAP) for the experiments in this part since they are frequently used in the task of writer identification and retrieval.

Tab. 2: Performance comparison of DGMP model (baseline) trained on binarized images with (a) pre-trained U-Nets with either frozen weights or fine-tuned (trained using the provided color images), (b) using Deep-TEN encoding with or without integrated DGMP layer (trained on the binarized images), and (c) using the full pipeline (trained on the color images). All three runs conducted on the Historical-WI [Fi17] testset.

	Model	Top-1				mAP			
		1st	2nd	3rd	Avg.	1st	2nd	3rd	Avg.
	DGMP (baseline)	71.2	70.5	72.8	71.5	52.0	51.9	53.9	52.6
(a)	U-Net (frozen) + DGMP	72.4	72.9	71.8	**72.4**	53.0	53.7	52.9	**53.2**
	U-Net (fine-tuned) + DGMP	71.7	72.6	71.1	71.8	52.6	53.4	51.8	52.6
(b)	Deep-TEN	67.9	69.4	66.4	67.9	48.6	50.2	47.8	48.9
	Deep-TEN + DGMP	61.1	57.0	62.9	60.3	42.1	38.8	44.2	41.7
(c)	U-Net (frozen) + Deep-TEN + DGMP	63.3	65.6	63.4	64.1	44.6	47.2	44.1	45.3

The triplet loss [SKP15] is employed to train our networks, where we use hard-batch online triplet selection [SKP15]. That means each mini-batch consists of P writers with K patches each. In our work, we use the following parameters: $P = 7$, $K = 3$, and triplet loss margin $m = 0.1$. The hyperparameter λ for DGMP is set to $\lambda = 10^3$. We use the Adam optimizer [KB14] with a weight decay of 10^{-3}. The learning rate is initialized with 10^{-5} and multiplied by 0.1 when the mAP plateaus. All experiments are run for 80 epochs, and the models with best validation accuracy are selected for the testing phase.

We have designed three experiments to evaluate the performance of our end-to-end model. The first two experiments address to evaluate the effects of each deep learning-based stage of our end-to-end network against its non-deep learning method. In the final experiment, the integrated end-to-end model is implemented and evaluated. The model proposed in [Ch19] acts as the baseline for all of our experiments in this part. The baseline model consists of fine-tuning a pre-trained ResNet-50 followed by a DGMP layer, i. e. no Deep-TEN encoding or U-Net is used. It was trained on the provided binarized images of the dataset. Note that the results differ from the ones reported in the paper [Ch19] due to a different batch size and a reduced input size.

(a) Effect of pre-trained U-Net: We train two models to evaluate the effects of the pre-trained U-Net. The U-Net takes a 300×300 input patch and outputs an equally-sized segmentation. Both models are composed of a pre-trained U-Net followed by the baseline architecture. We freeze the parameters of the U-Net in the first model while we fine-tuning the U-Net in the second model. Note that both models are trained with color images. The results are given in Tab. 2a. We see that the performance increases when the pre-trained U-Net is transferred to a writer identification network in case of frozen U-Net weights.

(b) Effect of learnable Deep-TEN layer: The only difference between our Deep-TEN architecture and the DGMP model (baseline) is the learnable encoding layer. In this experiment, the models are evaluated on the binarized Historical-WI dataset. Tab. 2b shows the test accuracy of the models. Both the original and the optimized encoding layer make the results worse. The pure Deep-TEN model even works better than the optimized one (Deep-TEN + DGMP). It seems that the learnable residual encoding layer does not work well in the task of writer identification. Moreover, it seems that the encoding layer does not benefit from the DGMP aggregation in this task. To this end, we can only speculate why this is the case. Maybe the Deep-TEN encoding layer encourages overfitting on the training writers and hence does not generalize well on the unseen test samples. Note that we also experimented with different learning rates, weight decay, etc.

(c) Full model: Finally, we evaluate the whole model, cf. Tab. 2c. While the additional binarization step improves over Deep-TEN + DGMP, the encoding stage by Deep-TEN worsened the results too much that it could possibly outperform the baseline model.

5 Conclusion

This paper aimed to propose a fully end-to-end deep learning-based pipeline for writer identification. To achieve this goal, we mainly have investigated two kinds of technologies, image binarization and the use of global feature encoding. The U-Net, which is pre-trained on the DIBCO dataset, works as the binarization layer in our end-to-end model. The results of our experiments show that the pre-trained U-Net outperforms the traditional method. However, a fine-tuning of the U-Net was not beneficial, thus this could also be a separate pre-processing step. Moreover, we evaluated Deep-TEN, an encoding technique to compute the global descriptor. We also incorporated DGMP aggregation mechanism. This improved encoding layer worked fine for texture classification. However, for the historical writer identification dataset, both pure and enhanced encoding layer worsen the performance. Overall, a short pipeline of just binarization, feature extraction by triplet loss and just use the weighted average by means of DGMP seems to be better than using a sophisticated encoding layer.

Acknowledgement: This work has been partly supported by the Cross-border Cooperation Program Czech Republic – Free State of Bavaria ETS Objective 2014–2020 (project no. 211).

References

[Ar16] Arandjelovic, R.; Gronat, P.; Torii, A.; Pajdla, T.; Sivic, J.: NetVLAD: CNN architecture for weakly supervised place recognition. In: CVPR. Pp. 5297–5307, 2016.

[CBA15] Christlein, V.; Bernecker, D.; Angelopoulou, E.: Writer identification using VLAD encoded contour-Zernike moments. In: ICDAR. IEEE, pp. 906–910, 2015.

[Ch15] Christlein, V.; Bernecker, D.; Maier, A.; Angelopoulou, E.: Offline writer identification using convolutional neural network activation features. In: German Conference on Pattern Recognition. Springer, pp. 540–552, 2015.

[Ch17] Christlein, V.; Gropp, M.; Fiel, S.; Maier, A.: Unsupervised feature learning for writer identification and writer retrieval. In: ICDAR. Vol. 1, IEEE, pp. 991–997, 2017.

[Ch19] Christlein, V.; Spranger, L.; Seuret, M.; Nicolaou, A.; Král, P.; Maier, A.: Deep Generalized Max Pooling. In: ICDAR. IEEE, pp. 1090–1096, 2019.

[Fi17] Fiel, S.; Kleber, F.; Diem, M.; Christlein, V.; Louloudis, G.; Nikos, S.; Gatos, B.: Icdar2017 competition on historical document writer identification (historical-wi). In: ICDAR. Vol. 1, IEEE, pp. 1377–1382, 2017.

[FS15] Fiel, S.; Sablatnig, R.: Writer identification and retrieval using a convolutional neural network. In: International Conference on Computer Analysis of Images and Patterns. Springer, pp. 26–37, 2015.

[He16] He, K.; Zhang, X.; Ren, S.; Sun, J.: Deep residual learning for image recognition. In: CVPR. IEEE, pp. 770–778, 2016.

[Jé12] Jégou, H.; Perronnin, F.; Douze, M.; Sánchez, J.; Pérez, P.; Schmid, C.: Aggregating Local Image Descriptors into Compact Codes. IEEE Transactions on Pattern Analysis and Machine Intelligence 34/9, pp. 1704–1716, Sept. 2012.

[KB14] Kingma, D. P.; Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980/, 2014.

[KFS18] Keglevic, M.; Fiel, S.; Sablatnig, R.: Learning features for writer retrieval and identification using triplet cnns. In: 2018 16th International Conference on Frontiers in Handwriting Recognition (ICFHR). IEEE, pp. 211–216, 2018.

[KSH12] Krizhevsky, A.; Sutskever, I.; Hinton, G. E.: Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems. Pp. 1097–1105, 2012.

[Lo99] Lowe, D. G.: Object recognition from local scale-invariant features. In: ICCV. Vol. 2, Ieee, pp. 1150–1157, 1999.

[LSD15] Long, J.; Shelhamer, E.; Darrell, T.: Fully convolutional networks for semantic segmentation. In: CVPR. Pp. 3431–3440, 2015.

[MS18] Mangalam, K.; Salzamann, M.: On compressing u-net using knowledge distillation. arXiv preprint arXiv:1812.00249/, 2018.

[Mu16] Murray, N.; Jégou, H.; Perronnin, F.; Zisserman, A.: Interferences in match kernels. IEEE transactions on pattern analysis and machine intelligence 39/9, pp. 1797–1810, 2016.

[QT09] Quattoni, A.; Torralba, A.: Recognizing indoor scenes. In: CVPR. IEEE, pp. 413–420, 2009.

[RFB15] Ronneberger, O.; Fischer, P.; Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. Springer, pp. 234–241, 2015.

[SF16] Sudholt, S.; Fink, G. A.: Phocnet: A deep convolutional neural network for word spotting in handwritten documents. In: 2016 15th International Conference on Frontiers in Handwriting Recognition (ICFHR). IEEE, pp. 277–282, 2016.

[Sh13] Sharan, L.; Liu, C.; Rosenholtz, R.; Adelson, E. H.: Recognizing materials using perceptually inspired features. International journal of computer vision 103/3, pp. 348–371, 2013.

[SKP15] Schroff, F.; Kalenichenko, D.; Philbin, J.: Facenet: A unified embedding for face recognition and clustering. In: CVPR. IEEE, pp. 815–823, 2015.

[SLT10] Su, B.; Lu, S.; Tan, C. L.: Binarization of historical document images using the local maximum and minimum. In: Proceedings of the 9th IAPR International Workshop on Document Analysis Systems. Pp. 159–166, 2010.

[TM17] Tensmeyer, C.; Martinez, T.: Document image binarization with fully convolutional neural networks. In: ICDAR. Vol. 1, IEEE, pp. 99–104, 2017.

[WTB14] Wu, X.; Tang, Y.; Bu, W.: Offline text-independent writer identification based on scale invariant feature transform. IEEE Transactions on Information Forensics and Security 9/3, pp. 526–536, 2014.

[Yo14] Yosinski, J.; Clune, J.; Bengio, Y.; Lipson, H.: How transferable are features in deep neural networks? In: Advances in neural information processing systems. Pp. 3320–3328, 2014.

[ZXD17] Zhang, H.; Xue, J.; Dana, K.: Deep ten: Texture encoding network. In: CVPR. IEEE, pp. 708–717, 2017.

R. Reussner, A. Koziolek, R. Heinrich (Hrsg.): INFORMATIK 2020,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2021 1355

The Dissimilar in the Similar. An Attribute-guided Approach to the Subject-specific Classification of Art-historical Objects

Stefanie Schneider[1], Matthias Springstein[2], Javad Rahnama[3], Eyke Hüllermeier[3], Ralph Ewerth[2,4], Hubertus Kohle[1]

Abstract: Due to the increasingly unmanageable number of art-historical inventories made available in digital form, methods that computationally arrange larger amounts of objects are becoming more important. The category of similarity, which is fundamental in all areas of art-historical description, gains new relevance in this context. In this paper, we propose a novel approach to the subject-specific classification of art-historical objects that utilizes expert-based attributes, i.e., significant figurative motifs. We evaluate our procedure on a concrete use case, representations of saints in the visual arts. A representative data set of saints images is collected and a semi-supervised learning technique applied to enrich the data set with neural style transfer as well as to improve the joint training of saints and their attributes. We show that this technique outperforms other approaches.

Keywords: Semi-supervised Learning; Semi-supervised Image Classification; Art Analysis; Digital Humanities

1 Introduction

The category of similarity is fundamental in all areas of art-historical description: in the history of style, the specification of formal characteristics determines the assignment of artistic phenomena to stylistic attitudes; in iconography, definitions of content are constituted by the observation of comparable—or similar—conventions of representation. Similarity also plays a central role in art-historical practice. When Wölfflin compares a portrait of Albrecht Dürer with one of Frans Hals–inter alia, in the form of the categories of the "linear" and the "painterly"—, he is assuming that the two works were painted in different ways while belonging to the same genre [Wö15]. Decisive for the persuasiveness of this procedure is the determination of the 'dissimilar in the similar': for only (or especially) where a common set of phenomena exists do possible differences become visible and plausible.

Because of the increasingly unmanageable number of art-historical inventories made available in digital form [MG14], two questions arise. Firstly, how can the manifold concepts

[1] Ludwig-Maximilians-Universität München, Institut für Kunstgeschichte, Zentnerstr. 31, 80798 München, {stefanie.schneider@itg.uni-muenchen.de, hubertus.kohle@lmu.de}.
[2] Technische Informationsbibliothek (TIB), Welfengarten 1b, 30167 Hannover, {matthias.springstein@tib.eu, ralph.ewerth@tib.eu}.
[3] Universität Paderborn, Fachgruppe Intelligente Systeme und Maschinelles Lernen, Pohlweg 51, 33098 Paderborn, {javad.rahnama@uni-paderborn.de, eyke@upb.de}.
[4] Forschungszentrum L3S, Leibniz Universität Hannover, Appelstraße 9a, 30167 Hannover.

doi:10.18420/inf2020_127

of similarity be considered to relate larger amounts of objects computationally? Secondly, are existing methods suitable for such heterogeneous inventories and, if so, to what extent can they be adopted and optimized? Previous studies on the automatic detection, recognition, or identification of objects relevant to image science focus either on small visually distinctive sub-fields, e.g., ballad prints [TBO14] and tinted drawings [Ya11], or larger non-specialised data sets, e.g., WikiArt [HWS16], that predominantly feature well-known Western artists and art periods. They thus do only partially account for the great diversity of historical artefacts and lack the generalizability necessary for this domain.

