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
• seminar
• 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-ev.de/service/publikationen/lni/

ISSN 1617-5468
ISBN 978-3-88579-270-3

The volumes P175, P176 and P177 contain the Proceedings of INFORMATIK 2010 – the 40th Annual Conference of the Gesellschaft für Informatik e.V. (GI) which took place in Leipzig from September 27 to October 1, 2010.

The conference featured invited keynote talks of leading experts from industry and academia and selected workshops from different fields of computer science, mostly covering the general conference topic "Service Science – New Perspectives for Computer Science" ("Service Science – Neue Perspektiven für die Informatik").

K.-P. Fähnrich, B. Franczyk (Hrsg.): INFORMATIK 2010

176

GI-Edition

Lecture Notes in Informatics

Klaus-Peter Fähnrich,
Bogdan Franczyk (Hrsg.)

INFORMATIK 2010

Service Science – Neue Perspektiven für die Informatik
Band 2

27.09. - 01.10.2010
Leipzig

Proceedings



Klaus-Peter Fähnrich, Bogdan Franczyk (Hrsg.)

INFORMATIK 2010

Service Science –
Neue Perspektiven für die Informatik
Band 2

27.09. - 01.10.2010
Leipzig

Gesellschaft für Informatik e.V. (GI)

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

Volume P-176 ISBN 978-3-88579-270-3
ISSN 1617-5468

Volume Editors
Prof. Dr. Klaus-Peter Fähnrich
Universität Leipzig, Institut für Informatik, 04103 Leipzig, Germany, Email: faehnrich@informatik.uni-leipzig.de
Prof. Dr. Bogdan Franczyk
Universität Leipzig, Institut für Wirtschaftsinformatik, 04109 Leipzig, Germany, Email: franczyk@wifa.uni-leipzig.de

Series Editorial Board
Heinrich C. Mayr, Universität Klagenfurt, Austria (Chairman, mayr@ifit.uni-klu.ac.at)
Hinrich Bonin, Leuphana-Universität Lüneburg, Germany
Dieter Fellner, Technische Universität Darmstadt, Germany
Ulrich Flegel, SAP Research, Germany
Ulrich Frank, Universität Duisburg-Essen, Germany
Johann-Christoph Freytag, Humboldt-Universität Berlin, Germany
Thomas Roth-Berghofer, DFKI
Michael Goedicke, Universität Duisburg-Essen
Ralf Hofestädt, Universität Bielefeld
Michael Koch, Universität der Bundeswehr, München, Germany
Axel Lehmann, Universität der Bundeswehr München, Germany
Ernst W. Mayr, Technische Universität München, Germany
Sigrid Schubert, Universität Siegen, Germany
Martin Warnke, Leuphana-Universität Lüneburg, Germany

Dissertations
Dorothea Wagner, Universität Karlsruhe, Germany
Seminars
Reinhard Wilhelm, Universität des Saarlandes, Germany
Thematics
Andreas Oberweis, Universität Karlsruhe (TH)

© Gesellschaft für Informatik, Bonn 2010
printed by Köllen Druck+Verlag GmbH, Bonn

Steering Committee

Leitung:

Klaus-Peter Fähnrich	Universität Leipzig
Bogdan Franczyk	Universität Leipzig
Klaus Bastian	Hochschule für Technik, Wirtschaft und Kultur Leipzig (HTWK)
Reinhard Clemens	T-Systems Deutsche Telekom AG
Ulrich Frank	Universität Duisburg-Essen
Helmut Krcmar	Technische Universität München
Pierfrancesco La Mura	Handelshochschule Leipzig (HHL)
Achim Luhn	Siemens AG
Horst Meier	Ruhr-Universität Bochum
Kathrin Möslein	Universität Erlangen-Nürnberg
Gerhard Satzger	IBM / Karlsruhe Service Research Institute (KSRI)
Gerik Scheuermann	Universität Leipzig
Anette Weisbecker	Fraunhofer IAO
Sabine Wieland	Hochschule für Telekommunikation Leipzig (HfTL)

Chair 100. Geburtstag Konrad Zuse:

Martin Bogdan	Universität Leipzig

Chair Telematik und Mobile Dienste:

Volker Gruhn	Universität Leipzig

Chair E-Humanities:

Gerhard Heyer	Universität Leipzig

Chair Informationsintegration in Service Architekturen:

Erhard Rahm	Universität Leipzig

Chair Architekturen für Gesundheitsdienstleistungen:

Alfred Winter	Universität Leipzig

Chair International Symposium on Services Science (ISSS):

Rainer Alt	Universität Leipzig

Chair Business Process and Services Computing (BPSC):

Witold Abramowicz	Poznan University of Economics
Leszek Maciaszek	Macquarie University Sydney / Wroclaw University of Economics

Organisationskomitee

Leitung:	
Alexandra Gerstner	Universität Leipzig
Jana Braun	Universität Leipzig
Ekaterina Demidova	Handelshochschule Leipzig (HHL)
Viola Gaida	Hochschule für Technik, Wirtschaft und Kultur Leipzig (HTWK)
Martin Gebauer	Universität Leipzig
Janika Groß	Universität Leipzig
Karsten Hain	Hochschule für Telekommunikation Leipzig (HfTL)
Kyrill Meyer	Universität Leipzig
Stefanie Plehn	Universität Leipzig
Kathleen Schumann	Universität Leipzig
Robin Sommerweiß	Universität Leipzig
Sven Windisch	Universität Leipzig
Studierendenprogramm:	
Christopher Klinkmüller	Universität Leipzig
André Ludwig	Universität Leipzig
Thomas Riechert	Universität Leipzig
Schüler- und Lehrerprogramm:	
Hans-Gert Gräbe	Universität Leipzig
Finanzen/Aussteller:	
Tanja Langmesser	INTERPLAN München
Franziska Simon	INTERPLAN München
GI-Geschäftsstelle:	
Andrea Schlaus	Gesellschaft für Informatik e.V. (GI)
Cornelia Winter	Gesellschaft für Informatik e.V. (GI)

Vorwort Tagungsleitung

Der Dienstleistungssektor ist der am schnellsten wachsenden Bereich der Wirtschaft. Insbesondere auf Informations- und Kommunikationstechnik basierende Dienste und Dienstleistungen stehen hier an der Spitze. Seit einigen Jahren bemüht man sich auch in der akademischen Welt, diesem Trend Rechnung zu tragen. Folgerichtig wurde eine interdisziplinäre Disziplin – die sogenannte Service Science – begründet. Im deutschsprachigen Raum nennt man dieses Gebiet auch Dienstleistungsforschung. Sie bringt Forscher und Entwickler aus den Ingenieurwissenschaften und der Informatik, den Wirtschaftswissenschaften aber auch den Arbeitswissenschaften und der Psychologie zusammen.

Die Informatik 2010 – die 40. Jahrestagung der Gesellschaft für Informatik e.V. (GI) findet vom 27. September bis zum 01. Oktober 2010 auf dem Innenstadt-Campus der Universität Leipzig statt. Sie steht unter dem Motto:

„Service Science – Neue Perspektiven für die Informatik“

Leipzig ist ein internationales Zentrum der sich rasch entwickelnden Disziplin. Es verfügt über ein eigenes Forschungscluster zur Dienstleistungsforschung in der Verbindung zwischen der Universität Leipzig, der Hochschule für Technik, Wirtschaft und Kultur Leipzig (HTWK), der Deutsche Telekom Hochschule für Telekommunikation (HfTL) und der Leipzig Graduate School of Management (HHL).

Wie aktuell die Thematik ist, zeigt auch die außerordentlich starken Beachtung des diesjährigen Mottos in den eingereichten Tagungsbeiträgen. Über die Hälfte der Beiträge widmet sich einem Aspekt der Dienstleistungsforschung. Insgesamt haben wir über 350 Beiträge erhalten. Dazu kommen noch zahlreiche weitere akzeptierte Arbeiten, welche im Rahmen der parallel stattfindenden Konferenzen präsentiert werden.

Der erste Tagungsband beinhaltet die Beiträge zum Tagungsmotto „Service Science“. Im zweiten Band finden sich die Beiträgen der sonstigen Workshops wieder sowie der Tutorien und der Studentenkonferenz SKIL. Der dritte Tagungsband widmet sich den beiden internationalen Konferenzen Business Process and Services Computing (BPSC) sowie dem International Symposium on Services Science (ISSS).

Die beiden Konferenzen Business Process and Services Computing (BPSC) und German Conference on Multi-Agent System Technologies (MATES) finden am Montag und Dienstag statt, das International Symposium on Services Science (ISSS) lädt am Donnerstag interessierte Teilnehmer ein. Tutorials zu diversen Themen bieten die ganze Woche einen praktischen Einblick, so beispielsweise zur modellgetriebenen Anwendungssoftwareentwicklung oder Managementaspekte beim Software-Engineering. Ebenfalls gibt es ein reichhaltiges Schüler- und Studentenprogramm u.a. mit dem LSB-Microsoft SummerCamp 2010 und der IBM Summer School.

Workshops zum Tagungsmotto aber auch zu vielen weiteren spannenden Themen der Informatik runden am Dienstag und Donnerstag das Programm ab. Der diesjährige „Tag der Informatik“ am Mittwoch steht ganz im Zeichen des Tagungsmottos. Namhafte Experten aus Wirtschaft und Forschung werden ihren Blick auf die Thematik darstellen.

Zu Beginn würdigt Arndt Bode den Informatiker Konrad Zuse. Thomas Curran von der Deutschen Telekom AG und Johannes Helbig von der Deutschen Post AG sprechen im Anschluss zu aktuellen Entwicklungen in der Dienstleistungswirtschaft. Am Nachmittag kommt Matthias Jarke vom Exzellenzcluster UMIC zu Wort, der zu mobilen Informations- und Kommunikationsdienstleistungen sprechen wird. Das Thema wird ebenfalls von Hermann Friedrich von der Siemens AG aufgegriffen. Er gibt einen Einblick in das Internet der Dienste. Abschließend spricht Kris Singh von IBM Research Almaden über das Forschungs- und Innovationsinstitut als globales Netzwerk der Dienstleistungswissenschaft.

Weiterhin gedenken wir des 100. Geburtstags von Konrad Zuse, dem Miterfinder des modernen Computers. Aus diesem Anlass hoffen wir auf zahlreiche Besucher der Konrad-Zuse Ausstellung und laden Sie zum Empfang an der Universität Leipzig mit anschließender Konrad-Zuse-Party ein.

Wir freuen uns, die 40. Jahrestagung mit einem Rahmenprogramm abrunden zu können. Es ermöglicht den Tagungsteilnehmern, Impressionen von der dynamischen Stadt Leipzig mit ihrer bewegten Geschichte zu gewinnen. Möglich wird dies durch den traditionellen Abendempfang, der vor Ort in den historischen Gemäuern des Auerbachs Kellers stattfindet, sowie durch Stadtführungen und abendliche Kneipentouren mit Leipziger Spezialitäten.

Die Tagung wird begleitet durch eine Firmenausstellung, die einen Einblick in aktuelle Marktentwicklungen und Möglichkeiten zum Kontaktknüpfen bietet.

Wir wünschen allen Vortragenden und Teilnehmern ein gutes Gelingen für die diesjährigen Workshops, Konferenzen und Tutorials, erfolgreiche Präsentationen und interessante Begegnungen.

Für die hervorragende Vorbereitung der Tagung geht unser Dank an die Organisatoren der einzelnen Workshops und die Mitgliedern der einzelnen Programmkomitees für die Erstellung der Gutachten. Insbesondere möchten wir den Mitgliedern des Organisationskomitees für die Planung und Durchführung der INFORMATIK 2010 danken. Ein besonderer Dank geht hierbei an Frau Alexandra Gerstner, die die Planungen zur Veranstaltung von Beginn an ausdauernd und erfolgreich begleitet und in die Tat umgesetzt hat. Für die Gestaltung des „Tags der Informatik“ sowie der anderen Konferenztage danken wir den eingeladenen Vortragenden, Autoren und Rednern. Ein weiteres Anliegen ist es, unseren Dank der Gesellschaft für Informatik e.V., insbesondere Frau Cornelia Winter sowie der Firma INTERPLAN für das Teilnehmer- und Finanzhandling auszudrücken. Für die Durchführung einer Tagung dieser Größenordnung bedarf es weiterer zahlreicher Helfer und Unterstützer, denen wir für ihre Hilfe bei der Vorbereitung und Realisierung der Tagung danken möchten. Ohne die finanzielle Unterstützung von Firmen und Institutionen wäre diese Veranstaltung nicht möglich gewesen – ihnen gilt ebenfalls unser herzlicher Dank.

Leipzig, im September 2010 Klaus-Peter Fähnrich, Bogdan Franczyk

Mit unseren Lösungen können Sie von überall arbeiten. Müssen Sie aber nicht.

Mit Lösungen der Telekom für vernetztes Arbeiten haben Sie die Chance, immer und überall auf Kundenanfragen zu reagieren – aus Ihrem Büro, von zu Hause oder von unterwegs. Wir schenken Ihnen damit ein Stück Freiheit und Sie Ihren Kunden das Gefühl, immer für sie da zu sein. www.telekom.de/gk-center

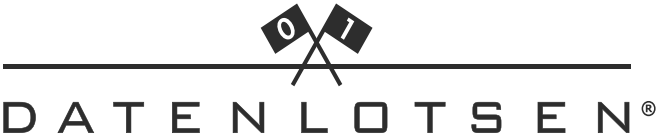


Erleben, was verbindet.