In this work, we concentrate on the broader category of iconographic similarity and propose a generic approach to the subject-specific classification of art-historical objects that utilizes expert-based attributes of the classification system Iconclass, i.e., figurative motifs significant from an art-historical point-of-view. This is the first attempt to actively exploit Iconclass in automatic classification tasks, to the best of our knowledge. We evaluate our procedure on a concrete use case, representations of saints in the visual arts. This example is advantageous because it is usually possible to clearly assign the saint and the attributes identifying him or her: the attributes are placed in a spatially comprehensible relationship to the person, i.e., they are positioned close to it, even if sometimes hidden. The latter is especially true for phases of art history in which, as in 16th-century Mannerism, the clear legibility of a picture's content was not the main focus. Since many art-historical narratives, especially those of Christian religion and classical mythology, feature sufficiently informative attributes (or attribute-like concepts), this approach is widely applicable.

The contributions are as follows: *(i)* collection of a representative data set of saints, *(ii)* a novel approach to attribute-guided classification that utilizes Iconclass, and *(iii)* application of a semi-supervised learning technique to enrich the data set with neural style transfer as well as to improve the joint training of saints and their attributes.

2 Related Work

Due to the recent growth in computerized analysis of cultural heritage, we primarily discuss studies that address the categorization of art-historical objects.

To classify art periods such as Baroque and Symbolism, Hentschel et al. [HWS16] contrast Fisher Vectors and a Support Vector Machine with a Convolutional Neural Network (CNN) pre-trained on ImageNet and fine-tuned on WikiArt. Anwer et al. [An16] extend on this methodology by also utilizing information about local regions of interest with a Deformable Part Model. In earlier and less relevant works, Gatys et al. [GEB15] train a CNN to capture, separate, and reconstruct the content of an object, and its style, whereas Saleh and Elgammal [SE15] combine low-level and high-level features to categorize style, genre, and artist. A CNN is trained on top of the last layer of an ImageNet-trained network to capture additional semantic features. More recently, Bianco et al. [Bi19] propose a multitask-multibranch CNN to simultaneously classify style, genre, and artist. In contrast, Yang et al. [Ya18] encode

Fig. 1: Images of the attributes “baptismal cup”, “book”, and “lamb”, retrieved from *Google Image Search* respectively.

complementary material to assist visual feature learning in CNNs for style classification. Sabatelli et al. [Sa18] investigate the general effect of fine-tuned CNNs in artist, material, and type classification tasks.

However, studies rarely incorporate concepts significant to iconography. In one of the few exceptions, Gonthier et al. [Go18] propose a multiple instance learning (MIL) technique for the weakly-supervised detection of art-historically specific objects. However, as image-level annotations are only gathered for 7 classes, the generalizability of the approach remains unclear, especially for concepts with high in-class variability. In this work, we focus entirely on a unified set of art-historically relevant classes that are of a comparably high visual and narrative complexity: representations of saints in the visual arts. Like Yang et al. [Ya18], we utilize historical context information—here expert-based attributes that are linked to the respective class, i.e., the respective saint—to improve the subject-specific classification of concepts with high in-class variability.

3 Data

Data set collection Our data set consists of two kinds of images: art-historical and non-art-historical, i.e., real-world imagery.

A total of 19 publicly available inventories, collections, institutions, and web portals are first harvested to gather depictions of saints in the visual arts.[5] The obtained reproductions are extremely varied and, e.g., include stained glass paintings of the Middle Ages, 16th-century emblems as well as Polish folk woodcuts. Each source is at least partially indexed by experts with the decimal classification system Iconclass that was specially conceived for the Western motifs of the visual arts [Wa85]. It thus also contains definitions of male and female saints, where each saint is provided with an explanatory textual correlate including a list of possible attributes.[6] This information is used to retrieve real images of the attributes from *Google*

[5] `artemis.uni-muenchen.de`, `https://www.bildindex.de/`, `http://ballads.bodleian.ox.ac.uk/`, `https://corpusvitrearum.de/`, `emblematica.library.illinois.edu`, `heartfield.adk.de`, `https://inkunabeln.digitale-sammlungen.de/`, `http://manuscripts.kb.nl/`, `http://www.museen.thueringen.de/`, `https://www.nga.gov/`, `https://datenbank.museum-kassel.de/`, `https://sammlung.belvedere.at/`, `http://pauart.pl/app`, `https://realonline.imareal.sbg.ac.at/`, `https://www.rijksmuseum.nl/en`, `https://rkd.nl/en/`, `sammlung.staedelmuseum.de`, `http://www.virtuelles-kupferstichkabinett.de/de/`, and `https://vitrosearch.ch/de`, respectively (all accessed April 28, 2020).

[6] All other notations are accompanied by a list of keywords, some of which can be defined as attributes, or at least have attribute-like properties.

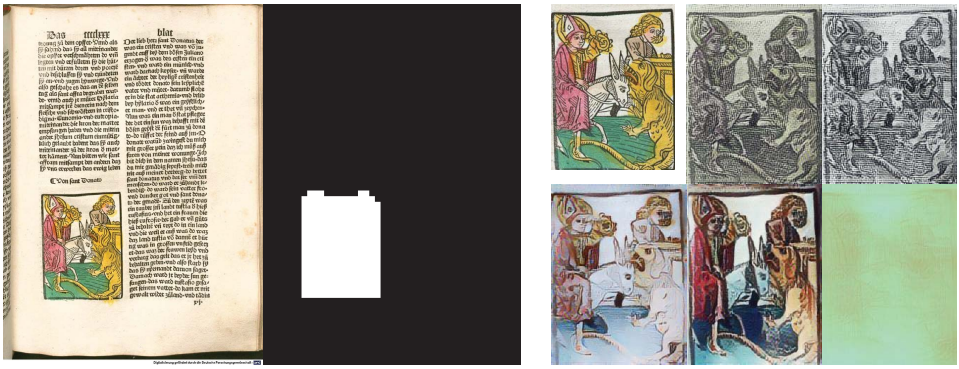


Fig. 2: Detection of bounding boxes (left) and application of style transfer to enrich the data set (right).

Image Search, i.e., photographs taken in recent years. As shown in Figure 1, not all images include the desired attribute in the narrow sense; e.g., a modern e-reader was found as well as lamb meat. In so doing, we collect 21,479 images of 239 saints and 124,133 images of 343 attributes for training and testing our procedures.

Data set preprocessing Many of the previously harvested representations are scans and contain background noise or further information, e.g., signatures of the artist or linear color control charts of the institution responsible for the reproduction. Two preprocessing steps are necessary to use these representations for training a neural network. Relevant image content is first detected using a DeepLabv3 image segmentation model trained on 100 examples from the afore-introduced saints data set [Ch17]. The overlapping image regions thus identified are then integrated into one rectangular region. If a region has a width or height of less than 100 pixels, it is discarded. An example prediction of the trained DeepLabv3 model is shown on the left in Figure 2. As the images of the saints and the images of the attributes originate from highly different domains, we deploy neural style transfer to enrich the data set and bridge the gap between domains [Gh17].[7] Up to 5 variations of the original image are created, where we choose a random image of a saint as a style image.

As depicted in Figure 2, not all images that are generated in this way are recognizable. On the one hand, this is due to the fact that style images are randomly selected and applied from all available saints images. On the other hand, the segmentation introduces errors; therefore, images are selected for style transfer that do not show any saint. On the basis of these steps, the number of images containing (representations of) saints increases to 25,667; the subsequent style transfer further increases the number to 120,626. The number of images depicting attributes increases to 403,788.

[7] Geirhos et al. [Ge19] also show that such techniques increase the robustness of neural versus textural change.

Fig. 3: Four representations of Saint John the Baptist with the exemplary selected attribute “lamb”.

4 Attribute-guided Classification

The idea behind our approach is as follows: generally, a saint cannot be identified exclusively by his or her physiognomy, but by a set of pictorial signs, *attributes*, that exemplify a special event in his life or take up characteristics of her status or profession. A distinction must be made between attributes characterizing a (larger) group of saints and attributes that are narratively significant for a particular saint. While, e.g., the staff serves as a general sign of holy abbots, John the Baptist is often accompanied by a lamb to recall the acclamation in which he refers to Christ as the “Lamb of God” (Figure 3). Since most attributes act as binding signifiers, they are often featured prominently in the fore- or background of an image and can thus support the computer-aided classification of saints. We assume that the joint appearance of even relatively trivial appearing or art-historically unspecific attributes, whose artistic depiction has hardly changed over time, is sufficient for this purpose.[8]

Two problems arise. On the one hand, a saint can be identified by more than one attribute; however, not *all* attributes need to be present in the image of a saint. On the other hand, the images found via *Google Image Search* do not always show the desired attribute, or solely modernized versions of it, as already illustrated in Section 3. We thus propose a semi-supervised learning technique based on FixMatch [So20]. The original objective of FixMatch is to use unlabeled data for training an image classifier. In doing so, unlabeled images for which the model predicts a high probability are automatically assigned to a concept and used for the training process. In our case, we use this technique to automatically annotate attributes in images of saints that were *not* originally annotated.

The training process for a batch is shown in Figure 4. During each iteration, the model forwards two batches of labeled images, $B_{l,s}$ for saints and $B_{l,a}$ for attributes, as well as two batches of unlabeled images, $B_{u,s}$ for saints and $B_{u,a}$ for attributes. It then determines the probability for the concept saints, p_s, and for the concept attributes, p_a, independently for each input image of the batch. The supervised loss L_l—applied to $B_{l,s}$ and $B_{l,a}$,

[8] This is in stark contrast to Gonthier et al., who state that “more specific objects or attributes such as ruins or nudity” are needed to detect [Go18, p. 2].

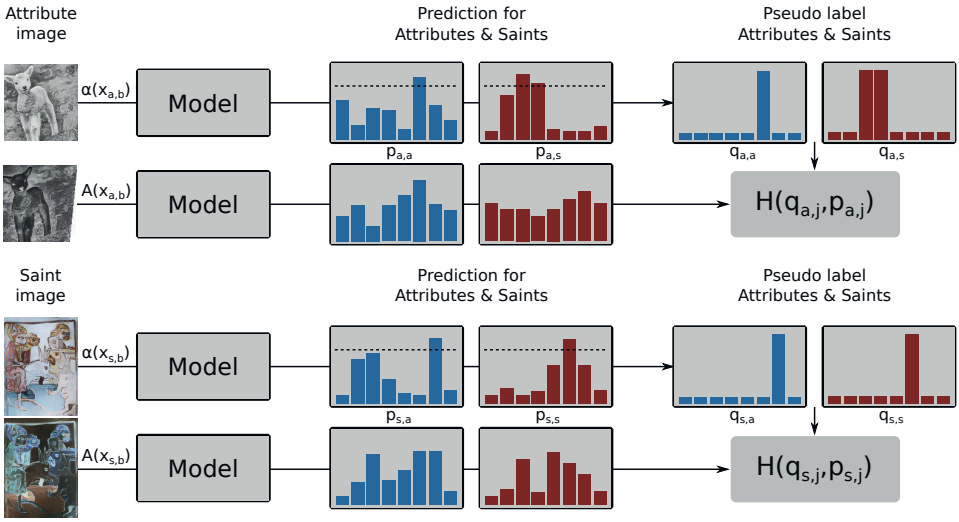


Fig. 4: Visualization of the semi-supervised learning technique. During each iteration, the system predicts a probability distribution for attributes (*blue*) and saints (*red*) that is used to generate pseudo-labels. These labels are then used as optimization targets for the same image with a different augmentation strategy.

respectively—results from the cross-entropy $H(\cdot)$ between the encoded label $\hat{y}_i$ and the prediction $p_{i,i}$ for an input $x_{i,b}$:

$$L_l = \sum_{i \in \{s,a\}} \frac{1}{B_{l,i}} \sum_{b=1}^{B_{l,i}} H\left(\hat{y}_{i,b}, p_{i,i}\left(y \mid \alpha\left(x_{i,b}\right)\right)\right) \tag{1}$$

For unlabeled data, we first compute the model's predicted class distribution for a *weakly*-augmented α-version of the sample $x_{i,b}$ in each subset $B_{u,s}$ and $B_{u,a}$. To create an artificial label, we assign a value of one for each prediction of a concept that is greater than a threshold τ; all other concepts are set to zero:

$$\hat{q}_{i,j,b} = \begin{cases} 1 & \text{if } p_{i,j}\left(y \mid \alpha\left(x_{i,b}\right)\right) \geq \tau \\ 0 & \text{if } p_{i,j}\left(y \mid \alpha\left(x_{i,b}\right)\right) < \tau \end{cases} \tag{2}$$

The unsupervised loss L_u results from the *strongly*-augmented version A of the image $x_{i,b}$ and the pseudo-label $\hat{q}_{i,j,b}$, as long as there is at least one prediction above the threshold τ:

$$L_u = \sum_{j \in \{s,a\}} \sum_{i \in \{s,a\}} \frac{1}{B_{u,i}} \sum_{b=1}^{B_{u,i}} \mathbb{1}\left(\max\left(p_{i,j}\left(y \mid \alpha\left(x_{i,b}\right)\right)\right) \geq \tau\right) H\left(\hat{q}_{i,j,b}, p_{i,j}\left(y \mid A\left(x_{i,b}\right)\right)\right) \tag{3}$$

The final loss L is simply the sum $L = L_l + L_u$. Since all images that contain neither a saint nor an attribute have a low probability of showing any relevant concept, they are

Attribute	AP	Attribute	AP	Attribute	AP	Attribute	AP
peacock feather	0.871	tablet	0.859	ducal hat	0.030	cope	0.016
scissors	0.870	hackle	0.845	net	0.026	stake	0.015
monstrance	0.867	tiara	0.844	Spes	0.026	Turk	0.014
staircase	0.866	broom	0.840	head	0.019	three	0.011
clog	0.865	wreath	0.837	mitre	0.017	two	0.007

Tab. 1: Best and worst classification results based on the data set with 343 attributes retrieved from *Google Image Search*. Average Precision (AP) is used to measure the retrieval performance.

automatically excluded during training. This procedure offers two advantages. When an attribute is recognized in the image of a saint, it is automatically annotated; in this way, there is feedback from attributes in images that were *not* originally annotated. Second, images that are not recognizable by the model after style transfer are excluded from training.

5 Experiments

We employ a ResNet-50 architecture pre-trained on ImageNet [He16]. The optimization is carried out using Stochastic gradient descent (SGD) with Nesterov momentum of 0.9 [Su13]. The initial learning rate is set to 0.01. The data set is split into training, validation, and test with a splitting ratio of 3:1:1. We evaluate the model with the highest accuracy on the validation set on the test set, respectively. Mean Average Precision (mAP) is used to measure the retrieval performance of our system for the entire test set.

Attribute classification We first evaluate whether the attributes data set is generally suitable for the prediction of saints. The model achieves a performance of 0.354 mAP. As shown in Table 1, attributes that are difficult to define ("three") or cannot be found by *Google Image Search* ("mitre") lead to poor classification performance, whereas objects still common in modern everyday life ("scissors") naturally show more promising results.

Joint training of saints and attributes Our approach to jointly train saints and attributes is compared to two baseline strategies, with and without style transfer, respectively. Thus, both saints classifiers do not use explicitly defined visual attributes during training. Random horizontal flip is used as augmentation step. In addition, we use *RandAugment* for the FixMatch approach, which applies a random transformation with a defined strength from a fixed set [Cu19]. We moreover use style-transferred images from the saints and attributes data set, respectively, as unlabeled input for FixMatch. The performance of the procedure is reported for the 49 saints with the most images, and only for images after the bounding box detection (see Section 3). As shown in Table 2, the proposed system performs best, mAP = 0.136, when a threshold of $\tau = 0.5$ is chosen. If the threshold is set too high, not

Method	$B_{l,s}$	$B_{u,s}$	$B_{l,a}$	$B_{u,a}$	mAP	Accuracy
Random					0.021	0.054
Saints (without Style Transfer)	16	0	0	0	0.131	0.250
Saints (with Style Transfer)	16	0	0	0	0.118	0.246
Saints and Attributes (without Style Transfer)	8	0	8	0	0.120	0.241
Saints and Attributes (with Style Transfer)	8	0	8	0	0.128	0.252
FixMatch ($\tau = 0.4$)	8	8	8	8	0.093	0.210
FixMatch ($\tau = 0.5$)	8	8	8	8	**0.136**	**0.260**
FixMatch ($\tau = 0.6$)	8	8	8	8	0.134	0.245

Tab. 2: Scores of the classification methods based on the data set with 49 saints. $B_{l,s}$ and $B_{l,a}$ denote the batch sizes of labeled images for saints and attributes, respectively, $B_{u,s}$ and $B_{u,a}$ the batch sizes of unlabeled images for saints and attributes, respectively. The best performing approach is bold.

enough images are selected for training or not all concepts in an image are selected. If the threshold is set too low, however, too many concepts are selected. We chose 0.5 as a starting point because it is commonly used to generate binary decisions after a sigmoid activation.

A closer look at the results shows that saints are more accurately classified if their depictions are limited to few narratives, or a certain stage of life is primarily illustrated, e.g., in the case of Jerome (AP = 0.432), even if differing materials or techniques are used. If, on the other hand, a saint can be represented in many strongly varying ways that are not related to any specific constellation of attributes, such as Bernard (AP = 0.036), classification results drop immensely. This is especially true for saints, like Helena (AP = 0.020), for whom there are few examples or many visually distinctive ones, e.g., engravings, stained glass paintings, or early preparatory drawings. These findings illustrate that the enormous complexity of the domain, in which an object can be depicted in various ways, is often only insufficiently manageable—even with common augmentation techniques and fine-tuned networks. The underlying phenomenon, referred to as the "cross-depiction problem" [WCH16, p. 1], might possibly be weakened by more sophisticated domain adaptation techniques [TK18]. Moreover, to mitigate the dependency on non-art-historical imagery and further improve classification, the harvested collections could be exploited more extensively, since many attributes are listed in Iconclass as separate notations.

6 Conclusion

In this work, we introduced a new data set and task for the identification of saints in the visual arts. We suggested a novel deep-learning approach that utilizes expert-based attributes to support the subject-specific classification especially of concepts with high in-class variability. The proposed semi-supervised joint training technique increases the performance compared to multiple baselines. In the future, we will apply this procedure to the classification of other art-historically relevant narratives and motifs that can possibly

also be improved by the use of visual attributes. To further improve the discrimination of saints (or other individuals relevant to art history), we plan to explore different loss functions, e.g., contrastive or triplet loss, as they are successfully used in face recognition tasks.

Acknowledgements

This work is financially supported by the German Research Foundation (DFG) under project number 415796915.

References

[An16] Anwer, R. M.; Khan, F. S.; van de Weijer, J.; Laaksonen, J.: Combining Holistic and Part-based Deep Representations for Computational Painting Categorization. In: Proceedings of the 2016 ACM International Conference on Multimedia Retrieval. Pp. 339–342, 2016.

[Bi19] Bianco, S.; Mazzini, D.; Napoletano, P.; Schettini, R.: Multitask Painting Categorization by Deep Multibranch Neural Network. In: Expert Systems with Applications. Vol. 135, pp. 90–101, 2019.

[Ch17] Chen, L.-C.; Papandreou, G.; Schroff, F.; Adam, H.: Rethinking Atrous Convolution for Semantic Image Segmentation, 2017, arXiv: 1706.05587, URL: http://arxiv.org/abs/1706.05587.

[Cu19] Cubuk, E. D.; Zoph, B.; Shlens, J.; Le, Q. V.: RandAugment. Practical Data Augmentation with No Separate Search, 2019, arXiv: 1909.13719, URL: http://arxiv.org/abs/1909.13719.

[Ge19] Geirhos, R.; Rubisch, P.; Michaelis, C.; Bethge, M.; Wichmann, F. A.; Brendel, W.: ImageNet-trained CNNs are Biased Towards Texture; Increasing Shape Bias Improves Accuracy and Robustness. In: 7th International Conference on Learning Representations. 2019.

[GEB15] Gatys, L. A.; Ecker, A. S.; Bethge, M.: A Neural Algorithm of Artistic Style, 2015, arXiv: 1508.06576, URL: https://arxiv.org/abs/1508.06576.

[Gh17] Ghiasi, G.; Lee, H.; Kudlur, M.; Dumoulin, V.; Shlens, J.: Exploring the Structure of a Real-time, Arbitrary Neural Artistic Stylization Network. In: British Machine Vision Conference. 2017.

[Go18] Gonthier, N.; Gousseau, Y.; Ladjal, S.; Bonfait, O.: Weakly Supervised Object Detection in Artworks, 2018, arXiv: 1810.02569, URL: https://arxiv.org/abs/1810.02569.

[He16] He, K.; Zhang, X.; Ren, S.; Sun, J.: Identity Mappings in Deep Residual Networks. In: Computer Vision – ECCV 2016. Vol. 9908, Springer, pp. 630–645, 2016.

[HWS16] Hentschel, C.; Wiradarma, T. P.; Sack, H.: An Approach to Large Scale Interactive Retrieval of Cultural Heritage. In: Proceedings of the 23th IEEE International Conference on Image Processing. Pp. 3693–3697, 2016.

[MG14] Mensink, T.; van Gemert, J.: The Rijksmuseum Challenge. Museum-centered Visual Recognition. In: Proceedings of the International Conference on Multimedia Retrieval. Pp. 451–454, 2014.

[Sa18] Sabatelli, M.; Kestemont, M.; Daelemans, W.; Geurts, P.: Deep Transfer Learning for Art Classification Problems. In: Computer Vision – ECCV 2018 Workshops. Vol. 48, Springer, 2018.

[SE15] Saleh, B.; Elgammal, A. M.: Large-scale Classification of Fine-art Paintings. Learning the Right Metric on the Right Feature, 2015, arXiv: 1505.00855, URL: https://arxiv.org/abs/1505.00855.

[So20] Sohn, K.; Berthelot, D.; Li, C.-L.; Zhang, Z.; Carlini, N.; Cubuk, E. D.; Kurakin, A.; Zhang, H.; Raffel, C.: FixMatch. Simplifying Semi-supervised Learning with Consistency and Confidence, 2020, arXiv: 2001.07685, URL: https://arxiv.org/abs/2001.07685.

[Su13] Sutskever, I.; Martens, J.; Dahl, G. E.; Hinton, G. E.: On the Importance of Initialization and Momentum in Deep Learning. In: Proceedings of the 30th International Conference on Machine Learning. Pp. 1139–1147, 2013.

[TBO14] Takami, M.; Bell, P.; Ommer, B.: An Approach to Large Scale Interactive Retrieval of Cultural Heritage. In: Eurographics Workshop on Graphics and Cultural Heritage. Pp. 87–95, 2014.

[TK18] Thomas, C.; Kovashka, A.: Artistic Object Recognition by Unsupervised Style Adaptation. In: Computer Vision – ACCV 2018. Vol. 29, Springer, 2018.

[Wa85] van de Waal, H.: Iconclass. An Iconographic Classification System. Completed and Edited by L. D. Couprie with R. H. Fuchs. North-Holland Publishing Company, Amsterdam, 1973–1985.

[WCH16] Westlake, N.; Cai, H.; Hall, P.: Detecting People in Artwork with CNNs. In: Computer Vision – ECCV 2016 Workshops. Vol. 9913, Springer, 2016.

[Wö15] Wölfflin, H.: Kunstgeschichtliche Grundbegriffe. Bruckmann, Munich, 1915.

[Ya11] Yarlagadda, P.; Monroy, A.; Carque, B.; Ommer, B.: Recognition and Analysis of Objects in Medieval Images. In: Proceedings of the ACCV Workshop on Computer Vision. Pp. 296–305, 2011.

[Ya18] Yang, J.; Chen, L.; Zhang, L.; Sun, X.; She, D.; Lu, S.; Cheng, M.-M.: Historical Context-based Style Classification of Painting Images via Label Distribution Learning. In: Proceedings of the 26th ACM International Conference on Multimedia. Pp. 1154–1162, 2018.

Autorenverzeichnis

GI-Edition Lecture Notes in Informatics

P-1 Gregor Engels, Andreas Oberweis, Albert Zündorf (Hrsg.): Modellierung 2001.

P-2 Mikhail Godlevsky, Heinrich C. Mayr (Hrsg.): Information Systems Technology and its Applications, ISTA'2001.

P-3 Ana M. Moreno, Reind P. van de Riet (Hrsg.): Applications of Natural Lan-guage to Information Systems, NLDB'2001.

P-4 H. Wörn, J. Mühling, C. Vahl, H.-P. Meinzer (Hrsg.): Rechner- und sensorgestützte Chirurgie; Workshop des SFB 414.

P-5 Andy Schürr (Hg.): OMER – Object-Oriented Modeling of Embedded Real-Time Systems.

P-6 Hans-Jürgen Appelrath, Rolf Beyer, Uwe Marquardt, Heinrich C. Mayr, Claudia Steinberger (Hrsg.): Unternehmen Hochschule, UH'2001.