Capgemini
CONSULTING.TECHNOLOGY.OUTSOURCING


SAP


DATENLOTSEN®


INTER
FORUM


Stadt Leipzig
Amt für Wirtschaftsförderung


ICS


VIEWEG+
TEUBNER


Netzwerk
Geschäftsprozessmanagement

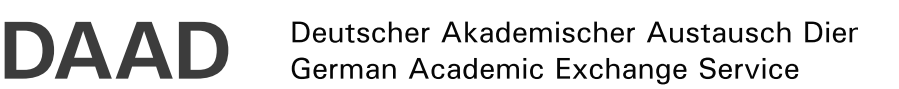

SENACOR


Unister
Holding


COMPETENCE CENTER FOR
APPLIED SECURITY TECHNOLOGY
CAST


MAGAZIN FÜR PROFESSIONELLE
INFORMATIONSTECHNIK


IT Jobs, Praktika & Ausbildung
ITmitte.de
Leipzig Halle Jena Erfurt Zwickau Chemnitz


DAAD
Deutscher Akademischer Austausch Dier
German Academic Exchange Service


LIV




Oldenbourg
Wissenschaftsverlag


PEARSON
Studium

Inhaltsverzeichnis

Band 2

Vorträge zum Tag der Informatik

Helbig Johannes, Gorriz Michael
SOA Innovation Lab and SOPERA – open innovation for a joint agenda of users and vendors 3

Jarke Matthias
Mobile Informations- und Kommunikationsdienste: Kooperativ, umweltschonend, sicher? 10

Workshops

Virtualisierung in der ICT-Branche 17

Joswig Jürgen
Virtualized DDI 19

Eren Evren, Breitländer Markus
Virtualisierung komplexer Netzwerke 26

Schmidt Benjamin
DatabaseHypervisor – ein Ansatz für relationale Datenbanken in der Cloud 32

Engelland Holger
Die Schwachstellen virtueller Systeme 38

Sorg Michael
Sichere Fernwartungszugriffe 43

Werner Mario
Intelligentes Backup/Recovery für virtuelle Umgebungen 49

Wieland Sabine
SOA in der Cloud 55

Krosta-Hartl Pamela, Burlaga Roland
Netzwerkvirtualisierung: Kosten senken im LAN & WAN 61

Applications of Semantic Technologies – AST2010 69

Abecker Andreas, Bloehdorn Stephan, Grimm Stephan
Application of Semantic Technologies. AST2010. 5th International Workshop (Vorwort) 71

Kleiner Frank, Abecker Andreas
Semantic MediaWiki as an Integration Platform for IT Service Management 73

Luczak-Rösch Markus, Coskun Gökhan, Paschke Adrian, Rothe Mario, Tolksdorf Robert
SVoNt - Version Control of OWL Ontologies on the Concept Level 79

Janning Ruth
Commonsense Ontologies and the Use of Words in Natural Language 85

Lucke Ulrike, Martens Alke
Utilization of Semantic Networks for Education: On the Enhancement of Existing Learning Objects with Topic Maps in <ML>³ 91

Kluge Rolf, Ludwig André, Belter Roman
Towards Semantic Matching of Business Services and Electronic Services 97

Michels Carolin, Agarwal Sudhir
Elicitation of Preferences for Web Service Compositions 103

Schulte Stefan, Siebenhaar Melanie, Eckert Julian, Steinmetz Ralf
Query Languages for Semantic Web Services 109
Langversion des Beitrags auf CD

Hess Claudia, Chen Willy, Syldatke Thomas
Business-oriented CAx Integration with Semantic Technologies Revisited 115

Konrad-Zuse-Workshop: Rechnender Raum und Zellulare Automaten 121

Klauer Bernd
Konrad-Zuse-Workshop. Rechnender Raum und Zellulare Automaten (Vorwort) 123

Zuse Horst
Der Rechnende Raum von Konrad Zuse 124

Schreckenberg Michael
Der Stau im Computer – Verkehrssimulationen mit Zellularautomaten 125

Fey Dietmar
Marching Pixels — Rechnen mit Hardware-Agenten im 2D Pixelraum 126

Hoßfeld Friedel
Konrad Zuses ”Rechnender Raum”— Kühne Betrachtungen nach 40 Jahren 127

Mahr Bernd
Zuses Auffassung vom Rechnen 130

2. Workshop zur IT-Unterstützung von Rettungskräften 133

Erfurth Christian, König-Ries Brigitta
Vorwort zum 2. Workshop IT-Unterstützung für Rettungskräfte (Vorwort) 135

Danowski, Kamen
Wissenschaftliche Optimierungsmethoden für die strategische Planung und die operative Einsatzführung in der Gefahrenabwehr – Implementierung und praktische Anwendung 137

Birkhäuser Benedikt, Schwenniger Jochen, Loke Ben, Niittylä Laura, Karampiperis Pythagoras, Marterer Robin
Entwicklung eines ereignisgetriebenen Systems für intelligentes Ressourcenmanagement – ‚Work-in-Progress‘-Report zum Stand des ersten Demonstrators 143

Hamp Quirin, Reindl Leonhard
Graph analysis of uncertain measures for position estimation 150
Langversion des Beitrags auf CD

Nestler Simon, Coskun Tayfur, Artinger Eva, Pichlmaier Peter, Klinker Gudrun
Indirect Tracking of Patients in Mass Casualty Incidents 156

Holl Stefan, Seyfried Armin, Boltes Maik
Hermes – Erforschung eines Evakuierungsassistenten für den Krisenfall bei Großveranstaltungen 162

Schneider Steffen, Pottebaum Jens, Becker Tobias, Koch Rainer
Räumungsszenarien bei Großveranstaltungen: Modellierung einer Datenbasis für Planung und Forschung 168

Kremers Horst
Risiko und Rettung: Die Rolle der Informatik in den Phasen des Desaster Managements 174

Brückner Markus, Ribe-Baumann Liz
Decentralized, Resource-Aware Information Management and Delay Tolerant Networks in Command-and-Control 175

Coskun Tayfur, Artinger Eva, Nestler Simon, Benzina Amal, Klinker Gudrun
Is it possible to interact with a handheld device while holding it in both hands? 181
Langversion des Beitrags auf CD

Schumann Sandy, Yildirim-Krannig Yeliz, Mähler M., Wucholt F., Strohschneider Stefan
Kultur und Technik - Ein kulturanalytischer Ansatz als Grundlage der Entwicklung eines mobilen Endgerätes für Einsatzkräfte in Großschadenslagen 187

Greiner-Mai Thomas, Donner Anton
Data Management in Mass Casualty Incidents: The e-Triage Project..... 192

5. Workshop Vorgehensmodelle in der Praxis – Lösung oder Ursache aller Probleme..... 199

Kuhrmann Marco, Keil Patrick, Schnackenburg André, Linssen Oliver
Vorgehensmodelle in der Praxis – Lösung oder Ursache aller Probleme? (Vorwort). 201

Wind Stefan, Schrödl Holger
Einführung agiler Projektmanagementmethoden für Softwareentwicklung in der IuK-Industrie – Fallstudie und Handlungsempfehlungen..... 203

Lewitz Olaf
Warum Projekte einen Business Case brauchen..... 209

Romahn Andreas
Einführung eines agilen Vorgehensmodells auf der Basis von Werkzeugen und eines Leitbildes..... 215

Bombosch Uwe, Bente Stefan, Hesselmann Margret, Seifert Gero
Applying scrum and XP in An Enterprise Context..... 221
Langversion des Beitrags auf CD

Risk Management, Compliance und Governance für widerstandsfähige Informationssysteme..... 227

Schermann Michael, Krcmar Helmut
Risk Management, Compliance und Governance für widerstandsfähige Informationssysteme (Vorwort)..... 229

Schwering Benjamin, Pellengahr Alexander
Die Relevanz von IT-Compliance Prüfungsstandards bei IT-Auslagerungen im deutschen Bankensektor 231

Wolf Patrick, Goeken Matthias
Integration von IT- und unternehmensweitem Risikomanagement auf Konzeptebene 239

Racz Nicolas, Weippl Edgar, Seufert Andreas
Questioning the need for separate IT risk management frameworks 245

Walser Konrad, Riedl Reinhard
Unternehmensarchitekturmanagement als Mittler zwischen IT-Governance und IT-Management 253

Pauli Manfred, Schermann Michael, Krcmar Helmut
The Risk-Aware Enterprise Architecture: Towards a Transparent Inventory of IT Risk Management Artifacts 259

Knittl Silvia, Hommel Wolfgang
Umgang mit Risiken für IT-Dienste im Hochschulumfeld am Beispiel des Münchner Wissenschaftsnetzes 265

Service Engineer, multidisziplinärer Allrounder oder Erfüllungsgehilfe auf Abruf Der Chief Social Media Officer (CSMO'10) 271

Wendland Karsten, Heinz Joachim, Wolf Frank
Service Engineer, multidisziplinärer Allrounder oder Erfüllungsgehilfe auf Abruf? Der „Chief Social Media Officer“ (Vorwort) 273

IT-Governance in verteilten Systemen (GVS) 275

Repp Nicolas, Schulte Stefan, Steffens Ulrike
Workshop IT-Governance in verteilten Systemen (GVS) 2010 (Vorwort) 277

Buckl Sabine, Matthes Florian, Roth Sascha, Schulz Christopher, Schweda Christian M.
A Method for Constructing Enterprise-wide Access Views on Business Objects 279

Looso Stefanie
Multi-Modell-Umgebung IT-Governance: Einsatz mehrerer Best-Practice-Referenzmodelle 285

Sommer Martin
Skills-Management - Schattendasein der SOA-Governance 291

Niemann Michael, Eckert Julian, Steinmetz Ralf
Semantische Analyse zur Unterstützung von SOA-Governance 299

Milicevic Danijel, Goeken Matthias
Konzepte der Informationssicherheit in Standards am Beispiel ISO 27001 305

Johannsen Wolfgang
Information-Governance: Herausforderungen in verteilten Umgebungen 311
Langversion des Beitrags auf CD

5. Modellbasiertes Testen und Testautomatisierung (MoTes 2010) . 317

Belli Fevzi, Budnik Christof J., Hollmann Axel, Saglietti Francesca, Winter Mario
MOTES2010 – Modellbasiertes Testen (Vorwort) 319

Weißleder Stephan, Lackner Hartmut
System Models vs. Test Models -Distinguishing the Undistinguishable? 321

Beyazıt Mutlu, Deistler Timon Sebastian, Gökçe Nida
Event-Based Mutation Testing vs. State-Based Mutation Testing – Comparison Using a Web-based System 327

Krause Jan, Diedrich Christian
Modellbasierte Testgenerierung aus Spezifikationen mit parallelem Verhalten 333

Duan Linshu, Hussmann Heinrich, Höfer Alexander
A Test-Oriented HMI Specification Model for Model-Based Testing of Automotive Human-Machine Interfaces 339

Cichos Harald, Schürr Andy
Dynamische Äquivalenzklassen im Klassifikationsbaum für zustandsbehaftete Systeme 345

Zitterell Dirk, Thiel Sebastian
Automatisierter funktionaler Steuergerätetest mit der EXtended Automation Method (EXAM) 351

Bauer Thomas, Eschbach Robert
Enabling statistical testing for component-based systems 357

Wischermann Dirk, Schröder-Preikschat Wolfgang
Modelle als Mittel zur Verstärkung von strukturellen Tests sicherheitskritischer Software 363

Schnelte Matthias, Güldali Baris
Test Case Generation for Visual Contracts Using AI Planning 369

Pinte Florian, Saglietti Francesca
Evaluierung des Fehlererkennungspotentials modellbasierter Komponenten- und Integrationstestfälle 375

Zukunftsfähigkeit unternehmensweiter Anwendungssysteme..... 381

Flieder Karl
Flexibilität für PLM-Services in der Digitalen Fabrik 383

Eggert Sandy
Gestaltung wandlungsfähiger ECM-Referenzprozesse unter Verwendung kartographischer Methoden 389
Langversion des Beitrags auf CD

Informationssysteme mit Open Source 395

Haubold Tobias, Golubski Wolfgang, Arnold Oliver, Beier Georg, Beine Gerrit
Software-Entwicklung mit Open Source Werkzeugen - Der GeneSEZ-Ansatz 397

Bolles André, Geesen Dennis, Grawunder Marco, Jacobi Jonas, Nicklas Daniela, Appelrath H.-Jürgen
Sensordatenverarbeitung mit dem Open Source Datenstrommanagementframework Odysseus 404

Müller Bernd
Java und Open-Source - Ein Geben und Nehmen 410

van Kempen Beate, Hogrebe Frank, Schranz Astrid
Toolunterstützte Transformation fachkonzeptioneller Verwaltungsprozesse mit der bflow Toolbox 418

Wirmer Lucas, Gehrke Nick, Nüttgens Markus, Peris Martina
Konzeption und Implementierung einer konfigurierbaren Komponente zum Modelldatenaustausch am Beispiel der bflow* Toolbox 424

Web Science 431

Auer Sören, Müller-Birn Claudia, Staab Steffen
Workshop Web Science (WSW 2010) –Social Computing Applications- (Vorwort) 433

Tolksdorf Robert
Towards Quality Measures for Web Network Extraction 435

Thurau Christian, Bauckhage Christian
Group Evolution Patterns in World of Warcraft 441

Hollink Laura, van Assem Mark
Estimating the relevance of search results in the Culture-Web: a study of semantic distance measures 447
Langversion des Beitrags auf CD

Orthuber Wolfgang, Dietze Stefan
Towards Standardized Vectorial Resource Descriptors on the Web 453

Neue Forschungen zu Konrad Zuse 459

Hellige, Hans Dieter
Neue Forschungen zu Konrad Zuse (Vorwort) 461

Seising Rudolf
Einführung: Die Zuse-Forschung und die Probleme der Pionier-Geschichten 463

Füßl Wilhelm
Der wissenschaftlich-technische Nachlass Konrad Zuses als Grundlage für neue Forschungsfragen und eine biografische Neubewertung 464

Hellige Hans Dieter
Die Kontrollnetze und 'Rechnenden Räume' des Konrad Zuse im „Dritten Reich" .. 466

Hashagen Ulf
Von Berlin ins Allgäu: Der Erfinderunternehmer Konrad Zuse 1945-1948 468

Zuse Horst
Unternehmenskultur, Entscheidungsstrukturen und Führungsprobleme in der Zuse KG 470

Petzold Hartmut
Konrad Zuse und N. Joachim Lehmann 471

Waldschmidt Klaus
Konrad Zuse und die parallele Rechnerarchitektur. Einige Gedanken und Beispiele 473

Vollmar Roland
Konrad Zuses informatisches Universum 474

8. Workshop Automotive Software Engineering 475

Schneider Jörn, Ziegenbein Dirk
8. Workshop Automotive Software Engineering (Vorwort) 477