P-7 Andy Evans, Robert France, Ana Moreira, Bernhard Rumpe (Hrsg.): Practical UML-Based Rigorous Development Methods – Countering or Integrating the extremists, pUML'2001.

P-8 Reinhard Keil-Slawik, Johannes Magenheim (Hrsg.): Informatikunterricht und Medienbildung, INFOS'2001.

P-9 Jan von Knop, Wilhelm Haverkamp (Hrsg.): Innovative Anwendungen in Kommunikationsnetzen, 15. DFN Arbeitstagung.

P-10 Mirjam Minor, Steffen Staab (Hrsg.): 1st German Workshop on Experience Management: Sharing Experiences about the Sharing Experience.

P-11 Michael Weber, Frank Kargl (Hrsg.): Mobile Ad-Hoc Netzwerke, WMAN 2002.

P-12 Martin Glinz, Günther Müller-Luschnat (Hrsg.): Modellierung 2002.

P-13 Jan von Knop, Peter Schirmbacher and Viljan Mahni_ (Hrsg.): The Changing Universities – The Role of Technology.

P-14 Robert Tolksdorf, Rainer Eckstein (Hrsg.): XML-Technologien für das Semantic Web – XSW 2002.

P-15 Hans-Bernd Bludau, Andreas Koop (Hrsg.): Mobile Computing in Medicine.

P-16 J. Felix Hampe, Gerhard Schwabe (Hrsg.): Mobile and Collaborative Business 2002.

P-17 Jan von Knop, Wilhelm Haverkamp (Hrsg.): Zukunft der Netze –Die Verletzbarkeit meistern, 16. DFN Arbeitstagung.

P-18 Elmar J. Sinz, Markus Plaha (Hrsg.): Modellierung betrieblicher Informationssysteme – MobIS 2002.

P-19 Sigrid Schubert, Bernd Reusch, Norbert Jesse (Hrsg.): Informatik bewegt – Informatik 2002 – 32. Jahrestagung der Gesellschaft für Informatik e.V. (GI) 30.Sept.-3. Okt. 2002 in Dortmund.

P-20 Sigrid Schubert, Bernd Reusch, Norbert Jesse (Hrsg.): Informatik bewegt – Informatik 2002 – 32. Jahrestagung der Gesellschaft für Informatik e.V. (GI) 30.Sept.-3. Okt. 2002 in Dortmund (Ergänzungs-band).

P-21 Jörg Desel, Mathias Weske (Hrsg.): Promise 2002: Prozessorientierte Methoden und Werkzeuge für die Entwicklung von Informationssystemen.

P-22 Sigrid Schubert, Johannes Magenheim, Peter Hubwieser, Torsten Brinda (Hrsg.): Forschungsbeiträge zur "Didaktik der Informatik" – Theorie, Praxis, Evaluation.

P-23 Thorsten Spitta, Jens Borchers, Harry M. Sneed (Hrsg.): Software Management 2002 – Fortschritt durch Beständigkeit

P-24 Rainer Eckstein, Robert Tolksdorf (Hrsg.): XMIDX 2003 – XML-Technologien für Middleware – Middleware für XML-Anwendungen

P-25 Key Pousttchi, Klaus Turowski (Hrsg.): Mobile Commerce – Anwendungen und Perspektiven – 3. Workshop Mobile Commerce, Universität Augsburg, 04.02.2003

P-26 Gerhard Weikum, Harald Schöning, Erhard Rahm (Hrsg.): BTW 2003: Datenbanksysteme für Business, Technologie und Web

P-27 Michael Kroll, Hans-Gerd Lipinski, Kay Melzer (Hrsg.): Mobiles Computing in der Medizin

P-28 Ulrich Reimer, Andreas Abecker, Steffen Staab, Gerd Stumme (Hrsg.): WM 2003: Professionelles Wissensmanagement – Er-fahrungen und Visionen

P-29 Antje Düsterhöft, Bernhard Thalheim (Eds.): NLDB'2003: Natural Language Processing and Information Systems

P-30 Mikhail Godlevsky, Stephen Liddle, Heinrich C. Mayr (Eds.): Information Systems Technology and its Applications

P-31 Arslan Brömme, Christoph Busch (Eds.): BIOSIG 2003: Biometrics and Electronic Signatures, Proceedings of the Special Interest Group on Biometrics and Electronic Signatures, 24. July 2003 in Darmstadt, Germany

P-32 Peter Hubwieser (Hrsg.): Informatische Fachkonzepte im Unterricht – INFOS 2003

P-33 Andreas Geyer-Schulz, Alfred Taudes (Hrsg.): Informationswirtschaft: Ein Sektor mit Zukunft

P-34 Klaus Dittrich, Wolfgang König, Andreas Oberweis, Kai Rannenberg, Wolfgang Wahlster (Hrsg.): Informatik 2003 – Innovative Informatikanwendungen (Band 1)

P-35 Klaus Dittrich, Wolfgang König, Andreas Oberweis, Kai Rannenberg, Wolfgang Wahlster (Hrsg.): Informatik 2003 – Innovative Informatikanwendungen (Band 2)

P-36 Rüdiger Grimm, Hubert B. Keller, Kai Rannenberg (Hrsg.): Informatik 2003 – Mit Sicherheit Informatik

P-37 Arndt Bode, Jörg Desel, Sabine Rathmayer, Martin Wessner (Hrsg.): DeLFI 2003: e-Learning Fachtagung Informatik

P-38 E.J. Sinz, M. Plaha, P. Neckel (Hrsg.): Modellierung betrieblicher Informationssysteme – MobIS 2003

P-39 Jens Nedon, Sandra Frings, Oliver Göbel (Hrsg.): IT-Incident Management & IT-Forensics – IMF 2003

P-40 Michael Rebstock (Hrsg.): Modellierung betrieblicher Informationssysteme – MobIS 2004

P-41 Uwe Brinkschulte, Jürgen Becker, Dietmar Fey, Karl-Erwin Großpietsch, Christian Hochberger, Erik Maehle, Thomas Runkler (Edts.): ARCS 2004 – Organic and Pervasive Computing

P-42 Key Pousttchi, Klaus Turowski (Hrsg.): Mobile Economy – Transaktionen und Prozesse, Anwendungen und Dienste

P-43 Birgitta König-Ries, Michael Klein, Philipp Obreiter (Hrsg.): Persistance, Scalability, Transactions – Database Mechanisms for Mobile Applications

P-44 Jan von Knop, Wilhelm Haverkamp, Eike Jessen (Hrsg.): Security, E-Learning. E-Services

P-45 Bernhard Rumpe, Wofgang Hesse (Hrsg.): Modellierung 2004

P-46 Ulrich Flegel, Michael Meier (Hrsg.): Detection of Intrusions of Malware & Vulnerability Assessment

P-47 Alexander Prosser, Robert Krimmer (Hrsg.): Electronic Voting in Europe – Technology, Law, Politics and Society

P-48 Anatoly Doroshenko, Terry Halpin, Stephen W. Liddle, Heinrich C. Mayr (Hrsg.): Information Systems Technology and its Applications

P-49 G. Schiefer, P. Wagner, M. Morgenstern, U. Rickert (Hrsg.): Integration und Datensicherheit – Anforderungen, Konflikte und Perspektiven

P-50 Peter Dadam, Manfred Reichert (Hrsg.): INFORMATIK 2004 – Informatik verbindet (Band 1) Beiträge der 34. Jahrestagung der Gesellschaft für Informatik e.V. (GI), 20.-24. September 2004 in Ulm

P-51 Peter Dadam, Manfred Reichert (Hrsg.): INFORMATIK 2004 – Informatik verbindet (Band 2) Beiträge der 34. Jahrestagung der Gesellschaft für Informatik e.V. (GI), 20.-24. September 2004 in Ulm

P-52 Gregor Engels, Silke Seehusen (Hrsg.): DELFI 2004 – Tagungsband der 2. e-Learning Fachtagung Informatik

P-53 Robert Giegerich, Jens Stoye (Hrsg.): German Conference on Bioinformatics – GCB 2004

P-54 Jens Borchers, Ralf Kneuper (Hrsg.): Softwaremanagement 2004 – Outsourcing und Integration

P-55 Jan von Knop, Wilhelm Haverkamp, Eike Jessen (Hrsg.): E-Science und Grid Adhoc-Netze Medienintegration

P-56 Fernand Feltz, Andreas Oberweis, Benoit Otjacques (Hrsg.): EMISA 2004 – Informationssysteme im E-Business und E-Government

P-57 Klaus Turowski (Hrsg.): Architekturen, Komponenten, Anwendungen

P-58 Sami Beydeda, Volker Gruhn, Johannes Mayer, Ralf Reussner, Franz Schweiggert (Hrsg.): Testing of Component-Based Systems and Software Quality

P-59 J. Felix Hampe, Franz Lehner, Key Pousttchi, Kai Ranneberg, Klaus Turowski (Hrsg.): Mobile Business – Processes, Platforms, Payments

P-60 Steffen Friedrich (Hrsg.): Unterrichtskonzepte für inforrmatische Bildung

P-61 Paul Müller, Reinhard Gotzhein, Jens B. Schmitt (Hrsg.): Kommunikation in verteilten Systemen

P-62 Federrath, Hannes (Hrsg.): „Sicherheit 2005“ – Sicherheit – Schutz und Zuverlässigkeit

P-63 Roland Kaschek, Heinrich C. Mayr, Stephen Liddle (Hrsg.): Information Systems – Technology and ist Applications

P-64 Peter Liggesmeyer, Klaus Pohl, Michael Goedicke (Hrsg.): Software Engineering 2005

P-65 Gottfried Vossen, Frank Leymann, Peter Lockemann, Wolffried Stucky (Hrsg.): Datenbanksysteme in Business, Technologie und Web

P-66 Jörg M. Haake, Ulrike Lucke, Djamshid Tavangarian (Hrsg.): DeLFI 2005: 3. deutsche e-Learning Fachtagung Informatik

P-67 Armin B. Cremers, Rainer Manthey, Peter Martini, Volker Steinhage (Hrsg.): INFORMATIK 2005 – Informatik LIVE (Band 1)

P-68 Armin B. Cremers, Rainer Manthey, Peter Martini, Volker Steinhage (Hrsg.): INFORMATIK 2005 – Informatik LIVE (Band 2)

P-69 Robert Hirschfeld, Ryszard Kowalcyk, Andreas Polze, Matthias Weske (Hrsg.): NODe 2005, GSEM 2005

P-70 Klaus Turowski, Johannes-Maria Zaha (Hrsg.): Component-oriented Enterprise Application (COAE 2005)

P-71 Andrew Torda, Stefan Kurz, Matthias Rarey (Hrsg.): German Conference on Bioinformatics 2005

P-72 Klaus P. Jantke, Klaus-Peter Fähnrich, Wolfgang S. Wittig (Hrsg.): Marktplatz Internet: Von e-Learning bis e-Payment

P-73 Jan von Knop, Wilhelm Haverkamp, Eike Jessen (Hrsg.): “Heute schon das Morgen sehen“

P-74 Christopher Wolf, Stefan Lucks, Po-Wah Yau (Hrsg.): WEWoRC 2005 – Western European Workshop on Research in Cryptology

P-75 Jörg Desel, Ulrich Frank (Hrsg.): Enterprise Modelling and Information Systems Architecture

P-76 Thomas Kirste, Birgitta König-Riess, Key Pousttchi, Klaus Turowski (Hrsg.): Mobile Informationssysteme – Potentiale, Hindernisse, Einsatz

P-77 Jana Dittmann (Hrsg.): SICHERHEIT 2006

P-78 K.-O. Wenkel, P. Wagner, M. Morgenstern, K. Luzi, P. Eisermann (Hrsg.): Land- und Ernährungswirtschaft im Wandel

P-79 Bettina Biel, Matthias Book, Volker Gruhn (Hrsg.): Softwareengineering 2006

P-80 Mareike Schoop, Christian Huemer, Michael Rebstock, Martin Bichler (Hrsg.): Service-Oriented Electronic Commerce

P-81 Wolfgang Karl, Jürgen Becker, Karl-Erwin Großpietsch, Christian Hochberger, Erik Maehle (Hrsg.): ARCS´06

P-82 Heinrich C. Mayr, Ruth Breu (Hrsg.): Modellierung 2006

P-83 Daniel Huson, Oliver Kohlbacher, Andrei Lupas, Kay Nieselt and Andreas Zell (eds.): German Conference on Bioinformatics

P-84 Dimitris Karagiannis, Heinrich C. Mayr, (Hrsg.): Information Systems Technology and its Applications

P-85 Witold Abramowicz, Heinrich C. Mayr, (Hrsg.): Business Information Systems

P-86 Robert Krimmer (Ed.): Electronic Voting 2006

P-87 Max Mühlhäuser, Guido Rößling, Ralf Steinmetz (Hrsg.): DELFI 2006: 4. e-Learning Fachtagung Informatik