Loos Andreas, Sum Richard
Geschäftsprozessautomatisierung im Nutzfahrzeuge-Support-Bereich 479

Ronneberger Torsten, Thum Daniela, Kratzsch Claudia
Konzepte zur Entwicklung, Integration und Vernetzung von Car-to-X Funktionen .. 485

Grossmann Jürgen, Neumann Carsten, Hinnerichs Andreas, Rechner Horst, Hecker Thomas, Radusch Ilja
Test von verteilten C2X-Applikationen 491
Langversion des Beitrags auf CD

Jaensch Martin, Hedenetz Bernd, Conrath Markus, Müller-Glaser Klaus D.
Transfer von Prozessen des Software-Produktlinien Engineering in die Elektrik/Elektronik-Architekturentwicklung von Fahrzeugen 497

Bunzel Stefan, Heidary Khosrau, Fürst Simon, Lajtkep Andre, Mössinger Jürgen, Cordes Jürgen, Schmerler Stefan, Kühn Christian, Kirschke-Biller Frank, Frielingsdorf Bernd, Rimkus Robert, Kacel Rick, Gilberg Alain, Delord Bertrand, Nishikawa Kenji, Hirano Hiroyuki, Titze Andreas, Kunkel Bernd
Hardware-independent Software Development with AUTOSAR 503

Scheible Jan, Kreuz Ingo
Ein Qualitätsmodell zur automatisierten Ermittlung der Modellqualität bei eingebetteten Systeme 509

Nörenberg Ralf , Reissing Ralf, Weber Jörg
ISO 26262 conformant Verification Plan 515

eHumanities – Welchen Nutzen hat die Informatik 521

Heyer Gerhard, Büchler Marco
GI Annual Meeting 2010. Workshop eHumanities – How does computer science benefit? (Vorwort) 523

Heyer Gerhard, Büchler Marco
Some Challenges Posed to Computer Science by the eHumanities 524

Wittenburg Peter, Hinrichs Erhard, Broeder Dan, Zastrow Thomas
eHumanities – können wir die Komplexität beherrschen? 530

Bamman David, Crane Gregory
Corpus Linguistics, Treebanks and the Reinvention of Philology 542

Becker Sven, Borowski Marion, Gnasa Melanie, Stalmann Kai, Wrobel Stefan
eHumanities: Intelligent Analysis and Information System for Humanities and Culture 552

Kahler Helge
eHumanities aus Sicht eines Förderers 557

Bornes Varol Marie-Christine , Ortola Marie-Sol, Gronoff Jean-Daniel
Specific polysemy of the brief sapiential units 559

Hafemann Ingelore
Der Thesaurus Linguae Aegyptiae – die Verknüpfung eines elektronischen Corpus ägyptischer Texte und dem Wortschatz der ägyptischen Sprache 566

Büchler Marco, Heyer Gerhard
Salton und Wittgenstein in den Humanities: Über die Semantik in Philosophischen Texten 572

Schlieder Christoph, Wullinger Peter
Semantic Ageing of Complex Documents: A Case Study from Built Heritage Preservation 580

Mehler Alexander, Diewald Nils, Gleim Rüdiger, Waltinger Ulli
Time Series of Linguistic Networks in the Patrologia Latina 586

Protokoll-basierte Modellierung von Geschäftsinteraktionen 595

Speck Andreas, Reich Johannes
Vorwort zum Workshop "Protokoll basierte Modellierung von Geschäftsinteraktionen" 597

Ramacher René, Mönch Lars
A Computational Study of an Automated Negotiation Scheme to Solve Multiple Criterion Single Machine Scheduling Problems 599

Unger Herwig, Akinalp Coskun
Berücksichtigung menschlicher Charaktereigenschaften in geschaftlichen Transaktionen 601

Reich Johannes
Finite system composition and interaction 603
Langversion des Beitrags auf CD

Speck Andreas, Rusnjak Andreas, Schulz Marcel
Realization of Business Processes 604
Langversion des Beitrags auf CD

Pokahr Alexander, Brauchbach Lars
Reusable Interaction Protocols for Workflows
Beitrag nur auf CD verfügbar

Runte Wolfgang
Formalizing Dependencies in Business Process Models Using Constraint Satisfaction 610

Wirtschaftlichkeit und Wertbeitrag von Informationssystemen 611

Strecker Stefan, Kütz Martin
Wirtschaftlichkeit und Wertbeitrag von Informationssystemen (WWI 2010) (Vorwort) 613

Gemlik Arkadius, Neumann Markus, Sprenger Jon, Breitner Michael H.
Praxisrelevanz des Modells von DeLone und McLean zur Erfolgsmessung von Informationssystemen 615

Meier Christian, Kundisch Dennis
Project interactions in value based IT project portfolio management 621

Martignoni Robert, Stimmer Jörg
Wirtschaftlichkeitsanalyse einer ausgelagerten Anwendungsentwicklung 627

Heise David
Potentiale der Unternehmensmodellierung für das IT-Kostenmanagement 633
Langversion des Beitrags auf CD

Koppe Roland, Häusler Stefan, Hahn Axel
Ein Modell zur Einflussanalyse von Änderungen in Entwicklungsprozessen 639

3. Workshop Informatik und Nachhaltigkeitsmanagement.... 645

Funk Burkhardt, Niemeyer Peter
Nachhaltigkeitsinformatik (Vorwort) 647

Niggemann Stephan
Nachhaltigkeit durch Energiecontrolling 649
Langversion des Beitrags auf CD

Bachmann Erik
REACH, RoHS, LCA… – Managing several complex material compliance requirements and evaluation of the environmental effects of a product during the whole lifecycle in SAP 660

Wohlgemuth Volker, Krehahn Peter, Ziep Tobias
Mobile Anwendungen als Datenquelle für das Stoffstrommanagement..... 673

Haak Liane, Giesen Nils, Gómez Jorge Marx
Bi-nationale Promotionsprogramme in der Betrieblichen Umweltinformatik am Beispiel des DEEBIS-Net Programms 681
Langversion des Beitrags auf CD

Deklarative Modellierung und effiziente Optimierung (MOC2010) 689

Geske Ulrich, Wolf Armin
Deklarative Modellierung und effiziente Optimierung (Vorwort)..... 691

Goltz Hans-Joachim
Interaktive Komponenten in constraint-basierten Planungssystemen 693
Langversion des Beitrags auf CD

Hofstedt Petra
The Multiparadigm Programming Language CCFL 699
Langversion des Beitrags auf CD

Troegner Dagi
Combination of fuzzy sets with the Object Constraint Language (OCL) 705

Wolf Armin, Geske Ulrich, Finsterbusch Andrej, Rothe Mario
Constraintbasierte Behandlungsplanung in der Dialyse..... 711
Langversion des Beitrags auf CD

2. Workshop über Daten in den Lebenswissenschaften. Informatik-Technologien und Support 717

Müller Wolfgang, Hofestädt Ralf, Kuhn Klaus A., Türker Can
Daten in den Lebenswissenschaften. Vom Papier über Datenbanken zur integrierten Informationsquelle (Vorwort) 719

Hippe Klaus, Kormeier Benjamin, Töpel Thoralf, Janowski Sebastian, Hofestädt Ralf
DAWIS-M.D. - a data warehouse system for metabolic data 720
Langversion des Beitrags auf CD

Mathäß Tobias, Haase Peter
Integrierte Informationsverwaltung für die Lebenswissenschaften mit der Information Workbench 726

Löbe Matthias, Stäubert Sebastian
Bringing Semantics to the i2b2 Framework 734

Lamla Gregor, Blaser Rainer, Prasser Fabian, Wurst Sebastian H. R., Rechl Hans, Gradinger Reiner, Kuhn Klaus A.
Eine Umsetzung des IHE Single Source Konzepts für die translationale Forschung bei Knochen- und Weichteilsarkomen 739
Langversion des Beitrags auf CD

Kohlmayer Florian M., Lautenschläger Ronald R., Wurst Sebastian H. R., Klopstock Thomas, Prokisch Holger, Meitinger Thomas, Eckert Claudia, Kuhn Klaus A.
Konzept für ein deutschlandweites Krankheitsnetz am Beispiel von mitoREGISTER 746

Wittig Ulrike, Kania Renate , Rojas Isabel, Müller Wolfgang
Herausforderungen bei der Extraktion von biochemischen Daten aus der Literatur .. 752

Mittag Kristin, Hinneburg Alexander
Ranked Set Search in Medline Documents 758

5th German Workshop SDF 2010 - Sensor Data Fusion: Trends, Solutions, Applications 765

Mertens Michael, Ulmke Martin
Multi-Target Tracking using Signal Strength Measurements 767

Painsky Amichai
First Order Multiple Hypothesis Testing for the Global Nearest Neighbor Data Correlation Approach 773
Langversion des Beitrags auf CD

Konle Wolfgang
Track heading and Speed Estimation 785
Langversion des Beitrags auf CD

Streit Roy L.
Connections Between Multitarget Intensity Filters and Medical Imaging 792

Guerriero Marco, Coraluppi Stefano, Carthel Craig, Willett Peter
Analysis of AIS Intermittency and Vessel Characterization using a Hidden Markov Model
Beitrag nur auf CD verfügbar

Battistello Giulia, Koch Wolfgang
Knowledge-aided Multi-Sensor Data Processing for Maritime Surveillance 796
Langversion des Beitrags auf CD

Engel Ulrich
Benefits of a Multi-GNSS Receiver in an Interference Environment 800
Langversion des Beitrags auf CD

Niewiejska Julia, Govaers Felix, Aschenbruck Nils
Position Tracking in Urban Environments using Linear Constraints and Bias Pseudo Measurements 806
Langversion des Beitrags auf CD

Brinkmann Kevin, Hurka Jörg
Narrowband Passive Sonar Tracking 812

Seget Kathrin, Schulz Arne, Heute Ulrich
Maneuver-Adaptive Multi-Hypothesis Tracking for Active Sonar Systems 818
Langversion des Beitrags auf CD

Ehlers Frank
A Multi-agent Concept for Multistatic Sonar 824

Lindgren David, Gustafsson Fredrik, Habberstad Hans
Microphone Networks for Shooter Localization 831

Kaune Regina
Performance analysis of passive emitter tracking using TDOA, AOA and FDOA measurements 838
Langversion des Beitrags auf CD

Hörst Julian, Oispuu Marc
Localization of Piecewise Curvilinearly Moving Targets Using Azimuth and Azimuth Rates 844
Langversion des Beitrags auf CD

Algeier Vadim, Thomä Reiner, Koch Wolfgang
Robust Likelihood Function for Blind Mobile Terminal Localization 850
Langversion des Beitrags auf CD

Wieneke Monika
Localization of Radioactive Source Carriers in Person Streams 856

Dästner Kaeye, Köhler Bastian, Opitz Felix
Information Fusion: Protection of Military Forces and Homeland Security 862
Langversion des Beitrags auf CD

Müller Reinhard, Demuth Michaela
Knowledge based Information Evaluation for Integrated Ship Bridge Systems 868
Langversion des Beitrags auf CD

Schikora Marek, Schikora Adam, Kogel Karl-Heinz, Koch Wolfgang, Cremers Daniel
Probabilistic Classification of Disease symptoms caused by Salmonella on Arabidopsis Plants 874

Dambreville Frederic
A tripartite architecture based on referee function for generic implementations of evidence combination rules 880

Pannetier, Benjamin, Dezert Jean
Multiple ground target tracking and classification with DSmT 886
Langversion des Beitrags auf CD

Schlangen Isabel, Daun Martina
Different tools for clutter mapping 892
Langversion des Beitrags auf CD

Rick Martin
Spatial Design of Transmitter-Receiver-Constellations for PCL-Systems using GIS based Visibility Analysis 898
Langversion des Beitrags auf CD

Baum Marcus, Feldmann Michael, Fränken Dietrich, Hanebeck Uwe D., Koch Wolfgang
Extended Object and Group Tracking: A Comparison of Random Matrices and Random Hypersurface Models 904
Langversion des Beitrags auf CD

Watson Alexander, Mihaylova Lyudmila, Vorley Dave
Bearings Only Versus Bearings and Extent Tracking for Missile Guidance Systems Utilising Particle Methods 907
Langversion des Beitrags auf CD

Svensson Daniel, Ulmke Martin, Danielsson Lars
Multi-Target Tracking with Partially Unresolved Measurements 913

Tutorien

Crestani Marcus, Sperber Michael
Erfolgreiche Anfängerausbildung im Programmieren 921

Weßel Christa
Semi-strukturierte Interviews im Software-Engineering: Indikationsstellung, Vorbereitung, Durchführung und Auswertung – Ein Fall-basiertes Tutorium 927

Weßel Christa
Software-Engineering Culture. Vom Umgehen miteinander und mit externen Partnern in IT-Projekten – Seminare in Aus- und Weiterbildung 938

Petrasch Roland
Model-to-Code-Transformation mit der Model to Text Transformation Language ... 950

Nolte Siegfried
M2m-Transformation mit der QVT Relations Language 957
Langversion des Beitrags auf CD

Nolte Siegfried
M2M-Transformation mit der QVT Operational Mappings 964
Langversion des Beitrags auf CD

SKIL 2010

Riechert Thomas, Auer Sören
Studentenkonferenz Informatik Leipzig (SKIL 2010) 973

Rausch Jan
Evolutionary Graph Clustering using Graph and Cluster Mixtures 975

Brekle Jonas
Abstracting Queries 979

Nitzschke Marcus, Hoffmann Clemens, Schöne Thomas, Baumheier Konrad, Mitjagin Marina
jQuery Plugin für eine erweiterte Autovervollständigung auf der Basis von SPARQL Endpunkten 983

Vierig Enrico
Requirements Engineering für Berichtsapplikationen im Produktionsprozess des BMW Werk Leipzig mittels Technologien des agilen Informationsmanagement - Anwendung der SoftWiki-Methodologie zur Realisierung von Produktionscockpits 985

Gerber Daniel
MM2RDF: A Framework for Extraction of Semantic Multimedia-Metadata 987

Rieß Christoph
Patterns for Semi-automatic Evolution and Refactoring of RDF Knowledge Bases . 989

Navarro Piris Rubén
Extracting Knowledge Bases from table-structured Web Resources applied to the semantic based Requirements Engineering Methodology SoftWiki 992

Golemo Florian
Unterstützung von Innovationsprozessen mit IT-Werkzeugen – Der InnoRadar als Teil eines Innovationswerkzeugs..... 996

Frommhold Marvin
Modellierung von touristischen Merkmalen am Anwendungsfall Vakantieland..... 1002

Arndt Natanael
Entwicklung eines mobilen Social Semantic Web Clients 1004

Vorträge zum Tag der Informatik

SOA Innovation Lab and SOPERA – open innovation for a joint agenda of users and vendors

Johannes Helbig, Michael Gorriz

Mitglied des Bereichsvorstands
Deutsche Post AG, 53250 Bonn

Chief Information Officer
Daimler AG, 70546 Stuttgart

J.Helbig@DeutschePost.de
michael.gorriz@Daimler.com
info@soa-lab.de

Abstract: Service Oriented Architectures (SOA) strengthen the vendor independence of companies, freeing them from the lock-in monolithic architectures tend to incur. Deutsche Post has externalized their SOA knowledge, skills, and infrastructure to the community, engaging in open innovation to respond to new challenges in the logistics industries. One example is SOPERA, an enterprise service bus that was open-sourced to the Eclipse foundation in order to lower the investment threshold for companies engaging in a SOA approach. A second example is the SOA Innovation Lab, Germany's first business driven EAM initiative. Here one of the current workstreams, led by Daimler as one of the currently sixteen prominent member companies, investigates into the implications of standard software packages on enterprise architecture, soliciting highly positive response from ISVs towards a joint strategic agenda between users and providers.

1 Business imperatives for innovation

The logistic industry is experiencing new challenges and chances. On the one side there is increasing pressure through deregulation, European and global competition and substitution (e.g. industry discontinuity for letter business). On the other side there is Growth potential through globalization, E-business growth and the unbundling of value chains due to specialization and industrialization.

For Deutsche Post this has led to three strategic IT-imperatives and chances:

- Flexibilization under uncertainty: the application landscape needs to be transformed into a robust basis that is adaptive to strategic business change

- From barrier to enabler: the IT function needs to be the strategic responsibility of each business manager and ensure world class delivery and source of differentiation
- Externalization: the foundation for internal and external collaboration needs to be build to enable the IT to use new business models

2 IT Strategy for consistent delivery

The critical success factors that enable the IT to consistently deliver against the business challenges are high business responsibility, focus on executional excellence, and an emphasis on enterprise architecture as enabling platform.

The basis for IT delivery is the underlying enterprise architecture. The goal should be to develop a "Lego" approach of modular services that leads to flexible application landscapes (exhibit 1). As this approach links internal and external domains in the same fashion, the boundaries between enterprise and those of its partners begin to blur. This is a crucial success factor in the automotive industry, especially for large enterprises like Daimler with high growth and innovation rates.

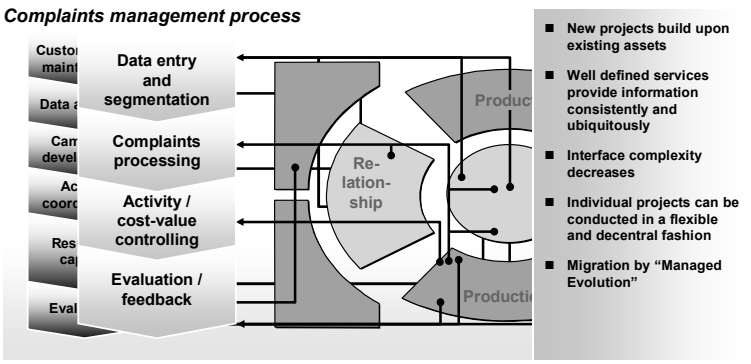


EXHIBIT 1 – "Lego" approach of modular services

Companies seeking to adopt a Service Oriented Architecture (SOA) need to address three strategic dimensions

- Governance & organization: SOA domains as master plan for orientation in a federal governance environment
- Skills & capabilities: Elaborate set of architecture management and service design processes as part of CIO function
- Applications & architecture: SOA framework: Deployable integration infrastructure to ease implementation by projects

Deutsche Post has gone through a substantial investment into SOA-enabling its application architecture through the last ten years. Deutsche Post has decided to share and externalize the knowledge, skills and infrastructure to further develop them with initiatives like the SOA Innovation Lab and open sourcing of SOPERA (exhibit 2).

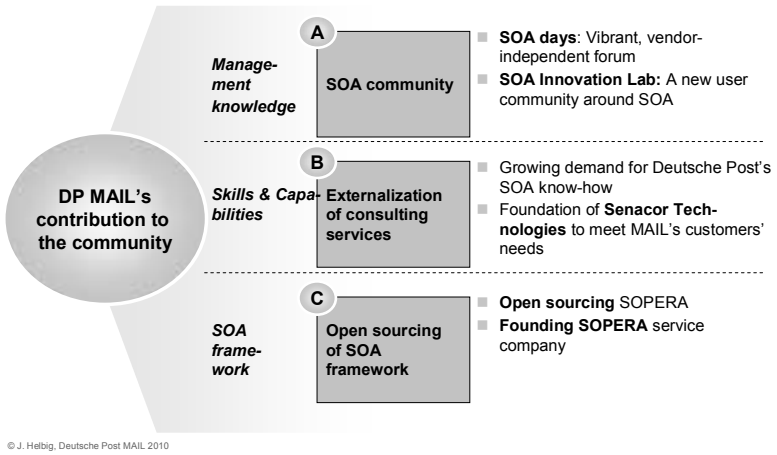


EXHIBIT 2 – Externalization of knowledge, skills, and infrastructure at Deutsche Post

3 Open innovation: SOPERA

SOPERA is the enterprise service bus that Deutsche Post developed to modularize its enterprise architecture. In 2007 Deutsche Post contributed SOPERA to Eclipse. Three years later, SOPERA comprises a complete integration suite, expanding the pure ESB functionality with BPM, security, governance and service management.

Due to a different development and sales process, SOPERA has significant lower cost. Deutsche Post reduces its costs of integration infrastructure by 75% within 4 year (exhibit 3). The eco system of customers and partners has reached a critical mass.

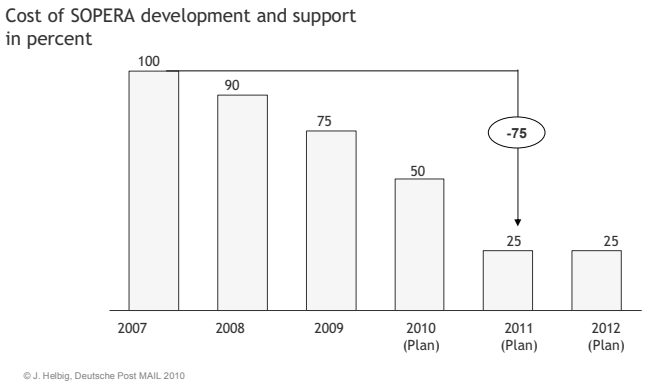


EXHIBIT 3 – Deutsche Post has reduced its costs of integration infrastructure by 75% within 4 years

SOPERA offers vendor Independence by enabling loose coupling of business applications as a SOA principle. Through its modular architecture SOPERA avoids vendor lock-in (exhibit 4). The Eclipse open source approach makes users even independent from the SOPERA service provider.

SOPERA GmbH offers all enterprise relevant services for the SOA platform: Support & Maintenance, Technical Consulting, Development sponsored by a customer which becomes part of the standard distribution, comprehensive training and certification programs

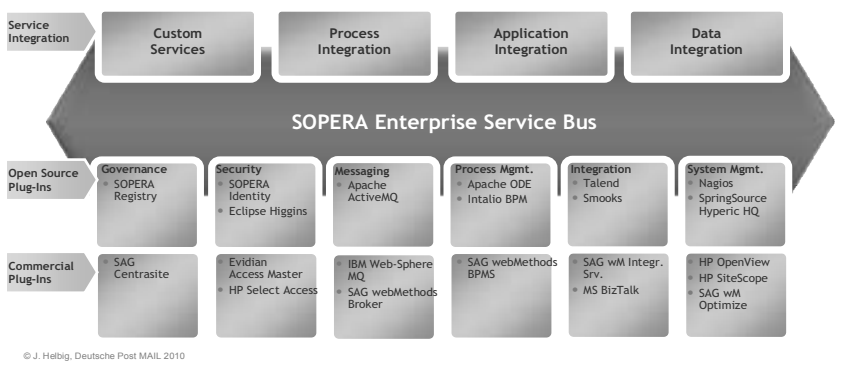


EXHIBIT 4 – Through its modular architecture, the open source ESB SOPERA avoids vendor lock-in

4 Open innovation: SOA Innovation Lab

SOA Innovation Lab is the first EAM business initiative in Germany (exhibit 5). Inflexible and complex IT landscapes represent a major challenge of modern large enterprises. SOA and EAM have grown into robust approaches, which, however, have to be enhanced and operationalized. User companies are at the top of SOA development. The agenda, however, is still driven by providers – therefore the idea of a user SOA community was born.

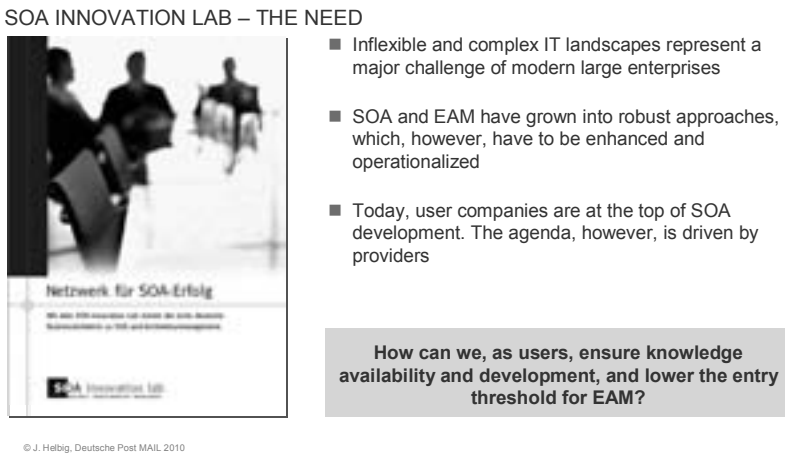


EXHIBIT 5 – SOA Innovation Lab is the first EAM business initiative in Germany

The mission of the SOA innovation lab is to build expertise and explore new approaches within large corporate IT users to develop flexible and efficient IT application landscapes by means of Service-oriented Architectures (SOA) and other methods of Enterprise Architecture Management (EAM). In the SOA Innovation Lab, employees are educated and trained in SOA methods, best practices are shared between members and new procedures and approaches are developed and tested.

The SOA Innovation Lab is a growing community with currently over 16 members representing large companies in all major industries (exhibit 6). Every company can define the degree of its individual commitment; the only prerequisite for membership are active contribution and a vital role of EAM for business success.

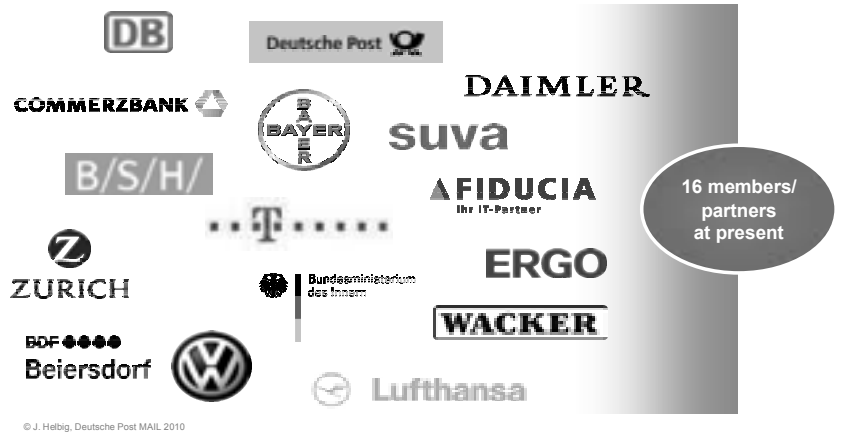


EXHIBIT 6 – Growing number of members

Current workstreams of the SOA Innovation Lab are addressing all relevant management questions with regard to Enterprise Architecture Management, the management of target architectures and SOA methods and tools (exhibit 7).

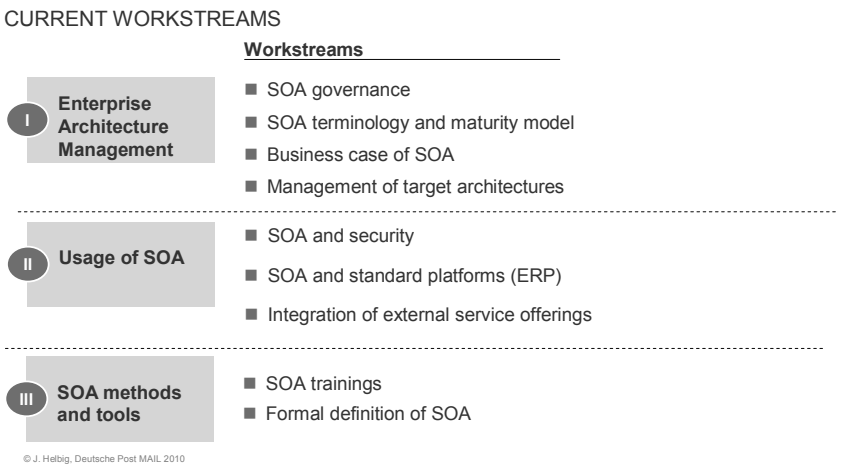


EXHIBIT 7 – Workstreams focus on practical relevance

As an example, the SOA and standard platform (ERP) workstream lead by Daimler investigates the use of standard software packages in a service-oriented context. As a part of this investigation the SOA capabilities of products from different vendors need to be evaluated (exhibit 8). For this purpose a SOA architecture maturity framework has been developed, leveraging and extending CMMI and TOGAF, as well as other state-of-the art frameworks and methods. Besides details about our framework, we present and analyze first results from the assessment of various vendor platforms.