P-88 Robert Hirschfeld, Andreas Polze, Ryszard Kowalczyk (Hrsg.): NODe 2006, GSEM 2006

P-90 Joachim Schelp, Robert Winter, Ulrich Frank, Bodo Rieger, Klaus Turowski (Hrsg.): Integration, Informationslogistik und Architektur

P-91 Henrik Stormer, Andreas Meier, Michael Schumacher (Eds.): European Conference on eHealth 2006

P-92 Fernand Feltz, Benoît Otjacques, Andreas Oberweis, Nicolas Poussing (Eds.): AIM 2006

P-93 Christian Hochberger, Rüdiger Liskowsky (Eds.): INFORMATIK 2006 – Informatik für Menschen, Band 1

P-94 Christian Hochberger, Rüdiger Liskowsky (Eds.): INFORMATIK 2006 – Informatik für Menschen, Band 2

P-95 Matthias Weske, Markus Nüttgens (Eds.): EMISA 2005: Methoden, Konzepte und Technologien für die Entwicklung von dienstbasierten Informationssystemen

P-96 Saartje Brockmans, Jürgen Jung, York Sure (Eds.): Meta-Modelling and Ontologies

P-97 Oliver Göbel, Dirk Schadt, Sandra Frings, Hardo Hase, Detlef Günther, Jens Nedon (Eds.): IT-Incident Mangament & IT-Forensics – IMF 2006

P-98 Hans Brandt-Pook, Werner Simonsmeier und Thorsten Spitta (Hrsg.): Beratung in der Softwareentwicklung – Modelle, Methoden, Best Practices

P-99 Andreas Schwill, Carsten Schulte, Marco Thomas (Hrsg.): Didaktik der Informatik

P-100 Peter Forbrig, Günter Siegel, Markus Schneider (Hrsg.): HDI 2006: Hochschuldidaktik der Informatik

P-101 Stefan Böttinger, Ludwig Theuvsen, Susanne Rank, Marlies Morgenstern (Hrsg.): Agrarinformatik im Spannungsfeld zwischen Regionalisierung und globalen Wertschöpfungsketten

P-102 Otto Spaniol (Eds.): Mobile Services and Personalized Environments

P-103 Alfons Kemper, Harald Schöning, Thomas Rose, Matthias Jarke, Thomas Seidl, Christoph Quix, Christoph Brochhaus (Hrsg.): Datenbanksysteme in Business, Technologie und Web (BTW 2007)

P-104 Birgitta König-Ries, Franz Lehner, Rainer Malaka, Can Türker (Hrsg.) MMS 2007: Mobilität und mobile Informationssysteme

P-105 Wolf-Gideon Bleek, Jörg Raasch, Heinz Züllighoven (Hrsg.) Software Engineering 2007

P-106 Wolf-Gideon Bleek, Henning Schwentner, Heinz Züllighoven (Hrsg.) Software Engineering 2007 – Beiträge zu den Workshops

P-107 Heinrich C. Mayr, Dimitris Karagiannis (eds.) Information Systems Technology and its Applications

P-108 Arslan Brömme, Christoph Busch, Detlef Hühnlein (eds.) BIOSIG 2007: Biometrics and Electronic Signatures

P-109 Rainer Koschke, Otthein Herzog, Karl-Heinz Rödiger, Marc Ronthaler (Hrsg.) INFORMATIK 2007 Informatik trifft Logistik Band 1

P-110 Rainer Koschke, Otthein Herzog, Karl-Heinz Rödiger, Marc Ronthaler (Hrsg.) INFORMATIK 2007 Informatik trifft Logistik Band 2

P-111 Christian Eibl, Johannes Magenheim, Sigrid Schubert, Martin Wessner (Hrsg.) DeLFI 2007: 5. e-Learning Fachtagung Informatik

P-112 Sigrid Schubert (Hrsg.) Didaktik der Informatik in Theorie und Praxis

P-113 Sören Auer, Christian Bizer, Claudia Müller, Anna V. Zhdanova (Eds.) The Social Semantic Web 2007 Proceedings of the 1st Conference on Social Semantic Web (CSSW)

P-114 Sandra Frings, Oliver Göbel, Detlef Günther, Hardo G. Hase, Jens Nedon, Dirk Schadt, Arslan Brömme (Eds.) IMF2007 IT-incident management & IT-forensics Proceedings of the 3rd International Conference on IT-Incident Management & IT-Forensics

P-115 Claudia Falter, Alexander Schliep, Joachim Selbig, Martin Vingron and Dirk Walther (Eds.) German conference on bioinformatics GCB 2007

P-116 Witold Abramowicz, Leszek Maciszek (Eds.) Business Process and Services Computing 1st International Working Conference on Business Process and Services Computing BPSC 2007

P-117 Ryszard Kowalczyk (Ed.) Grid service engineering and manegement The 4th International Conference on Grid Service Engineering and Management GSEM 2007

P-118 Andreas Hein, Wilfried Thoben, Hans-Jürgen Appelrath, Peter Jensch (Eds.) European Conference on ehealth 2007

P-119 Manfred Reichert, Stefan Strecker, Klaus Turowski (Eds.) Enterprise Modelling and Information Systems Architectures Concepts and Applications

P-120 Adam Pawlak, Kurt Sandkuhl, Wojciech Cholewa, Leandro Soares Indrusiak (Eds.) Coordination of Collaborative Engineering - State of the Art and Future Challenges

P-121 Korbinian Herrmann, Bernd Bruegge (Hrsg.) Software Engineering 2008 Fachtagung des GI-Fachbereichs Softwaretechnik

P-122 Walid Maalej, Bernd Bruegge (Hrsg.) Software Engineering 2008 - Workshopband Fachtagung des GI-Fachbereichs Softwaretechnik

P-123 Michael H. Breitner, Martin Breunig, Elgar Fleisch, Ley Pousttchi, Klaus Turowski (Hrsg.)
Mobile und Ubiquitäre Informationssysteme – Technologien, Prozesse, Marktfähigkeit
Proceedings zur 3. Konferenz Mobile und Ubiquitäre Informationssysteme (MMS 2008)

P-124 Wolfgang E. Nagel, Rolf Hoffmann, Andreas Koch (Eds.)
9th Workshop on Parallel Systems and Algorithms (PASA)
Workshop of the GI/ITG Speciel Interest Groups PARS and PARVA

P-125 Rolf A.E. Müller, Hans-H. Sundermeier, Ludwig Theuvsen, Stephanie Schütze, Marlies Morgenstern (Hrsg.)
Unternehmens-IT:
Führungsinstrument oder Verwaltungsbürde
Referate der 28. GIL Jahrestagung

P-126 Rainer Gimnich, Uwe Kaiser, Jochen Quante, Andreas Winter (Hrsg.)
10th Workshop Software Reengineering (WSR 2008)

P-127 Thomas Kühne, Wolfgang Reisig, Friedrich Steimann (Hrsg.)
Modellierung 2008

P-128 Ammar Alkassar, Jörg Siekmann (Hrsg.)
Sicherheit 2008
Sicherheit, Schutz und Zuverlässigkeit
Beiträge der 4. Jahrestagung des Fachbereichs Sicherheit der Gesellschaft für Informatik e.V. (GI)
2.-4. April 2008
Saarbrücken, Germany

P-129 Wolfgang Hesse, Andreas Oberweis (Eds.)
Sigsand-Europe 2008
Proceedings of the Third AIS SIGSAND European Symposium on Analysis, Design, Use and Societal Impact of Information Systems

P-130 Paul Müller, Bernhard Neumair, Gabi Dreo Rodosek (Hrsg.)
1. DFN-Forum Kommunikations-technologien Beiträge der Fachtagung

P-131 Robert Krimmer, Rüdiger Grimm (Eds.)
3rd International Conference on Electronic Voting 2008
Co-organized by Council of Europe, Gesellschaft für Informatik and E-Voting. CC

P-132 Silke Seehusen, Ulrike Lucke, Stefan Fischer (Hrsg.)
DeLFI 2008:
Die 6. e-Learning Fachtagung Informatik

P-133 Heinz-Gerd Hegering, Axel Lehmann, Hans Jürgen Ohlbach, Christian Scheideler (Hrsg.)
INFORMATIK 2008
Beherrschbare Systeme – dank Informatik
Band 1

P-134 Heinz-Gerd Hegering, Axel Lehmann, Hans Jürgen Ohlbach, Christian Scheideler (Hrsg.)
INFORMATIK 2008
Beherrschbare Systeme – dank Informatik
Band 2

P-135 Torsten Brinda, Michael Fothe, Peter Hubwieser, Kirsten Schlüter (Hrsg.)
Didaktik der Informatik –
Aktuelle Forschungsergebnisse

P-136 Andreas Beyer, Michael Schroeder (Eds.)
German Conference on Bioinformatics GCB 2008

P-137 Arslan Brömme, Christoph Busch, Detlef Hühnlein (Eds.)
BIOSIG 2008: Biometrics and Electronic Signatures

P-138 Barbara Dinter, Robert Winter, Peter Chamoni, Norbert Gronau, Klaus Turowski (Hrsg.)
Synergien durch Integration und Informationslogistik
Proceedings zur DW2008

P-139 Georg Herzwurm, Martin Mikusz (Hrsg.)
Industrialisierung des Software-Managements
Fachtagung des GI-Fachausschusses Management der Anwendungsentwicklung und -wartung im Fachbereich Wirtschaftsinformatik

P-140 Oliver Göbel, Sandra Frings, Detlef Günther, Jens Nedon, Dirk Schadt (Eds.)
IMF 2008 - IT Incident Management & IT Forensics

P-141 Peter Loos, Markus Nüttgens, Klaus Turowski, Dirk Werth (Hrsg.)
Modellierung betrieblicher Informations-systeme (MobIS 2008)
Modellierung zwischen SOA und Compliance Management

P-142 R. Bill, P. Korduan, L. Theuvsen, M. Morgenstern (Hrsg.)
Anforderungen an die Agrarinformatik durch Globalisierung und Klimaveränderung

P-143 Peter Liggesmeyer, Gregor Engels, Jürgen Münch, Jörg Dörr, Norman Riegel (Hrsg.)
Software Engineering 2009
Fachtagung des GI-Fachbereichs Softwaretechnik

P-144 Johann-Christoph Freytag, Thomas Ruf, Wolfgang Lehner, Gottfried Vossen (Hrsg.)
Datenbanksysteme in Business, Technologie und Web (BTW)

P-145 Knut Hinkelmann, Holger Wache (Eds.)
WM2009: 5th Conference on Professional Knowledge Management

P-146 Markus Bick, Martin Breunig, Hagen Höpfner (Hrsg.)
Mobile und Ubiquitäre Informationssysteme – Entwicklung, Implementierung und Anwendung
4. Konferenz Mobile und Ubiquitäre Informationssysteme (MMS 2009)

P-147 Witold Abramowicz, Leszek Maciaszek, Ryszard Kowalczyk, Andreas Speck (Eds.)
Business Process, Services Computing and Intelligent Service Management
BPSC 2009 · ISM 2009 · YRW-MBP 2009

P-148 Christian Erfurth, Gerald Eichler, Volkmar Schau (Eds.)
9th International Conference on Innovative Internet Community Systems
I^{2}CS 2009

P-149 Paul Müller, Bernhard Neumair, Gabi Dreo Rodosek (Hrsg.)
2. DFN-Forum Kommunikationstechnologien
Beiträge der Fachtagung

P-150 Jürgen Münch, Peter Liggesmeyer (Hrsg.)
Software Engineering
2009 - Workshopband

P-151 Armin Heinzl, Peter Dadam, Stefan Kirn, Peter Lockemann (Eds.)
PRIMIUM
Process Innovation for Enterprise Software

P-152 Jan Mendling, Stefanie Rinderle-Ma, Werner Esswein (Eds.)
Enterprise Modelling and Information Systems Architectures
Proceedings of the 3rd Int'l Workshop EMISA 2009

P-153 Andreas Schwill, Nicolas Apostolopoulos (Hrsg.)
Lernen im Digitalen Zeitalter
DeLFI 2009 – Die 7. E-Learning Fachtagung Informatik

P-154 Stefan Fischer, Erik Maehle Rüdiger Reischuk (Hrsg.)
INFORMATIK 2009
Im Focus das Leben

P-155 Arslan Brömme, Christoph Busch, Detlef Hühnlein (Eds.)
BIOSIG 2009:
Biometrics and Electronic Signatures
Proceedings of the Special Interest Group on Biometrics and Electronic Signatures

P-156 Bernhard Koerber (Hrsg.)
Zukunft braucht Herkunft
25 Jahre »INFOS – Informatik und Schule«

P-157 Ivo Grosse, Steffen Neumann, Stefan Posch, Falk Schreiber, Peter Stadler (Eds.)
German Conference on Bioinformatics 2009