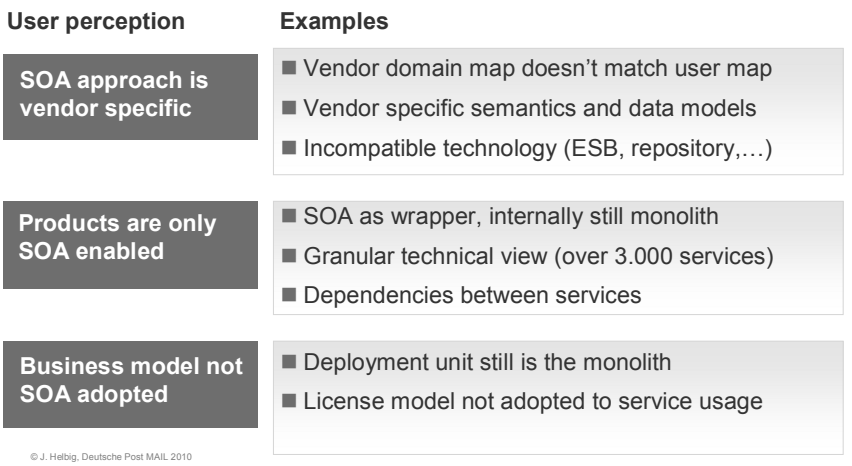


EXHIBIT 8 – SOA and Standard Software: current obstacles

Results from this workstream are that across vendors, there are still obstacles to apply standard software in a heterogeneous SOA environment. Often, a vendor's SOA approach is specific to the vendor. For most vendors, products are only SOA *enabled*. This means that SOA is implemented as wrapper around existing interfaces, and the internal structure is still monolithic. Finally, most vendors have not adopted a business model that supports the usage of standard software through services.

In summary, many vendors have invested early in SOA, long before users were ready to use new SOA enabled components in an appropriate way. The investment therefore was mostly to SOA enable products from a technical point of view, without considering the business scenarios that they should support. Therefore the adoption on the user side is slow, with only a few (100-300 per vendor) pilot SOA cases, mostly focusing on GUI integration. This is a tiny fraction of the overall installed base for standard software.

The most important result from the assessment workshops with vendors is that there is a strong interest from vendors to work together with the SOA Innovation Lab to further refine SOA methods and to develop solutions for the SOA use cases. We think that this is a great asset and we will continue to build on these relationships to further develop the maturity of SOA and standard software.

Experiences gained in the first two years of the SOA Innovation Lab are that sharing stories of success and failure, joining forces and learning from real experience are the key success factors for a growing SOA community. Given the success from the SOA Innovation Lab, the next step is now to grow internationally in order to expand the experience base and match the global footprint of most providers.

Mobile Informations- und Kommunikationsdienste: Kooperativ, umweltschonend, sicher?

Matthias Jarke

Lehrstuhl Informatik 5 (Informationssysteme)
RWTH Aachen & Fraunhofer FIT
Ahornstr. 55
52074 Aachen
jarke@cs.rwth-aachen.de

Abstract: Das explosive Wachstum der Mobilkommunikation trägt die Informatik in alle Welt. Es verändert damit nicht nur die Kultur bei uns, sondern vor allem auch in Schwellenländern und in der Dritten Welt. Gleichzeitig stellt es aber auch die Forschung vor neue Herausforderungen. Die Beherrschung von Heterogenität und Informationsüberflutung, Benutzerfreundlichkeit, Sicherheit und Datenschutz bei der Kommunikation, und nicht zuletzt der hohe Energie- und natürliche Ressourcenverbrauch erfordern zum Teil völlig neuartige informatische Konzepte im interdisziplinären Kontext mit Informationstechnik und Anwendungen. Im Vortrag werden diese Herausforderungen sowie einige Lösungen vorgestellt, die im Umfeld des Aachener Exzellenzclusters UMIC (Ultra-Highspeed Mobile Information and Communication) erarbeitet wurden.

1 Einleitung

Die Zahl der Mobiltelefone hat schon vor mehreren Jahren die der PCs auf der Welt überholt. Bei rund vier Milliarden Endgeräten kommt mittlerweile rein rechnerisch ein Handy auf zwei Erdbewohner. Da die Geräte immer mehr Rechenleistung anbieten, sind Mobiltelefone auch der Weg, auf dem das Informationszeitalter in große Teile der Schwellen- und Entwicklungsländer findet. Die Mobilkommunikation ist jedoch nicht auf das Handy beschränkt, sondern wird zunehmend auch im Internet der Dinge und Dienste zur Maschine-Maschine-Kommunikation und Informationsanalyse genutzt.

Das explosive Wachstum, das aus dieser Konfluenz zwischen Kommunikationstechnik, Informatik und zahlreichen Anwendungswissenschaften von Medien bis Soziologie entstanden ist, führt jedoch auch zu neuen Herausforderungen an die Informatik. Die strenge Separierung von Protokollschichten, wie sie in der Kommunikationstechnik zur Komplexitätsbewältigung traditionell verfolgt wird, muss ebenso durchbrochen werden wie die Grenzen zwischen den Disziplinen, um diese Herausforderungen zu bewältigen. Das Konzept des DFG-Exzellenzclusters bietet den Forschern in UMIC hierzu einen optimalen Rahmen. Drei Herausforderungen seien im Folgenden kurz skizziert.

2 Herausforderung Coverage

Zwar ist von GSM zu UMTS und demnächst zu LTE die Datenrate stark gestiegen und wird zunehmend multimedia-tauglich. Immer noch aber liegen Performanz und Verfügbarkeit mobiler Kommunikation weit hinter derjenigen des Festnetzes zurück. Auch der Schritt von GSM zu UMTS bedeutete nur dort einen wirklich großen Fortschritt, wo nur wenige Benutzer in einer infrastrukturmäßig gut erschlossenen Region parallel arbeiten – fast ein Widerspruch in sich. Die Erschließung ländlicher Regionen durch schnelle Mobilkommunikation scheiterte lange an zu kleinen Zellengrößen, die sich aus ungünstigen verfügbaren Frequenzspektren und damit kurzen Reichweiten der Sendemasten ergaben. Die „digitale Dividende“ aus der Digitalisierung von Rundfunk und Fernsehen soll hier Abhilfe schaffen; die dazu gehörigen Frequenzversteigerungen haben im Frühjahr 2010 stattgefunden. Ergänzend werden in UMIC adaptive Methoden des Cognitive Radio erforscht, um Handys in die Lage zu versetzen, durch Kooperation eine bessere Nutzung der vorhandenen Frequenzen zu erreichen.

Ein kurzfristigerer Ansatz zur regionalen Verbesserung der Situation in Gebieten, wo sich für die großen Anbieter bisher eine mobile Breitbandversorgung nicht lohnt, ist die Selbsthilfe: Durch eine homogenisierende Overlayschicht werden die vielen mittlerweile im lokalen Bereich vorhandenen Basisstationen (etwa WLAN) zusammengebracht, um etwa eine Gemeinde mit mobilen Daten zu versorgen. In dem UMIC angeschlossenen Anwendungsprojekt MobileACcess haben sich zahlreiche Aachener öffentliche und private Einrichtungen sowie Nachbargemeinden in der Eifel zusammengeschlossen, um eine solche Lösung auf Basis von UMIC-Technologien zu realisieren, wobei nicht nur die Infrastruktur, sondern auch ein reichhaltiges Repertoire mobiler Internetdienste aus Handel, Tourismus und eGovernment bereitgestellt werden soll.

3 Herausforderung Energie und Umwelt

Sowohl die Infrastruktur von Basisstationen als auch die mobilen Endgeräte machen mittlerweile einen merklichen Teil des gesamten Stromverbrauchs der Industrieländer aus, selbst wenn man von den riesigen Serverfarmen der großen stationären Internetanbieter wie Google absieht, die zunehmend von mobiler Kundschaft leben. Unsere Untersuchungen zeigten, dass die Ursachen steigenden Stromverbrauchs außerordentlich vielfältig und auf den verschiedensten Ebenen der Systemarchitekturen angesiedelt sind. Dementsprechend vielfältig und disziplinübergreifend sind auch die neuen informatischen Lösungsansätze, die hier benötigt werden.

Ein typischer Stromfresser ist bekanntlich schneller Multimedia-Einsatz. Lösungen für dieses Problem kombinieren neue Bildverarbeitungs- und Graphikalgorithmen mit Load Balancing der aufwändigen Berechnungen in der „Cloud“ jeweils lokal verfügbarer Server oder anderer Mobilgeräte und mit erweiterten Kompressionstechniken bis hin zu spezialisierten Prozessoren. Auch das Scheduling der Datenübertragung spielt eine Rolle, beispielsweise das Vermeiden von nutzlosen Übertragungsversuchen bei schlechter Übertragungsqualität oder das Pre-Loading von Daten. Ähnliche Lösungen werden auch für die mobile Analyse von Datenströmen im Gesundheitswesen oder in der Car-to-Car-Kommunikation untersucht.

Ein zweites, weniger offensichtliches Beispiel ist die Sicherheitstechnik. Gute Verschlüsselung war traditionell algorithmisch zu aufwändig für die Einheitsprozessoren eines Handy’s; heterogener Spezialprozessoren führen hier gemeinsam mit Broadcast-Technologien und angepassten Security-Protokollen zu wesentlich besseren Lösungen.

Als drittes Beispiel sei das scheinbar simple Problem der Mehrbandhandy’s genannt, mittels derer ein Endgerät weltweit für die verschiedenen Frequenzspektren genutzt werden kann. Durch die Kombination digitaler und analoger Aspekte entstehen hier potenziell Verfahren, die hardware- und abwärmeseitig exponentiell mit der Zahl nutzbarer Frequenzbänder wachsen. Verschiedene Hardware- und Softwareinnovationen mussten kombiniert werden, um dieses Problem zu lösen.

Neben der mit dem Stromverbrauch verbundenen CO2-Belastung stoßen auch die Mobilfunkmasten selbst wegen befürchteter Gesundheitsfolgen auf Akzeptanzprobleme. In UMIC werden daher multikriterielle parallele Optimierungsverfahren entwickelt, mittels derer die Standortwahl nicht nur unter Kosten- und Versorgungsgesichtspunkten, sondern auch unter dem Gesichtspunkt sozialverträglicher Abstände etwa zu Wohngebieten oder Kindergärten ganzheitlich optimiert wird.

4 Herausforderung Multimedia-Kooperation

Anwendungsseitig konzentriert sich die Aufmerksamkeit der Öffentlichkeit gern auf die großen Anbieter von Suchmaschinen, eCommerce und „Social Software“ wie Google, eBay oder Facebook. Dabei wird übersehen, dass der Großteil der Internetaktivitäten, auch der mobilen Internet-Aktivitäten, sich in den kleinen bis mittelgroßen Communities des sogenannten Long Tails abspielt, da die Anwendergemeinde sich wie fast alle großen sozialen Gruppierungen entsprechend des Power Law verhält. Diese Beobachtung hat erhebliche Konsequenzen für viele informatische, sozial-wissenschaftliche, volks- und betriebswirtschaftliche Theorien, die empirisch durch Studien solcher Communities hinterlegt werden können.

Die Bedeutung der Communities und die zunehmende Mobilität ihrer Mitglieder fordert jedoch auch dazu heraus, ihnen neue Technologien zur Verfügung zu stellen, mit denen sie zum einen operativ unterstützt werden, zum anderen aber auch ihre eigenen Spielregeln festsetzen, überwachen und anpassen können. Die in UMIC verfolgten Forschungsansätze kombinieren medientheoretische Modellierungskonzepte mit leichtgewichtigen semantischen Beschreibungen der in solchen Communities verwendeten medienübergreifenden Interaktionen. Forschungsergebnisse aus diesem Bereich decken ein breites Spektrum ab, das von neuartigen Endbenutzerschnittstellen über mobile Augmented Reality bis hin zu mobilen Webservern und Verfahren des Data Minings sozialer Netzwerke reicht.

Die Ergebnisse werden anhand dreier, teils bereits in der Praxis erprobter Szenarien demonstriert: dem „Virtual Campfire“ zum Multimedia-Storytelling aus mobil gesammelten Daten (exemplifiziert u.a. an einem Wiederaufbauprojekt im afghanischen Bamyantal); einem mobilen multimedialen und multimodalen Stadtführer für Aachen auf Basis neuer Interaktionsgeräte und Augmented Reality; und das HealthNet-Szenario zur kooperativen mobilen Gesundheitsüberwachung. Alle drei Szenarien werden in intensiver Zusammenarbeit mit Endanwendern entwickelt und validiert.

4 Fazit

Die Mobilkommunikation war technisch wie gesellschaftlich eine zentrale Herausforderung an das vergangene erste Jahrzehnt des 21. Jahrhunderts vor allem aufgrund der Konvergenz zwischen Funktechnik, Informationstechnik und Informatik. Sie wird auch in den kommenden Jahren unsere Kultur weiter verändern, wobei als zusätzliches Element vernetzte eingebettete Softwaresysteme in die Konvergenz einzubeziehen sind. Die mit der Maschine-Maschine-Kommunikation verbundene höhere erforderliche Präzision stößt dann mit der Freiheit des Sozialen Netzwerks zusammen, und wir dürfen gespannt sein, welche neuen Probleme und Lösungen sich daraus ergeben.

Weitere Informationen über den Exzellenzcluster UMIC, der seit 2006 im Rahmen der Exzellenzinitiative von Bund und Ländern gefördert wird, finden sich einschließlich eines umfassenden Publikationsverzeichnisses unter www.umic.rwth-aachen.de.

Workshops

Workshop
Virtualisierung in der ICT-Branche

Profn. Sabine Wieland

VIRTUALIZED DDI

Dipl.-Math. Jürgen Joswig

EMEA Sales Director
Nixu Software
Immenkamp 3
30926 Seelze OT Velber
Jurgen.Joswig@nixu.com
emea-sales@nixusoftware.com

Abstract: When talking about virtualized computing environments, virtualization of the IP and core network services such as DNS (Domain Name Service), DHCP (Dynamic Host Configuration Protocol) and IP Address Management are rarely thought of. This document provides a brief introduction to the Why & How of an end-to-end virtualized DDI solution with related technical, administrative and commercial benefits. It further discusses the advantages of adhering to emerging networking standards, including significantly increased level of security.

1 Introduction to Server Virtualization

The hardware of today's x86 computer was designed to run a single operating system and a single application, leaving resources of these machines most time in idle mode. Using virtualization multiple virtual machines are available on one single physical computer and all virtual machines share the real resources of this one physical computer. High flexibility is given as different virtual machines can run different operating systems and multiple applications on the same physical computer.

Today's virtualization platforms provide a business-ready architecture and "virtualize" hardware resources of x86-based machines - including CPU, RAM, HDD, Controller and Interfaces - to create fully functional virtual machines that can run their own applications on different operating systems and so behave like 'real' computers with each of these virtual machines providing a complete system.

A thin layer of software dynamically allocates hardware resources and multiple operating systems run concurrently on a single physical computer, sharing hardware resources with each other. By encapsulating an entire machine, including CPU, memory, operating system and network devices, a virtual machine is completely compatible with all standard x86 operating systems, applications and device drivers. Thanks to this innovative approach, one can build an entire virtual infrastructure, scaling across hundreds of interconnected physical computers and storage devices utilizing proven virtualization systems. This results in a platform that can be used to build clouds, without the need to assign servers, storage, network bandwidth and priority permanently to each

application. The virtualization 'Hypervisor' dynamically allocates hardware resources when and where they are needed within the cloud.

Because of this flexibility, server virtualization simplifies platform management by separating the hardware and thin software layer allocating system resources, thereby translating to significant savings in terms of computing platform TCO and OPEX. Furthermore, by separating the hardware platform from the applications and services run on the virtualized computing platform, virtualization environments allow organizations to increase the level of availability of network services and applications and cut costs.

2 Introduction to Emerging Network Standards: IPv6, DNSSEC

2.1 IPv6

According to recent reports by NRO (Number Resource Organization), the allocation of available IPv4 blocks has seen a significant increase during Q1 2010. In early Q2 2010, only 7.8% of IPv4 space remained free for allocation, and the available IPv4 space is now expected to exhaust by as soon as 2011. Although IPv4 is expected to co-exist with IPv6 well into the 2010s and early 2020s, a gradual shift to IPv6-based networking seems inevitable, especially as organizations such as OECD are now promoting global adaptation of the second generation IP protocol.

In comparison to IPv4, IPv6 will add a significant amount of complexity into managing DDI services. The added complexity stems from the fact that while IPv4 addresses were simple enough for network administrators and other users to remember and enter manually, the syntax of IPv6 addresses and the vast size of IPv6 spaces introduces new complexities into network management routines and processes that cannot be coped with manually. Due to this, the introduction of IPv6 and dual-stack networks is likely to bring about the use of next generation DDI management solutions and tools.

2.2 DNSSEC

After the Kaminsky vulnerability was made public in the summer of 2008, it became widely accepted that the inherent design of DNS is not secure, making this core network service vulnerable to man-in-the-middle, DNS cache poisoning and other similar forms of malicious attacks. While the related risks were mitigated for the time being by randomizing the source ports used in name resolution, the mitigation was in fact bypassed by Russian physicist Polyakov shortly after patches to widely used DNS server types were released. While Polyakov's approach to bypass the randomization of source ports was based on brute force possible only in laboratory environment, it is likely that the introduced method can be also used against production DNS servers in the future as more powerful CPUs and higher amount of bandwidth become available.

To address this issue, the global networking community is now making a move torwards the worldwide deployment of DNSSEC (DNS Security Extensions), a standard allowing DNS zone data to be signed in order to assure the authenticity of the name resolution process. The implementation of DNSSEC relies on public and private keys used to sign zone data – much in the same way as in PKI (Public Key Infrastructure) – allowing a chain of trust to be established all the way from the root DNS servers to the client resolving domain names. The DNS root will be signed on July 2010, preceded and followed by a number of ccTLDs (country code Top Level Domain) and gTLDs (generic Top Level Domains) such as .se, .bg and .org.

At the time of the writing, it seems quite likely that most of the 296 TLDs will be signed by the end of 2012.

From the network and DNS management perspective, DNSSEC presents three challenges to system and network administrators. First, when zones are signed, the number of resource records increases by the factor of 4-5, increasing the number of resource records in one's DNS. Second, formerly static DNS entries will become dynamic when signed, because the keys used to sign zones have to be rolled-over on regular intervals. Third, as signed zones contain resource records that are effectively very long keys, adding the keys manually becomes increasingly error-prone and tedious. Given these complexities, successful deployment of DNSSEC requires that the DNS management processes are automated. However, once automated, managing DNSSEC enabled environments is not more difficult than managing traditional DNS environments. In fact, the introduction of DNSSEC support to one's DNS management process is actually likely to reduce the related network management overheads, because the automation is not restricted to DNSSEC only but covers the entire process.

3 Virtualized DNS and IP Addressing Environments

Much of the discussion around virtualization has insofar focused on APPLICATIONS and STORAGE; virtualized DEVELOPMENT PLATFORM; and the NETWORK itself through virtual switches. Core network services such as DNS, DHCP and IP Address Management rarely come up in this context. However, considering the mission critical nature of core network services as well as the scalability and reliability requirements associated with them – and the fact that dedicated servers used to run network services utilize a rather low percentage of their CPU most of the time – running core network services in virtualized environments that optimize the use of computing resources available to them is often the most cost-efficient way of tackling these requirements.

As organizations initiate server virtualization projects in their operating environments, core network services such as **DNS** and **DHCP** make a perfect candidate for virtualization. Because of the high costs and inflexibilities associated with traditional, hardware based DNS and IP addressing computing appliances, most organizations have implemented existing D-services (DNS, DHCP) by utilizing unsupported open source software (BIND, DHCPD) or generic Microsoft server products (Microsoft AD) on industry standard servers. These home-grown servers typically utilize a very low percentage of their CPU. They are also prone to vulnerabilities, unnecessarily tedious to

manage, and require on-going maintenance. Virtualization is an excellent way to streamline these inefficiencies.

Furthermore, when preparing the introduction of emerging network technologies such as DNSSEC and IPv6, organizations running TCP/IP networks are required to re-design and re-think their existing network architecture and network management processes as far as DNS, DHCP and IP Address Management is concerned. This creates a logical discontinuity point during which the DDI environment can easily be migrated to virtual computing environment. When deploying a virtualized DDI environment, rather than re-building new virtual machines manually, the easiest way to virtualize any network service is to use productized software packages known as **virtual** or **software appliances**. A software appliance is a self-installing server optimized for a specific task. It contains all software components required for an installation from a hardened operating system to all necessary tools and application(s). Distributed as an installation media (ISO image) that can be used to boot up new virtual machines in just minutes, software appliances automate installation and maintenance processes thereby translating to higher productivity and lower total cost of ownership.

3.1 Benefits of Virtual DNS and DHCP Infrastructure

1. Optimization of CPU utilization and consolidation of computing resources as DNS and DHCP servers are migrated on virtual platforms
2. Savings in hardware and maintenance costs through decreased number of physical servers in the network, software appliance usage, and centralized management
3. Enhanced availability through continuous uptime offered by advanced virtual environments
4. Rapid and flexible deployment, improved scalability through software appliance usage
5. Higher level of information security thanks to purpose-built, hardened software appliances optimized for a specific service

3.2 Virtualization Enables Better Security, Efficiencies

In traditional network environments, DNS and DHCP services are typically run on industry standard servers that host a number of different network services. While pragmatic, this deployment strategy involves several disadvantages:

1. Scalability issues: running several network services on a single server translates to limited scalability during busy-hours and traffic peaks.
2. Single point of failure: if the hardware platform fails, all network services provided by that server will become unavailable.
3. Vulnerabilities: when running several network services on a single server, vulnerability in any one piece of software can make all services running on that server vulnerable to exploitations.

4. Compromises in server and system management: running several network services on a single server translates to compromises in management and maintenance processes.
5. Inflexible deployment processes: making changes or new additions to network architecture is unnecessarily complicated as different network services have been bundled together.

The easiest and the most cost-efficient way to eliminate these problems is to install and run DNS and DHCP software appliances in virtual infrastructure as dedicated virtual machines. Thanks to their purpose-built design and automated software updates, software appliances are less prone to vulnerabilities and provide efficient tools for uncompromised server and system management. By auto-installing on advanced virtualization platforms they offer better scalability, higher availability, and flexible deployment options.

3.3 Software Appliances for DNS, DHCP and IP Address Management

- **Nixu NameSurfer Suite (NSS)** is a secure management system designed for efficient, distributed management of organizations' DNS data and IP address space, supporting emerging technologies such as DNSSEC and IPv6.
- **Nixu Secure Name Server (SNS)** is a secure stand-alone server that can be operated as authoritative and/or as caching/recursive DNS server, supporting emerging network technologies such as IPv6 and DNSSEC.
- **Nixu DHCP Server** is a secure, stand-alone DHCP server with configuration validations and built-in support for DHCP failover pairs, supporting emerging network technologies such as DHCPv6.

3.4 Solution Architecture and Related Considerations

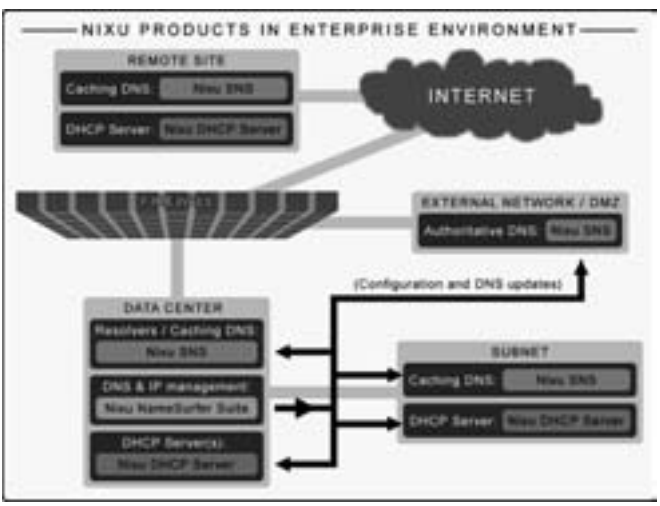


Illustration 1: *DNS and DHCP services in a typical enterprise network*

The above diagram provides a simplified illustration of DNS and DHCP services in an enterprise network environment. The descriptions below provide general guidelines on how DNS and DHCP services can be virtualized in a typical enterprise network environment.

1. **Data Centre:** A virtual infrastructure provides an excellent foundation for virtualizing DNS and/or DHCP services in enterprises. In order to ensure redundancy and availability, Nixu Software recommends the following solution architecture:

   I. Centralized DNS, DHCP and IP Address Management system that is used to manage the entire name and address space. This system would consist of a single virtual machine with support for DNS views, so that internal and external authoritative DNS zones can be managed separately.

   II. Two or more virtual machines for internal authoritative DNS

   III. Two or more virtual machines for internal recursive/caching DNS

   IV. Two or more virtual machines running DHCP servers as failover pair

   If an enterprise has several data centers, it is recommended that a number of DNS and/or DHCP servers are distributed to different geographical locations for added redundancy.

2. **Subnets:** oftentimes, subnets are workstation networks that rely on Microsoft AD for DNS and DHCP services. An alternate solution for providing DHCP and caching/recursive DNS services in subnets is to run software appliances as dedicated virtual machines. If use of Microsoft AD is mandatory, management of the Microsoft AD zone as well as the administrative management of the clients in MS AD subnet can be carried out centrally using Nixu IPAM (Nixu NameSurfer Suite) described in paragraph 1 above.

3. **External network / DMZ:** external authoritative DNS servers residing in DMZ can be run in virtual infrastructure as dedicated virtual machines. The data (authoritative external DNS zones) on these servers is managed using the centralized management system described in paragraph 1.

4. **Remote Site(s):** typically, remote sites are small(ish) networks that rely on Microsoft AD for DNS and DHCP services. An alternate solution for providing DHCP and caching/recursive DNS services in subnets is to run software appliances as dedicated virtual machines. By adopting the alternate approach, also other network services and resources can be run as virtual machines on the same virtualization platform.

4 About Nixu Software

Nixu Software is a VMware Technology Alliance Partner and Citrix Ready Partner specializing in virtualization-ready DNS, DHCP and IP Address Management (DDI) software appliances. All Nixu Products are certified as VMware and Citrix Ready, and can be used to build centrally managed end-to-end DDI environments.

Headquartered in Espoo, Finland, and with regional sales offices in Central Europe and Asia-Pacific, Nixu Software's mission is to offer the best value in industry by creating virtualization-ready DDI solutions that set the benchmark for combined security, ease of use and low cost of ownership.

Nixu Products have an installed base of more than 3.000 instances worldwide, used by:

- one third of all 2.5G and 3G mobile operators worldwide
- enterprises large and small from practically all industry verticals
- well known research and education institutions all over the world.

For further details and contact requests, please visit www.nixusoftware.com.

Virtualisierung komplexer Netzwerke

Prof. Dr. -Ing. Evren Eren, Dipl. -Inf. Markus Breitländer

Fachhochschule Dortmund
Emil-Figge-Str. 42
44227 Dortmund
eren@fh-dortmund.de
markus@breitländer.de

Abstract: Der Begriff Virtualisierung hat viele Ausprägungen und wird in verschiedenen Kontexten der Informationstechnik benutzt. Der wesentliche Gesichtspunkt von Virtualisierung ist es, die Abhängigkeit von Hardware und Software aufzutrennen. Diese Abstraktion führt dazu, dass vorhandene IT-Ressourcen flexibel genutzt und eine höhere Auslastung erzielt werden kann.

Die Virtualisierung von Betriebssystemen hat sich bereits seit geraumer Zeit etabliert. Mit Hilfe moderner Virtualisierungssoftware bestehen heute Möglichkeiten, sogar komplexe IT-Infrastrukturen und deren Komponenten (Subnetze, Router, Switches, Firewalls, DMZ, etc.) zu virtualisieren.