P-158 W. Claupein, L. Theuvsen, A. Kämpf, M. Morgenstern (Hrsg.)
Precision Agriculture Reloaded – Informationsgestützte Landwirtschaft

P-159 Gregor Engels, Markus Luckey, Wilhelm Schäfer (Hrsg.)
Software Engineering 2010

P-160 Gregor Engels, Markus Luckey, Alexander Pretschner, Ralf Reussner (Hrsg.)
Software Engineering 2010 – Workshopband
(inkl. Doktorandensymposium)

P-161 Gregor Engels, Dimitris Karagiannis Heinrich C. Mayr (Hrsg.)
Modellierung 2010

P-162 Maria A. Wimmer, Uwe Brinkhoff, Siegfried Kaiser, Dagmar Lück-Schneider, Erich Schweighofer, Andreas Wiebe (Hrsg.)
Vernetzte IT für einen effektiven Staat
Gemeinsame Fachtagung Verwaltungsinformatik (FTVI) und Fachtagung Rechtsinformatik (FTRI) 2010

P-163 Markus Bick, Stefan Eulgem, Elgar Fleisch, J. Felix Hampe, Birgitta König-Ries, Franz Lehner, Key Pousttchi, Kai Rannenberg (Hrsg.)
Mobile und Ubiquitäre Informationssysteme
Technologien, Anwendungen und Dienste zur Unterstützung von mobiler Kollaboration

P-164 Arslan Brömme, Christoph Busch (Eds.)
BIOSIG 2010: Biometrics and Electronic Signatures Proceedings of the Special Interest Group on Biometrics and Electronic Signatures

P-165 Gerald Eichler, Peter Kropf, Ulrike Lechner, Phayung Meesad, Herwig Unger (Eds.)
10th International Conference on Innovative Internet Community Systems (I^2CS) – Jubilee Edition 2010 –

P-166 Paul Müller, Bernhard Neumair, Gabi Dreo Rodosek (Hrsg.)
3. DFN-Forum Kommunikationstechnologien
Beiträge der Fachtagung

P-167 Robert Krimmer, Rüdiger Grimm (Eds.)
4th International Conference on Electronic Voting 2010
co-organized by the Council of Europe, Gesellschaft für Informatik and E-Voting.CC

P-168 Ira Diethelm, Christina Dörge, Claudia Hildebrandt, Carsten Schulte (Hrsg.)
Didaktik der Informatik
Möglichkeiten empirischer Forschungsmethoden und Perspektiven der Fachdidaktik

P-169 Michael Kerres, Nadine Ojstersek Ulrik Schroeder, Ulrich Hoppe (Hrsg.)
DeLFI 2010 - 8. Tagung der Fachgruppe E-Learning der Gesellschaft für Informatik e.V.

P-170 Felix C. Freiling (Hrsg.)
Sicherheit 2010
Sicherheit, Schutz und Zuverlässigkeit

P-171 Werner Esswein, Klaus Turowski, Martin Juhrisch (Hrsg.)
Modellierung betrieblicher Informationssysteme (MobIS 2010)
Modellgestütztes Management

P-172 Stefan Klink, Agnes Koschmider Marco Mevius, Andreas Oberweis (Hrsg.)
EMISA 2010
Einflussfaktoren auf die Entwicklung flexibler, integrierter Informationssysteme
Beiträge des Workshops der GI-Fachgruppe EMISA (Entwicklungsmethoden für Informationssysteme und deren Anwendung)

P-173 Dietmar Schomburg, Andreas Grote (Eds.)
German Conference on Bioinformatics 2010

P-174 Arslan Brömme, Torsten Eymann, Detlef Hühnlein, Heiko Roßnagel, Paul Schmücker (Hrsg.)
perspeGKtive 2010
Workshop „Innovative und sichere Informationstechnologie für das Gesundheitswesen von morgen“

P-175 Klaus-Peter Fähnrich, Bogdan Franczyk (Hrsg.)
INFORMATIK 2010
Service Science – Neue Perspektiven für die Informatik
Band 1

P-176 Klaus-Peter Fähnrich, Bogdan Franczyk (Hrsg.)
INFORMATIK 2010
Service Science – Neue Perspektiven für die Informatik
Band 2

P-177 Witold Abramowicz, Rainer Alt, Klaus-Peter Fähnrich, Bogdan Franczyk, Leszek A. Maciaszek (Eds.)
INFORMATIK 2010
Business Process and Service Science – Proceedings of ISSS and BPSC

P-178 Wolfram Pietsch, Benedikt Krams (Hrsg.)
Vom Projekt zum Produkt
Fachtagung des GI-Fachausschusses Management der Anwendungsentwicklung und -wartung im Fachbereich Wirtschafts-informatik (WI-MAW), Aachen, 2010

P-179 Stefan Gruner, Bernhard Rumpe (Eds.)
FM+AM`2010
Second International Workshop on Formal Methods and Agile Methods

P-180 Theo Härder, Wolfgang Lehner, Bernhard Mitschang, Harald Schöning, Holger Schwarz (Hrsg.)
Datenbanksysteme für Business, Technologie und Web (BTW)
14. Fachtagung des GI-Fachbereichs „Datenbanken und Informationssysteme“ (DBIS)

P-181 Michael Clasen, Otto Schätzel, Brigitte Theuvsen (Hrsg.)
Qualität und Effizienz durch informationsgestützte Landwirtschaft, Fokus: Moderne Weinwirtschaft

P-182 Ronald Maier (Hrsg.)
6th Conference on Professional Knowledge Management
From Knowledge to Action

P-183 Ralf Reussner, Matthias Grund, Andreas Oberweis, Walter Tichy (Hrsg.)
Software Engineering 2011
Fachtagung des GI-Fachbereichs Softwaretechnik

P-184 Ralf Reussner, Alexander Pretschner, Stefan Jähnichen (Hrsg.)
Software Engineering 2011
Workshopband
(inkl. Doktorandensymposium)

P-185 Hagen Höpfner, Günther Specht, Thomas Ritz, Christian Bunse (Hrsg.)
MMS 2011: Mobile und ubiquitäre Informationssysteme Proceedings zur 6. Konferenz Mobile und Ubiquitäre Informationssysteme (MMS 2011)

P-186 Gerald Eichler, Axel Küpper, Volkmar Schau, Hacène Fouchal, Herwig Unger (Eds.)
11th International Conference on Innovative Internet Community Systems (I^2CS)

P-187 Paul Müller, Bernhard Neumair, Gabi Dreo Rodosek (Hrsg.)
4. DFN-Forum Kommunikations-technologien, Beiträge der Fachtagung 20. Juni bis 21. Juni 2011 Bonn

P-188 Holger Rohland, Andrea Kienle, Steffen Friedrich (Hrsg.)
DeLFI 2011 – Die 9. e-Learning Fachtagung Informatik der Gesellschaft für Informatik e.V. 5.–8. September 2011, Dresden

P-189 Thomas, Marco (Hrsg.)
Informatik in Bildung und Beruf
INFOS 2011
14. GI-Fachtagung Informatik und Schule

P-190 Markus Nüttgens, Oliver Thomas, Barbara Weber (Eds.)
Enterprise Modelling and Information Systems Architectures (EMISA 2011)

P-191 Arslan Brömme, Christoph Busch (Eds.)
BIOSIG 2011
International Conference of the Biometrics Special Interest Group

P-192 Hans-Ulrich Heiß, Peter Pepper, Holger Schlingloff, Jörg Schneider (Hrsg.)
INFORMATIK 2011
Informatik schafft Communities

P-193 Wolfgang Lehner, Gunther Piller (Hrsg.)
IMDM 2011

P-194 M. Clasen, G. Fröhlich, H. Bernhardt, K. Hildebrand, B. Theuvsen (Hrsg.)
Informationstechnologie für eine nachhaltige Landbewirtschaftung
Fokus Forstwirtschaft

P-195 Neeraj Suri, Michael Waidner (Hrsg.)
Sicherheit 2012
Sicherheit, Schutz und Zuverlässigkeit
Beiträge der 6. Jahrestagung des Fachbereichs Sicherheit der Gesellschaft für Informatik e.V. (GI)

P-196 Arslan Brömme, Christoph Busch (Eds.)
BIOSIG 2012
Proceedings of the 11th International Conference of the Biometrics Special Interest Group

P-197 Jörn von Lucke, Christian P. Geiger, Siegfried Kaiser, Erich Schweighofer, Maria A. Wimmer (Hrsg.)
Auf dem Weg zu einer offenen, smarten und vernetzten Verwaltungskultur
Gemeinsame Fachtagung Verwaltungsinformatik (FTVI) und Fachtagung Rechtsinformatik (FTRI) 2012

P-198 Stefan Jähnichen, Axel Küpper, Sahin Albayrak (Hrsg.)
Software Engineering 2012
Fachtagung des GI-Fachbereichs Softwaretechnik

P-199 Stefan Jähnichen, Bernhard Rumpe, Holger Schlingloff (Hrsg.)
Software Engineering 2012
Workshopband

P-200 Gero Mühl, Jan Richling, Andreas Herkersdorf (Hrsg.)
ARCS 2012 Workshops

P-201 Elmar J. Sinz Andy Schürr (Hrsg.)
Modellierung 2012

P-202 Andrea Back, Markus Bick, Martin Breunig, Key Pousttchi, Frédéric Thiesse (Hrsg.)
MMS 2012:Mobile und Ubiquitäre Informationssysteme

P-203 Paul Müller, Bernhard Neumair, Helmut Reiser, Gabi Dreo Rodosek (Hrsg.)
5. DFN-Forum Kommunikations-technologien
Beiträge der Fachtagung

P-204 Gerald Eichler, Leendert W. M. Wienhofen, Anders Kofod-Petersen, Herwig Unger (Eds.)
12th International Conference on Innovative Internet Community Systems (I2CS 2012)

P-205 Manuel J. Kripp, Melanie Volkamer, Rüdiger Grimm (Eds.)
5th International Conference on Electronic Voting 2012 (EVOTE2012)
Co-organized by the Council of Europe, Gesellschaft für Informatik and E-Voting.CC

P-206 Stefanie Rinderle-Ma, Mathias Weske (Hrsg.)
EMISA 2012
Der Mensch im Zentrum der Modellierung

P-207 Jörg Desel, Jörg M. Haake, Christian Spannagel (Hrsg.)
DeLFI 2012: Die 10. e-Learning Fachtagung Informatik der Gesellschaft für Informatik e.V.
24.–26. September 2012

P-208 Ursula Goltz, Marcus Magnor, Hans-Jürgen Appelrath, Herbert Matthies, Wolf-Tilo Balke, Lars Wolf (Hrsg.) INFORMATIK 2012

P-209 Hans Brandt-Pook, André Fleer, Thorsten Spitta, Malte Wattenberg (Hrsg.) Nachhaltiges Software Management

P-210 Erhard Plödereder, Peter Dencker, Herbert Klenk, Hubert B. Keller, Silke Spitzer (Hrsg.) Automotive – Safety & Security 2012 Sicherheit und Zuverlässigkeit für automobile Informationstechnik

P-211 M. Clasen, K. C. Kersebaum, A. Meyer-Aurich, B. Theuvsen (Hrsg.) Massendatenmanagement in der Agrar- und Ernährungswirtschaft Erhebung - Verarbeitung - Nutzung Referate der 33. GIL-Jahrestagung 20. – 21. Februar 2013, Potsdam

P-212 Arslan Brömme, Christoph Busch (Eds.) BIOSIG 2013 Proceedings of the 12th International Conference of the Biometrics Special Interest Group 04.–06. September 2013 Darmstadt, Germany

P-213 Stefan Kowalewski, Bernhard Rumpe (Hrsg.) Software Engineering 2013 Fachtagung des GI-Fachbereichs Softwaretechnik

P-214 Volker Markl, Gunter Saake, Kai-Uwe Sattler, Gregor Hackenbroich, Bernhard Mit schang, Theo Härder, Veit Köppen (Hrsg.) Datenbanksysteme für Business, Technologie und Web (BTW) 2013 13. – 15. März 2013, Magdeburg

P-215 Stefan Wagner, Horst Lichter (Hrsg.) Software Engineering 2013 Workshopband (inkl. Doktorandensymposium) 26. Februar – 1. März 2013, Aachen

P-216 Gunter Saake, Andreas Henrich, Wolfgang Lehner, Thomas Neumann, Veit Köppen (Hrsg.) Datenbanksysteme für Business, Technologie und Web (BTW) 2013 – Workshopband 11. – 12. März 2013, Magdeburg

P-217 Paul Müller, Bernhard Neumair, Helmut Reiser, Gabi Dreo Rodosek (Hrsg.) 6. DFN-Forum Kommunikations- technologien Beiträge der Fachtagung 03.–04. Juni 2013, Erlangen

P-218 Andreas Breiter, Christoph Rensing (Hrsg.) DeLFI 2013: Die 11 e-Learning Fachtagung Informatik der Gesellschaft für Informatik e.V. (GI) 8. – 11. September 2013, Bremen

P-219 Norbert Breier, Peer Stechert, Thomas Wilke (Hrsg.) Informatik erweitert Horizonte INFOS 2013 15. GI-Fachtagung Informatik und Schule 26. – 28. September 2013

P-220 Matthias Horbach (Hrsg.) INFORMATIK 2013 Informatik angepasst an Mensch, Organisation und Umwelt 16. – 20. September 2013, Koblenz

P-221 Maria A. Wimmer, Marijn Janssen, Ann Macintosh, Hans Jochen Scholl, Efthimios Tambouris (Eds.) Electronic Government and Electronic Participation Joint Proceedings of Ongoing Research of IFIP EGOV and IFIP ePart 2013 16. – 19. September 2013, Koblenz

P-222 Reinhard Jung, Manfred Reichert (Eds.) Enterprise Modelling and Information Systems Architectures (EMISA 2013) St. Gallen, Switzerland September 5. – 6. 2013

P-223 Detlef Hühnlein, Heiko Roßnagel (Hrsg.) Open Identity Summit 2013 10. – 11. September 2013 Kloster Banz, Germany

P-224 Eckhart Hanser, Martin Mikusz, Masud Fazal-Baqaie (Hrsg.) Vorgehensmodelle 2013 Vorgehensmodelle – Anspruch und Wirklichkeit 20. Tagung der Fachgruppe Vorgehensmodelle im Fachgebiet Wirtschaftsinformatik (WI-VM) der Gesellschaft für Informatik e.V. Lörrach, 2013

P-225 Hans-Georg Fill, Dimitris Karagiannis, Ulrich Reimer (Hrsg.) Modellierung 2014 19. – 21. März 2014, Wien

P-226 M. Clasen, M. Hamer, S. Lehnert, B. Petersen, B. Theuvsen (Hrsg.) IT-Standards in der Agrar- und Ernährungswirtschaft Fokus: Risiko- und Krisenmanagement Referate der 34. GIL-Jahrestagung 24. – 25. Februar 2014, Bonn

P-227 Wilhelm Hasselbring,
Nils Christian Ehmke (Hrsg.)
Software Engineering 2014
Fachtagung des GI-Fachbereichs Softwaretechnik
25. – 28. Februar 2014
Kiel, Deutschland

P-228 Stefan Katzenbeisser, Volkmar Lotz, Edgar Weippl (Hrsg.)
Sicherheit 2014
Sicherheit, Schutz und Zuverlässigkeit
Beiträge der 7. Jahrestagung des Fachbereichs Sicherheit der Gesellschaft für Informatik e.V. (GI)
19. – 21. März 2014, Wien

P-229 Dagmar Lück-Schneider, Thomas Gordon, Siegfried Kaiser, Jörn von Lucke,Erich Schweighofer, Maria A.Wimmer, Martin G. Löhe (Hrsg.)
Gemeinsam Electronic Government ziel(gruppen)gerecht gestalten und organisieren
Gemeinsame Fachtagung Verwaltungsinformatik (FTVI) und Fachtagung Rechtsinformatik (FTRI)
2014, 20.-21. März 2014 in Berlin

P-230 Arslan Brömme, Christoph Busch (Eds.)
BIOSIG 2014
Proceedings of the 13th International Conference of the Biometrics Special Interest Group
10. – 12. September 2014 in Darmstadt, Germany

P-231 Paul Müller, Bernhard Neumair, Helmut Reiser, Gabi Dreo Rodosek (Hrsg.)
7. DFN-Forum Kommunikationstechnologien
16. – 17. Juni 2014
Fulda

P-232 E. Plödereder, L. Grunske, E. Schneider, D. Ull (Hrsg.)
INFORMATIK 2014
Big Data – Komplexität meistern
22. – 26. September 2014
Stuttgart

P-233 Stephan Trahasch, Rolf Plötzner, Gerhard Schneider, Claudia Gayer, Daniel Sassiat, Nicole Wöhrle (Hrsg.)
DeLFI 2014 – Die 12. e-Learning Fachtagung Informatik
der Gesellschaft für Informatik e.V.
15. – 17. September 2014
Freiburg

P-234 Fernand Feltz, Bela Mutschler, Benoît Otjacques (Eds.)
Enterprise Modelling and Information Systems Architectures
(EMISA 2014)
Luxembourg, September 25-26, 2014

P-235 Robert Giegerich, Ralf Hofestädt, Tim W. Nattkemper (Eds.)
German Conference on Bioinformatics 2014
September 28 – October 1
Bielefeld, Germany

P-236 Martin Engstler, Eckhart Hanser, Martin Mikusz, Georg Herzwurm (Hrsg.)
Projektmanagement und Vorgehensmodelle 2014
Soziale Aspekte und Standardisierung
Gemeinsame Tagung der Fachgruppen Projektmanagement (WI-PM) und Vorgehensmodelle (WI-VM) im Fachgebiet Wirtschaftsinformatik der Gesellschaft für Informatik e.V., Stuttgart 2014

P-237 Detlef Hühnlein, Heiko Roßnagel (Hrsg.)
Open Identity Summit 2014
4.–6. November 2014
Stuttgart, Germany

P-238 Arno Ruckelshausen, Hans-Peter Schwarz, Brigitte Theuvsen (Hrsg.)
Informatik in der Land-, Forst- und Ernährungswirtschaft
Referate der 35. GIL-Jahrestagung
23. – 24. Februar 2015, Geisenheim

P-239 Uwe Aßmann, Birgit Demuth, Thorsten Spitta, Georg Püschel, Ronny Kaiser (Hrsg.)
Software Engineering & Management 2015
17.-20. März 2015, Dresden

P-240 Herbert Klenk, Hubert B. Keller, Erhard Plödereder, Peter Dencker (Hrsg.)
Automotive – Safety & Security 2015
Sicherheit und Zuverlässigkeit für automobile Informationstechnik
21.–22. April 2015, Stuttgart

P-241 Thomas Seidl, Norbert Ritter, Harald Schöning, Kai-Uwe Sattler, Theo Härder, Steffen Friedrich, Wolfram Wingerath (Hrsg.)
Datenbanksysteme für Business, Technologie und Web (BTW 2015)
04. – 06. März 2015, Hamburg

P-242 Norbert Ritter, Andreas Henrich, Wolfgang Lehner, Andreas Thor, Steffen Friedrich, Wolfram Wingerath (Hrsg.)
Datenbanksysteme für Business, Technologie und Web (BTW 2015) – Workshopband
02. – 03. März 2015, Hamburg

P-243 Paul Müller, Bernhard Neumair, Helmut Reiser, Gabi Dreo Rodosek (Hrsg.)
8. DFN-Forum Kommunikationstechnologien
06.–09. Juni 2015, Lübeck

P-244 Alfred Zimmermann, Alexander Rossmann (Eds.)
Digital Enterprise Computing (DEC 2015)
Böblingen, Germany June 25-26, 2015

P-245 Arslan Brömme, Christoph Busch , Christian Rathgeb, Andreas Uhl (Eds.)
BIOSIG 2015
Proceedings of the 14th International Conference of the Biometrics Special Interest Group
09.–11. September 2015
Darmstadt, Germany

P-246 Douglas W. Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.)
INFORMATIK 2015
28.9.-2.10. 2015, Cottbus

P-247 Hans Pongratz, Reinhard Keil (Hrsg.)
DeLFI 2015 – Die 13. E-Learning Fachtagung Informatik der Gesellschaft für Informatik e.V. (GI)
1.–4. September 2015
München

P-248 Jens Kolb, Henrik Leopold, Jan Mendling (Eds.)
Enterprise Modelling and Information Systems Architectures
Proceedings of the 6th Int. Workshop on Enterprise Modelling and Information Systems Architectures, Innsbruck, Austria
September 3-4, 2015

P-249 Jens Gallenbacher (Hrsg.)
Informatik allgemeinbildend begreifen
INFOS 2015 16. GI-Fachtagung Informatik und Schule
20.–23. September 2015

P-250 Martin Engstler, Masud Fazal-Baqaie, Eckhart Hanser, Martin Mikusz, Alexander Volland (Hrsg.)
Projektmanagement und Vorgehensmodelle 2015
Hybride Projektstrukturen erfolgreich umsetzen
Gemeinsame Tagung der Fachgruppen Projektmanagement (WI-PM) und Vorgehensmodelle (WI-VM) im Fachgebiet Wirtschaftsinformatik der Gesellschaft für Informatik e.V., Elmshorn 2015

P-251 Detlef Hühnlein, Heiko Roßnagel, Raik Kuhlisch, Jan Ziesing (Eds.)
Open Identity Summit 2015
10.–11. November 2015
Berlin, Germany

P-252 Jens Knoop, Uwe Zdun (Hrsg.)
Software Engineering 2016
Fachtagung des GI-Fachbereichs Softwaretechnik
23.–26. Februar 2016, Wien

P-253 A. Ruckelshausen, A. Meyer-Aurich, T. Rath, G. Recke, B. Theuvsen (Hrsg.)
Informatik in der Land-, Forst- und Ernährungswirtschaft
Fokus: Intelligente Systeme – Stand der Technik und neue Möglichkeiten
Referate der 36. GIL-Jahrestagung
22.-23. Februar 2016, Osnabrück

P-254 Andreas Oberweis, Ralf Reussner (Hrsg.)
Modellierung 2016
2.–4. März 2016, Karlsruhe

P-255 Stefanie Betz, Ulrich Reimer (Hrsg.)
Modellierung 2016 Workshopband
2.–4. März 2016, Karlsruhe

P-256 Michael Meier, Delphine Reinhardt, Steffen Wendzel (Hrsg.)
Sicherheit 2016
Sicherheit, Schutz und Zuverlässigkeit
Beiträge der 8. Jahrestagung des Fachbereichs Sicherheit der Gesellschaft für Informatik e.V. (GI)
5.–7. April 2016, Bonn

P-257 Paul Müller, Bernhard Neumair, Helmut Reiser, Gabi Dreo Rodosek (Hrsg.)
9. DFN-Forum Kommunikationstechnologien
31. Mai – 01. Juni 2016, Rostock

P-258 Dieter Hertweck, Christian Decker (Eds.)
Digital Enterprise Computing (DEC 2016)
14.–15. Juni 2016, Böblingen

P-259 Heinrich C. Mayr, Martin Pinzger (Hrsg.)
INFORMATIK 2016
26.–30. September 2016, Klagenfurt

P-260 Arslan Brömme, Christoph Busch, Christian Rathgeb, Andreas Uhl (Eds.)
BIOSIG 2016
Proceedings of the 15th International Conference of the Biometrics Special Interest Group
21.–23. September 2016, Darmstadt

P-261 Detlef Rätz, Michael Breidung, Dagmar Lück-Schneider, Siegfried Kaiser, Erich Schweighofer (Hrsg.)
Digitale Transformation: Methoden, Kompetenzen und Technologien für die Verwaltung
Gemeinsame Fachtagung Verwaltungsinformatik (FTVI) und Fachtagung Rechtsinformatik (FTRI) 2016
22.–23. September 2016, Dresden

P-262 Ulrike Lucke, Andreas Schwill, Raphael Zender (Hrsg.)
DeLFI 2016 – Die 14. E-Learning Fachtagung Informatik
der Gesellschaft für Informatik e.V. (GI)
11.–14. September 2016, Potsdam

P-263 Martin Engstler, Masud Fazal-Baqaie, Eckhart Hanser, Oliver Linssen, Martin Mikusz, Alexander Volland (Hrsg.)
Projektmanagement und Vorgehensmodelle 2016
Arbeiten in hybriden Projekten: Das Sowohl-als-auch von Stabilität und Dynamik
Gemeinsame Tagung der Fachgruppen Projektmanagement (WI-PM) und Vorgehensmodelle (WI-VM) im Fachgebiet Wirtschaftsinformatik
der Gesellschaft für Informatik e.V., Paderborn 2016

P-264 Detlef Hühnlein, Heiko Roßnagel, Christian H. Schunck, Maurizio Talamo (Eds.)
Open Identity Summit 2016
der Gesellschaft für Informatik e.V. (GI)
13.–14. October 2016, Rome, Italy

P-265 Bernhard Mitschang, Daniela Nicklas,Frank Leymann, Harald Schöning, Melanie Herschel, Jens Teubner, Theo Härder, Oliver Kopp, Matthias Wieland (Hrsg.)
Datenbanksysteme für Business, Technologie und Web (BTW 2017)
6.–10. März 2017, Stuttgart

P-266 Bernhard Mitschang, Norbert Ritter, Holger Schwarz, Meike Klettke, Andreas Thor, Oliver Kopp, Matthias Wieland (Hrsg.)
Datenbanksysteme für Business, Technologie und Web (BTW 2017)
Workshopband
6.–7. März 2017, Stuttgart

P-267 Jan Jürjens, Kurt Schneider (Hrsg.)
Software Engineering 2017
21.–24. Februar 2017, Hannover