Dieser Beitrag stellt das Projekt "VISA - Virtual IT-Security Architectures" vor, dass im LISA-Labor (www.lisa.fh-dortmund.de) des Fachbereiches Informatik an der FH-Dortmund entwickelt wurde. In diesem Projekt wurde auf Basis von Open Source Software eine komplexe IT-Sicherheitsinfrastruktur für mittelständische Unternehmen virtualisiert. Hierbei wurde die Software KVM (GPL) eingesetzt, welche Vollvirtualisierung auf x86-Hardware ermöglicht. KVM nutzt die Befehlssatzerweiterungen der neueren Intel- und AMD-Prozessoren (Intel VT und AMD-V). Vollvirtualisierung erlaubt den parallelen Betrieb mehrerer Betriebssysteme in einem unveränderten Modus. Die virtuellen Maschinen "wissen" also nicht, dass sie in einer virtuellen Umgebung laufen und müssen nicht angepasst werden. KVM ist eine Abspaltung vom Emulator „QEMU“.

Zur Realisierung der Netzwerkkommunikation zwischen den virtuellen Maschinen dient die Software VDE (Virtual Distributed Ethernet). VDE stellt wichtige Elemente einer virtuellen Infrastruktur, wie Switches und Kabel zur Verfügung.

Im Weiteren wird der Einsatz speziell konfigurierten virtuellen Maschinen, sowie von Hardware-Emulatoren, als virtuelle Router und Firewalls beschrieben. In diesem Zusammenhang wird die Router- und Firewall-Distribution Vyatta und der Router-Emulator Dynamips vorgestellt.

Durch die Kombination der beiden Softwarelösungen KVM und VDE sowie Emulatoren wie Dynamips lassen sich komplexe Netzwerkstrukturen flexibel und "realitätsnah" virtualisieren.

1 LISA (Laboratory for IT-Security Architectures)

Das „Laboratory for IT-Security Architectures – LISA" im Fachbereich Informatik der Fachhochschule Dortmund stellt eine modulare Entwicklungs- und Evaluations-Plattform für IT-Sicherheitsarchitekturen zur Verfügung. Hier können Sicherheitsmodelle und -architekturen exemplarisch diskutiert, erprobt und validiert werden. LISA wird sowohl für die praxisorientierte Lehre und Forschung, als auch als Demo-Center für externe Unternehmen eingesetzt. Dabei werden insbesondere Sicherheitsprobleme und -architekturen von Klein- und Mittelständischen Unternehmen adressiert.

2 VISA

Im Rahmen des Projektes VISA wurde eine komplexe IT-Sicherheitsinfrastruktur virtualisiert. VISA soll in diesem Beitrag als ein Beispiel für eine komplexe virtuelle Netzwerktopologie dienen.

Im Folgenden werden die verschiedenen Softwarelösungen und Techniken vorgestellt, die zur Virtualisierung der verschiedenen Komponenten der LISA-Infrastruktur eingesetzt wurden. Zunächst ging es darum, die bestehenden Systeme samt ihrer Konfiguration in virtuelle Maschinen (VM) zu überführen. Bei diesem Verfahren spricht man von der sog. Physical-to-Virtual Migration (P2V). Hierbei wird eine exakte Kopie der Festplatte erstellt, welche dann später als virtueller Datenträger in die VMs eingebunden wird. In der Regel können hierfür gängige Imaging-Tools wie Clonezilla oder Acronis True Image genutzt werden. Es existieren allerdings auch spezielle P2V-Tools wie beispielsweise Platespin Powerconvert oder VMware Converter, mit denen es möglich ist, Systeme während des laufenden Betriebs zu migrieren. Die so erstellten Images lassen sich von modernen Vollvirtualisierungslösungen meist ohne weitere Anpassungen gebootet. In VISA wurden die Images der Linux-Systeme mit Clonezilla erstellt und zur Migration der Windows-Systeme kam VMware Converter zum Einsatz.

3 Kernel-based Virtual Machine (KVM)

KVM [KVM] ist eine Open Source Virtualisierungslösung (GPL) für Linux, welche Vollvirtualisierung auf x86-Hardware ermöglicht. Hierbei werden die Befehlssatz-erweiterungen der modernen Intel- und AMD-Prozessoren (Intel VT und AMD-V) genutzt. Das Verfahren der Vollvirtualisierung ermöglicht es, mehrere unveränderte Linux- und Windows-Betriebssysteme parallel auf einem System zu betreiben.

KVM wurde 2006 von der israelischen Firma Qumranet veröffentlicht und bereits ein halbes Jahr später in den Linux-Kernel (> 2.6.20) aufgenommen. Qumranet wurde im September 2008 von Red Hat gekauft [HEI]. KVM ist heutzutage in den meisten Distributionen als Standard-Virtualisierungslösung integriert. Sie ist die einzige wirklich „freie“ Virtualisierungslösung, welche die volle Unterstützung der Linux-Community erfährt und profitiert direkt von Weiterentwicklungen am Linux-Kernel. Auch in der Industrie erfährt KVM eine breite Unterstützung. Es kooperieren die Firmen AMD, IBM, Intel, Suse und RedHat bei der Weiterentwicklung der Software.

KVM baut auf dem Emulator QEMU [WIK0] was verschiedene Prozessorarchitekturen emulieren kann. Hierzu zählen PowerPC, ARM, Alpha, m68k, MIPS und Sparc. Er stellt weiterhin die virtuelle Hardware (wie bspw. virtuelle Netzwerkinterfaces) den VMs zur Verfügung.

KVM besteht aus den ladbaren Kernel-Modulen kvm.ko sowie den prozessorabhängigen Modulen kvm-intel.ko und kvm-amd.ko. Bei entsprechender Hardwareunterstützung ermöglicht KVM das äußerst performante Verfahren der hardwarebasierten Vollvirtualisierung. Hierbei werden CPU-Instruktionen direkt auf der Hardware des Hostsystems ausgeführt. Performanz-kritische I/O-Zugriffe lassen sich zusätzlich durch angepasste (paravirtualisierte) Gerätetreiber für Netzwerk- sowie Blockdevices weiter optimiert. Hierbei baut KVM auf dem offenen Standard VirtIO [LIN] auf.

Für Netzwerkverbindungen zwischen VMs sowie zum Anschluss an reale Netzwerke bietet KVM mehrere Netzwerkoptionen. Zur vollständigen TCP/IP-Anbindung wird in der Praxis meist auf TUN/TAP-Interfaces zurückgegriffen. TUN und TAP sind virtuelle Netzwerk-Kerneltreiber, die Netzwerkgeräte über Software simulieren. TUN simuliert ein Point-to-Point-Netzwerkgerät, während TAP ein Ethernet-Gerät darstellt. Für jedes virtuelle Netzwerkinterface einer VM wird im Hostsystem jeweils ein TAP-Interface erstellt. Diese können dann über Netzwerk-Bridges (IEEE 802.1d) des Hostsystems verbunden werden. An eine Bridge lassen sich wiederum auch physikalische Netzwerkschnittstellen des Hostsystems anschließen, was eine Kommunikation der virtuellen Maschinen mit dem Netzwerk des Hostsystems ermöglicht.

Eine weitere Netzwerkoption stellt VDE (Virtual Distributed Ethernet) dar, welche im nächsten Kapitel vorgestellt wird. KVM-VMs können direkt mit VDE-Netzwerken verbunden werden.

4 Virtual Distributed Ethernet (VDE)

Die Software Virtual Distributed Ethernet (VDE) [WIK1] stellt eine allgemeine, virtuelle Infrastruktur zur Verbindung verschiedener Softwarekomponenten zur Verfügung.

VDE ist Ethernet-konform und stellt virtuellen Infrastrukturen virtuelle Switche und virtuelle Kabel zur Verfügung. Es lassen sich virtuelle Maschinen verschiedener Virtualisierungslösungen, Emulatoren, reale Betriebssysteme und Netzwerke miteinander verbinden. Auf Basis von VDE lassen sich sehr einfach und flexibel virtuelle Netzwerke erstellen. Teilbereiche solcher virtuellen Netzwerke lassen sich auf mehrere physikalische Rechner verteilen.

VDE-Netzwerke bestehen aus den folgenden Hauptkomponenten:

VDE-Switch: In der virtuellen Umgebung übernimmt ein VDE-Switch die gleiche Funktion wie ein physikalischer Ethernet-Switch eines realen Netzwerkes. Er verfügt über mehrere Ports, über die verschiedenen Systeme (Computer, Router und andere Ethernet-kompatiblen Endgeräte) einer virtuellen Infrastruktur miteinander verbunden werden können. Intern arbeitet ein solcher virtueller Switch wie ein realer Layer-2-Switch. Er „lernt“ durch Analyse der Paketheader und verwaltet an Hand einer Adresstabelle (Source Address Table) dynamisch die Verbindungen zwischen Hardwareadressen (MAC) und den Switch-Ports.

VDE-Plug: Ein VDE-Plug (dt. Stecker) ist ein Hilfsprogramm, welches sich mit einem Port eines VDE-Switches verbinden kann. Ein VDE-Plug stellt ein universelles Tool dar, mit dem sich der Datenstrom des virtuellen Netzwerkes auf die Betriebssystem-Schnittstellen „stdin“ und „stdout“ umleiten lässt.

VDE-Wire: Als VDE-Wire (dt. Leitung) können Programme eingesetzt werden, die in der Lage sind, Datenströme bidirektional miteinander zu verbinden. Für diesen Zweck hat VDE das Tool dpipe. Jedoch sind auch andere Tools wie bspw. netcat oder ssh einsetzbar.

VDE-Cable: Ein VDE-Cable (dt. Kabel) besteht aus der Kombination zweier VDE-Plugs und einem VDE-Wire. Dieses Konstrukt stellt das virtuelle Pendant zum Netzwerkkabel eines realen Netzwerks dar.

Die o.g. Hauptkomponenten stellen die Basis zum Aufbau von VDE-Netzwerken dar. Die Terminologie orientiert sich hierbei an den Komponenten realer Netzwerke.

VDE-Switches lassen sich durch VDE-Cables untereinander verbinden. Um Loop-Verbindungen zwischen Switches zu verhindern, unterstützten VDE-Switches das Fast-Spanning-Tree Protocol. Da, wie beschrieben, ein VDE-Wire sich auch über Netzwerk-protokolle realisieren lässt, können Verbindungen zwischen VDE-Switchen auch über ein physikalisches Netzwerk hergestellt werden. Somit sind virtuelle Netzwerke, bspw. über eine verschlüsselte SSH-Verbindung sehr einfach zu verbinden. Auch lassen sich somit VPNs realisieren. Darüber hinaus unterstützt VDE die Realisierung von VLANs nach dem IEEE 802.1q-Standard.

Da sich TUN/TAP-Interfaces an einen VDE-Switch anschließen lassen, besteht die Möglichkeit, virtuelle Netzwerke mit dem realen Netzwerk des Hostsystems zu verbinden.

Mit dem VDE-Tool wirefilter besteht eine weitere interessante Möglichkeit, verschiedene Eigenschaften einer Netzwerkverbindung zu simulieren. Dabei werden die Eigenschaften eines VDE-Cables herangezogen. Solchen virtuellen Kabeln können QoS-Parameter wie Bandbreite, Verzögerung, Paketverlust, etc. zugeordnet werden.

Alle wichtigen Komponenten von VDE sind User-Mode-Prozesse. Somit sind auch unprivilegierte Benutzer in der Lage, eigene virtuelle Netzwerke zu erstellen, ohne die Netzwerkkonfiguration des Hosts verändern zu müssen. Über die Rechteverwaltung des Betriebssystems des Hosts kann zudem geregelt werden, welche Benutzer sich mit welchen virtuellen Netzwerken verbinden dürfen. (vgl. [WIK2])

5 Virtuelle Router

Wichtige Komponenten einer Netzwerkinfrastruktur stellen Router- und Firewall-Systeme dar. Je nach Sichtweise dienen sie dazu, Netzwerke bzw. Netzwerksegmente miteinander zu verbinden oder zu trennen. Neben spezieller Hardware können auch normale PCs als sogenannte Software-Router eingesetzt werden, da viele moderne Betriebssysteme über entsprechende Routingdienste verfügen. Gleiches gilt auch für den Einsatz als Software-Firewall. Für den Einsatz als Software-Router sowie als Software-Firewall existieren allerdings auch spezielle, vorkonfigurierte Distributionen mit teilweise mächtigem Funktionsumfang. Bekannte Open Source Lösungen sind beispielsweise SmoothWall, Edian Firewall, pfSense oder Vyatta.

Durch den Einsatz von Virtualisierungssoftware können virtuelle Router und Firewalls in Form von entsprechend konfigurierten virtuellen Maschinen realisiert werden. Hierbei bieten sich die erwähnten Distributionen als Betriebssysteme solcher VMs an.

Vyatta

Im Projekt VISA wurde hierfür die Software Vyatta [VYA1] eingesetzt. Vyatta ist eine umfangreiche Open Source Router- und Firewall-Distribution auf Basis eines angepassten Debian-Linux Systems. Vyatta bietet zahlreiche Netzwerkfunktionen, welche auch auf professionellen Hardware-Routern zum Einsatz kommen. Unterstützt werden diverse Routingprotokolle wie bspw. OSPF, BGP und RIP. Vyatta bietet jedoch nicht nur Routingfunktionen, es lässt sich auch als VPN-Server und Firewall einsetzen. Weitere Features sind u.a. DHCP, NAT, PPPoE, QoS, site-to-site IPsec VPN, SSL-based OpenVPN, RADIUS, 802.1q VLANs, Stateful Inspection Firewalling, High Availability (VRRP), SNMP sowie Intrusion Detection/Prevention und Anti-Virus. (vgl. [VYA2])

Das System lässt sich komfortabel über eine Management-Konsole, sehr ähnlich der von Juniper JUNOS oder Cisco IOS, sowie über eine Web-basierte GUI und üblichen Linux-Systembefehlen konfigurieren und verwalten. Nicht nur virtualisiert stellt Vyatta eine leistungsfähige und erstzunehmende Alternative zu Hardwarerouter wie die von Cisco oder Juniper dar.