P-268 A. Ruckelshausen, A. Meyer-Aurich, W. Lentz, B. Theuvsen (Hrsg.)
Informatik in der Land-, Forst- und Ernährungswirtschaft
Fokus: Digitale Transformation – Wege in eine zukunftsfähige Landwirtschaft
Referate der 37. GIL-Jahrestagung
06.–07. März 2017, Dresden

P-269 Peter Dencker, Herbert Klenk, Hubert Keller, Erhard Plödereder (Hrsg.)
Automotive – Safety & Security 2017
30.–31. Mai 2017, Stuttgart

P-270 Arslan Brömme, Christoph Busch, Antitza Dantcheva, Christian Rathgeb, Andreas Uhl (Eds.)
BIOSIG 2017
20.–22. September 2017, Darmstadt

P-271 Paul Müller, Bernhard Neumair, Helmut Reiser, Gabi Dreo Rodosek (Hrsg.)
10. DFN-Forum Kommunikations-technologien
30. – 31. Mai 2017, Berlin

P-272 Alexander Rossmann, Alfred Zimmermann (eds.)
Digital Enterprise Computing (DEC 2017)
11.–12. Juli 2017, Böblingen

P-273 Christoph Igel, Carsten Ullrich, Martin Wessner (Hrsg.)
BILDUNGSRÄUME
DeLFI 2017
Die 15. e-Learning Fachtagung Informatik der Gesellschaft für Informatik e.V. (GI)
5. bis 8. September 2017, Chemnitz

P-274 Ira Diethelm (Hrsg.)
Informatische Bildung zum Verstehen und Gestalten der digitalen Welt
13.–15. September 2017, Oldenburg

P-275 Maximilian Eibl, Martin Gaedke (Hrsg.)
INFORMATIK 2017
25.–29. September 2017, Chemnitz

P276 Alexander Volland, Martin Engstler, Masud Fazal-Baqaie, Eckhart Hanser, Oliver Linssen, Martin Mikusz (Hrsg.)
Projektmanagement und Vorgehensmodelle 2017
Die Spannung zwischen dem Prozess und den Menschen im Projekt
Gemeinsame Tagung der Fachgruppen Projektmanagement und Vorgehensmodelle im Fachgebiet Wirtschaftsinformatik der Gesellschaft für Informatik e.V. in Kooperation mit der Fachgruppe IT-Projektmanagement der GPM e.V., Darmstadt 2017

P-277 Lothar Fritsch, Heiko Roßnagel, Detlef Hühnlein (Hrsg.)
Open Identity Summit 2017
5.–6. October 2017, Karlstad, Sweden

P-278 Arno Ruckelshausen, Andreas Meyer-Aurich, Karsten Borchard, Constanze Hofacker, Jens-Peter Loy, Rolf Schwerdtfeger, Hans-Hennig Sundermeier, Helga Floto, Brigitte Theuvsen (Hrsg.)
Informatik in der Land-, Forst- und Ernährungswirtschaft
Referate der 38. GIL-Jahrestagung
26.–27. Februar 2018, Kiel

P-279 Matthias Tichy, Eric Bodden, Marco Kuhrmann, Stefan Wagner, Jan-Philipp Steghöfer (Hrsg.)
Software Engineering und Software Management 2018
5.–9. März 2018, Ulm

P-280 Ina Schaefer, Dimitris Karagiannis, Andreas Vogelsang, Daniel Méndez, Christoph Seidl (Hrsg.)
Modellierung 2018
21.–23. Februar 2018, Braunschweig

P-281 Hanno Langweg, Michael Meier, Bernhard C. Witt, Delphine Reinhardt (Hrsg.)
Sicherheit 2018
Sicherheit, Schutz und Zuverlässigkeit
25.–27. April 2018, Konstanz

P-282 Arslan Brömme, Christoph Busch, Antitza Dantcheva, Christian Rathgeb, Andreas Uhl (Eds.)
BIOSIG 2018
Proceedings of the 17th International Conference of the Biometrics Special Interest Group
26.–28. September 2018
Darmstadt, Germany

P-283 Paul Müller, Bernhard Neumair, Helmut Reiser, Gabi Dreo Rodosek (Hrsg.)
11. DFN-Forum Kommunikationstechnologien
27.–28. Juni 2018, Günzburg

P-284 Detlef Krömker, Ulrik Schroeder (Hrsg.)
DeLFI 2018 – Die 16. E-Learning Fachtagung Informatik
10.–12. September 2018, Frankfurt a. M.

P-285 Christian Czarnecki, Carsten Brockmann, Eldar Sultanow, Agnes Koschmider, Annika Selzer (Hrsg.)
Workshops der INFORMATIK 2018 - Architekturen, Prozesse, Sicherheit und Nachhaltigkeit
26.–27. September 2018, Berlin

P-286 Martin Mikusz, Alexander Volland, Martin Engstler, Masud Fazal-Baqaie, Eckhart Hanser, Oliver Linssen (Hrsg.)
Projektmanagement und Vorgehensmodelle 2018
Der Einfluss der Digitalisierung auf Projektmanagementmethoden und Entwicklungsprozesse
Düsseldorf 2018

P-287 A. Meyer-Aurich, M. Gandorfer, N. Barta, A. Gronauer, J. Kantelhardt, H. Floto (Hrsg.)
Informatik in der Land-, Forst- und Ernährungswirtschaft
Fokus: Digitalisierung für landwirtschaftliche Betriebe in kleinstrukturierten Regionen – ein Widerspruch in sich?
Referate der 39. GIL-Jahrestagung
18.–19. Februar 2019, Wien

P-288 Arno Pasternak (Hrsg.)
Informatik für alle
18. GI-Fachtagung
Informatik und Schule
16.-18. September 2019 in Dortmund

P-289 Torsten Grust, Felix Naumann, Alexander Böhm, Wolfgang Lehner, Jens Teubner, Meike Klettke, Theo Härder, Erhard Rahm, Andreas Heuer, Holger Meyer (Hrsg.)
Datenbanksysteme für Business, Technologie und Web (BTW 2019)
4.–8. März 2019 in Rostock

P-290 Holger Meyer, Norbert Ritter, Andreas Thor, Daniela Nicklas, Andreas Heuer, Meike Klettke (Hrsg.)
Datenbanksysteme für Business, Technologie und Web (BTW 2019)
Workshopband
4.–8. März 2019 in Rostock

P-291 Michael Räckers, Sebastian Halsbenning, Detlef Rätz, David Richter, Erich Schweighofer (Hrsg.)
Digitalisierung von Staat und Verwaltung
Gemeinsame Fachtagung Verwaltungsinformatik (FTVI) und Fachtagung Rechtsinformatik (FTRI) 2019
6.–7. März 2019 in Münster

P-292 Steffen Becker, Ivan Bogicevic,Georg Herzwurm, Stefan Wagner (Hrsg.)
Software Engineering and Software Management 2019
18.–22. Februar 2019 in Stuttgart

P-293 Heiko Roßnagel, Sven Wagner, Detlef Hühnlein (Hrsg.)
Open Identity Summit 2019
28.–29. März 2019
Garmisch-Partenkirchen

P-294 Klaus David, Kurt Geihs, Martin Lange, Gerd Stumme (Hrsg.)
INFORMATIK 2019
50 Jahre Gesellschaft für Informatik – Informatik für Gesellschaft
23.–26. September 2019 in Kassel

P-295 Claude Draude, Martin Lange, Bernhard Sick (Hrsg.)
INFORMATIK 2019
50 Jahre Gesellschaft für Informatik – Informatik für Gesellschaft
Workshop-Beiträge
23.–26. September 2019 in Kassel

P-296 Arslan Brömme, Christoph Busch, Antitza Dantcheva, Christian Rathgeb, Andreas Uhl (Eds.)
BIOSIG 2019
Proceedings of the 18th International Conference of the Biometrics Special Interest Group
18.–20. September 2019
Darmstadt, Germany

P-297 Niels Pinkwart, Johannes Konert (Hrsg.)
DELFI 2019 –Die 17. Fachtagung Bildungstechnologien
16.–19. September 2019 in Berlin

P-298 Oliver Linssen, Martin Mikusz, Alexander Volland, Enes Yigitbas, Martin Engstler, Masud Fazal-Baqaie, Marco Kuhrmann (Hrsg.)
Projektmanagement und Vorgehensmodelle 2019 –Neue Vorgehensmodelle in Projekten – Führung, Kulturen und Infrastrukturen im Wandel
1Gemeinsame Tagung der Fachgruppen Projektmanagement (WI-PM), Vorgehensmodelle (WI-VM) und Software Produktmanagement (WI-ProdM) im Fachgebiet Wirtschaftsinformatik der Gesellschaft für Informatik e.V.
in Kooperation mit der Fachgruppe IT-Projektmanagement der GPM e.V.,
Lörrach 2019

P-299 M. Gandorfer, A. Meyer-Aurich, H. Bernhardt, F. X. Maidl, G. Fröhlich, H. Floto (Hrsg.)
Informatik in der Land-, Forst- und Ernährungswirtschaft
Fokus: Digitalisierung für Mensch, Umwelt und Tier
Referate der 40. GIL-Jahrestagung
17.–18. Februar 2020,
Campus Weihenstephan

P-300 Michael Felderer, Wilhelm Hasselbring, Rick Rabiser, Reiner Jung (Hrsg.)
Software Engineering 2020
24.–28. Februar 2020
Innsbruck, Austria

P-301 Delphine Reinhardt, Hanno Langweg, Bernhard C. Witt, Mathias Fischer (Hrsg.)
Sicherheit 2020
Sicherheit, Schutz und Zuverlässigkeit
17.–20. März 2020, Göttingen

P-302 Dominik Bork, Dimitris Karagiannis, Heinrich C. Mayr (Hrsg.)
Modellierung 2020
19.–21. Februar 2020, Wien

P-303 Peter Heisig, Ronald Orth, Jakob Michael Schönborn, Stefan Thalmann (Hrsg.)
Wissensmanagement in digitalen Arbeitswelten: Aktuelle Ansätze und Perspektiven
18.–20.03.2019, Potsdam

P-304 Heinrich C. Mayr, Stefanie Rinderle-Ma, Stefan Strecker (Hrsg.)
40 Years EMISA
Digital Ecosystems of the Future: Methodology, Techniques and Applications
May 15.–17. 2019
Tutzing am Starnberger See

P-305 Heiko Roßnagel, Christian H. Schunck, Sebastian Mödersheim, Detlef Hühnlein (Hrsg.)
Open Identity Summit 2020
26.–27. May 2020, Copenhagen

P-306 Arslan Brömme, Christoph Busch, Antitza Dantcheva, Kiran Raja, Christian Rathgeb, Andreas Uhl (Eds.)
BIOSIG 2020
Proceedings of the 19th International Conference of the Biometrics Special Interest Group
16.–18. September 2020
International Digital Conference

P-307 Ralf H. Reussner, Anne Koziolek, Robert Heinrich (Hrsg.)
INFORMATIK 2020
Back to the Future
28. September – 2. Oktober 2020, Karlsruhe

P-308 Raphael Zender, Dirk Ifenthaler, Thiemo Leonhardt, Clara Schumacher (Hrsg.)
DELFI 2020 –
Die 18. Fachtagung Bildungstechnologien der Gesellschaft für Informatik e.V.
14.–18. September 2020
Online

P-310 Anne Koziolek, Ina Schaefer, Christoph Seidl (Hrsg.)
Software Engineering 2021
22.–26. Februar 2021,
Braunschweig/Virtuell

The titles can be purchased at:

Köllen Druck + Verlag GmbH
Ernst-Robert-Curtius-Str. 14 · D-53117 Bonn
Fax: +49 (0)228/9898222
E-Mail: druckverlag@koellen.de

Gesellschaft für Informatik e.V. (GI)

publishes this series in order to make available to a broad public recent findings in informatics (i.e. computer science and information systems), to document conferences that are organized in cooperation with GI and to publish the annual GI Award dissertation.

Broken down into
• seminars
• proceedings
• dissertations
• thematics
current topics are dealt with from the vantage point of research and development, teaching and further training in theory and practice. The Editorial Committee uses an intensive review process in order to ensure high quality contributions.

The volumes are published in German or English.

Information: http://www.gi.de/service/publikationen/lni/

ISSN 1617-5468
ISBN 978-3-88579-701-2

The 50th annual conference of the Gesellschaft für Informatik (GI) was also the first virtual one, however, without shifting its thematic focus. The main track reflected the well established key subjects of informatics in Karlsruhe: data science, robotics and AI, Internet and society, secure and reliable systems, software engineering, as well as autonomous driving, mobility systems, energy informatics, and digital health. These topics also pointed out the role of informatics as a central problem solver for various challenges of modern societies. To enable a dialog beyond the purely technical point of view, the conference sessions and workshops contained a mix of scientific contributions as well as contributions by business representatives and public players.