Dynamips

Eine weitere Möglichkeit ist der Einsatz eines Hardware-Emulators, um Router in einer virtuellen Infrastruktur abzubilden. Hierbei sei erwähnt, dass sich Durchsatz und Performanz nicht mit Hardware-Routern und virtualisierten Software-Routern vergleichen lassen. In diese Kategorie fällt die Software Dynamips. Dynamips (GPL) wurde im Jahr 2005 zur Emulation von Cisco Router-Plattformen entwickelt. Sie emuliert die Hardware verschiedener Cisco Router und ermöglicht es, unangepasste Cisco IOS Betriebssysteme auf x86-Hardware zu betreiben. Der Emulator wurde mit der Zielsetzung entwickelt, Anwendern die Möglichkeit zu bieten, sich mit dem weitverbreiteten Router-Betriebssystem Cisco-IOS vertraut zu machen, ohne teure Cisco-Hardware zu benötigen. Er stellt somit eine freie Test- und Trainingsplattform für Cisco-IOS zur Verfügung. Die Software kann auch bspw. dazu genutzt werden, Konfigurationen in der emulierten Umgebung zu erstellen und zu testen, um diese dann später auf realen Routern einzusetzen.

Mit der Software GNS3 [GNS] existiert ein komfortables grafisches Frontend für Dynamips. Auch wird die Emulation von Cisco PIX- und ASA-Firewalls sowie das Juniper-Betriebssystem JunOS unterstützt. Hierzu werden speziell veränderte Versionen des Emulators Qemu benutzt. Diese inoffiziellen Versionen von Qemu sind auch unter den Namen PEMU [PEM] und JEMU (Codename Olive) bekannt [JUN].

6 Fazit

Dieser Beitrag zeigt auf, dass sich mit Hilfe moderner Virtualisierungssoftware sogar komplexe IT-Infrastrukturen bis auf Layer 1 des ISO-OSI-Modells virtuell abbilden lassen. Das Projekt "VISA - Virtual IT-Security Architectures" im LISA-Labor (www.lisa.fh-dortmund.de) des Fachbereiches Informatik an der FH-Dortmund entwickelte hierzu eine Lösung auf Basis von KVM sowie VDE. Des Weiteren wurde die Router- und Firewall-Distribution Vyatta und der Router-Emulator Dynamips eingesetzt.

Literaturverzeichnis

[KVM] http://www.linux-kvm.org/page/
[HEI] http://www.heise.de/open/Red-Hat-investiert-in-Virtualisierung-Update--/news/meldung/115452
[LIN] http://www.linux-kvm.org/page/Virtio
[WIK0] http://wiki.qemu.org/Main_Page
[WIK1] http://wiki.virtualsquare.org/index.php/Introduction#Virtual_Square_Networking
[WIK2] http://wiki.virtualsquare.org/index.php
[VYA1] http://www.vyatta.org/
[VYA2] http://www.vyatta.com/downloads/datasheets/vyatta_virtual_datasheet.pdf
[GNS] http://www.gns3.net/
[PEM] http://www.blindhog.net/pemu-cisco-pix-emulator/
[JUN] http://juniper.cluepon.net/index.php/Olive

DatabaseHypervisor – ein Ansatz für relationale Datenbanken in der Cloud

Benjamin Schmidt

Zimory GmbH
Stralauer Platz 33-34
10243 Berlin
schmidt@zimory.com

Abstract: Aktuelle Cloud-Produkte und Infrastrukturen haben mittlerweile eine breite Akzeptanz gefunden, wenn es um den Präsentations- und Applikations-Layer einer typischen 3-gliedrigen Web-Anwendung geht. Für den dritten Layer, der Datenbank, haben sich NoSQL oder Key-Value-Datenbanken wie z.B. Googles BigTable oder Amazons SimpleDB etabliert. Diese bieten die in der Cloud-Umgebung geforderte horizontale Skalierbarkeit, wenn auch mit einem Haken – sie haben ein nichtrelationales Design und orientieren sich deshalb nicht am SQL-Standard.

Bei traditionellen Enterprise-Anwendungen sind jedoch diese nicht-relationalen, verteilten Data Stores keine Option. Die Akzeptanz von Cloud Computing in die IT-Infrastruktur eines Großunternehmens mit großen Datenbankservern hat deshalb noch keinen nennenswerten Zustrom erhalten. Relationale Datenbanken sind bisher aus sowohl Virtualisierten-, wie auch Cloud-Umgebungen gehalten worden und sind immer noch zu sehr an einen einzigen Server (oder Cluster) und Ort gebunden.

Dieser Beitrag vergleicht bisherige Cloud-Datenbank-Implementierungen und stellt einen neuen Ansatz zur Datenbankvirtualisierung vor. Dieser Ansatz wird anhand der Architektur des Prototyps "Spree-DatabaseHypervisor" veranschaulicht.

1 Skalierung von Datenbanken

Die Architektur traditioneller Unternehmensapplikationen sind in der Regel in drei logischen Schichten unterteilt: Präsentation, Anwendung (oder Business Logic) und Data.

Auf der Präsentations- und Applikationsschicht ist eine horizontale Skalierung eine weit verbreitete Praxis, um eine Lastverteilung, erhöhte Ausfallsicherheit und größere Leistung zu gewährleisten. In horizontal skalierten Architekturen, hat jedes System seine eigenen, unabhängigen Ressourcen wie Prozessoren, Arbeitsspeicher und Storage. Diese

"shared nothing"-Architektur bietet eine höhere Verfügbarkeit da es keinen Single-Point-of-Failure (SPOF) enthält. Zur Erhöhung der Leistung bei horizontalen Architekturen werden einfach neue Knoten dem Cluster hinzugefügt.

Im Gegensatz dazu, wird bei der unteren Schicht, der Datenschicht, eher vertikal skaliert. SMP (Symmetrisches Multiprozessorsystem) Maschinen dominieren auf der Datenebene, bei der die Ressourcen, wie Prozessor, Speicher und Storage, geteilt werden. Zur Erhöhung der Verfügbarkeit von Anwendungen und um die Gefahr des SPOFs unter Verwendung von SMP-Systemen bei der Datenschicht zu mildern, wird oft eine von mehreren Datenreplikationsmethoden angewandt, um ein zusätzliches System für Failover-Zwecke zu erhalten. Um die Gesamtleistung zu steigern, müssen zusätzliche Ressourcen dem System hinzugefügt werden (Upgrade); ist das System nicht mehr Upgrade-fähig, muss ein größeres (und meistens teureres) System beschafft werden.

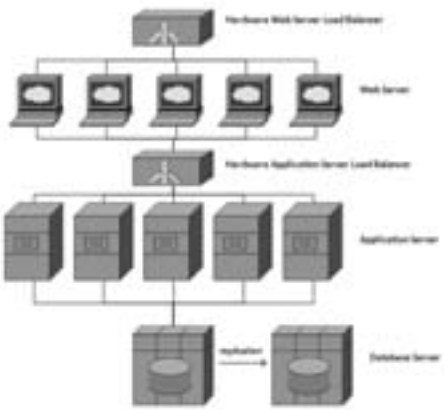


Abbildung 1: Schichten einer Unternehmensapplikation

Da beim Design heutiger Webapplikationen bereits auf eine Unterstützung von verteilten Systemen geachtet wird und somit eine horizontale Skalierung transparent ermöglicht, hat das Prinzip des Cloud Computing sehr schnell sehr großen Anklang gefunden, da es keine oder nur eine sehr geringe Adaptionszeit der Applikation benötigt. Beim Infrastructure-as-a-Service Cloud Computing wird dem Entwickler bzw. dem Applikationsmanager nicht nur die Verantwortung für die darunterliegende Hardwareinfrastruktur abgenommen, auch eine horizontale Skalierbarkeit, mit einer theoretisch unbegrenzten Anzahl von Knoten, ist ein Teil des Cloud Computing Angebots.

1.1 Konsistenzbetrachtung

Da es bisher unmöglich war, Datenkonsistenz über mehrere Server hinweg zu gewährleisten, ohne dabei Abstriche bei der Leistung zu machen, wurde die Brauchbarkeit einer horizontalen Architektur in der Datenschicht eingeschränkt.

Die Gültigkeit und Bedeutung von Brewers CAP Theorem [Br00] in Bezug auf Datenbanksysteme in der Cloud wurde sowohl von Prof. Michael Stonebraker [St10], dem ersten Entwickler von PostgreSQL, wie auch von Amazons CTO, Dr. Werner Vogels hervorgehoben.

Kurz zusammengefasst, besagt das CAP Theorem, dass es beim Entwerfen und der Implementierung von Applikationen in verteilten Umgebungen, drei bedeutende Anforderungen gibt, deren gleichzeitige Einhaltung unmöglich ist.

C: **Consistency** – Hierbei ist das Ziel, auf bei Transaktionen über mehrere Knoten hinweg, eine „all-or-nothing" Sematik einzuhalten. Multiple Werte für das gleiche Datenstück sind verboten. Traditionelle relationale Datenbank fokussieren auf ACID Eigenschaften (atomicity, consistency, isolation, durability) und gewährleisten Konsistenz z.B. mit Isolationsmethoden. Wird Replikation eingesetzt, müssen auch alle Replikate einen konsistenten Status besitzen.

A: **Availability** – Hier ist das Ziel einer permanenten Verfügbarkeit des Systems. Sollte ein Knoten ausfallen, kommt es zum Failover auf einen anderen Knoten ohne die Gesamtverfügbarkeit zu beeinflussen. Im Fall von DBMS kann Replikation eingesetzt werden.

P: **Partition Tolerance** – Im Falle einer Netzwerkunterbrechung können zwei oder mehrere Knotengruppen entstehen, wobei Knotengruppe A nicht mehr mit Knotengruppe B kommunizieren kann.

Das CAP Theorem besagt nun, das im Fehlerfall nicht alle drei Ziele zu verwirklichen sind. Eine Eigenschaft muss zwangsläufig kompromittiert werden.

Zum Leben wird das CAP Theorem erst bei der Skalierung von Applikationen erweckt. Bei einem niedrigen Transaktionsvolumen haben kurze Verzögerungen, verursacht durch die Konsistenzwahrung der Datenbank, keinen merklichen Einfluss auf die Gesamtleistung des Systems. Laut Vogels [Vo08] treten diese Herausforderungen erst bei einer Skalierung von verteilten Systemen auf weltweiter Ebene hervor. Ereignisse, die in der Regel eine geringe Eintrittswahrscheinlichkeit haben, treten bei Milliarden von Anfragen garantiert ein und müssen schon im Voraus beim Entwurf des Systems und in der Architektur berücksichtigt werden.

Vogels argumentiert, dass Dateninkonsistenz in einer großen verteilten Umgebung aus zwei Gründen toleriert werden muss, da:

1. Die Schreib- und Leseperformance unter hoch nebenläufigen Bedingungen verbessert werden muss

2. Partitionen aufgrund von Netzwerkproblemen nicht zu einem Ausfall eines Teilsystems führen dürfen obwohl die Knoten noch aktiv sind.

C wird also für A und P geopfert. Die Konsistenzverantwortung wird der Applikation übertragen. Im Gegensatz zum "strong consistency" Modell, bei dem nach einem erfolgreichen Update *jeder* Zugriff bereits den neuen Wert liefert, und dem „weak consistency" Modell, bei dem das System *keine* Garantie übernimmt bzw. erst nach einem „*inconsistency window"* garantiert, dass nachfolgende Zugriffe den neuen Wert liefern, befürwortet Vogels das "eventually consistent" Modell. Diese spezielle Form des weak consistency Modells garantiert, dass letztendlich (irgendwann) alle Zugriffe den neuen Wert liefern solange das Objekt keine neuen Updates erhält. Das inconsistency window wird beeinflusst von Faktoren wie Kommunikationsverzögerung (Latenzzeit), der Systemlast und die Anzahl der involvierten Replikate[1] – aber auch nur solange währenddessen keine Fehler auftreten!

Das eventual consistency Modell hat eine Paar Variationen zu bieten, wie z.B. „read-your-writes consistency" oder „monotonic read/write consistency" die versuchen, einige der Schwächen des Modells, zu kompensieren. Trotzdem ist es offensichtlich, dass Unternehmensapplikationen, die heute beispielsweise mit einem hochverfügbaren Oracle Database Cluster auf der Datenschicht arbeiten, mit dieser „irgendwann Philosophie" nicht zufrieden sein können. Eine berechtigte Frage auf Vogels Blogpost lautet daher: „Wie wird die New York Stock Exchange beispielsweise, konsistente Kurse an alle Investoren liefern können? Ist ‚irgendwann' gut genug?"

1.2 NoSQL Datenbanken

Amazons SimpleDB, ein verteiltes Datenbankmanagementsystem auf Basis von einfachen Key-Value Stores, basiert auf dem eventual consistency Modell. Da man sich aber bei Amazon der Schwächen bewusst ist, wurde Anfang 2010 eine „strong consistency" Option eingeführt – „consistent read". Bei dieser Option reflektieren Leseoperationen immer das vorherige Update. Damit will man bei Amazon speziell Applikationen adressieren, die ursprünglich eine klassische relationale Datenbank benützen, um so den Umstieg auf SimpleDB zu erleichtern.[2]

Alternativen zum bisherigen, fast vierzig Jahre altem Konzept einer relationalen Datenbank werden unter der Bezeichnung „NoSQL Datenbanken" vereint. Dabei umschließt dieser Begriff eine Reihe von Datenbankkonzepten und Technologien, viele davon zur Lösung eines speziellen Bedürfnisses. Gemein haben NoSQL Datenbanken, dass sie Konsistenz für Skalierbarkeit (Größen- wie auch Komplexitätsskalierung) opfern und keine relationale Datenbank sind.

[1] Als Beispiel für das „eventually consistent" Modell dient DNS (Domain Name System), bei dem Updates nach einem bestimmten Schema und in Verbindung mit einem zeitgesteuertem Cache propagiert werden; letztendlich werden alle Clients die Updates sehen.

[2] Auch diese Option wird aber von Vogels relativiert: "Even under extreme failure scenarios, such as complete datacenter failures, SimpleDB is architected to continue to operate reliably. However when one of these extreme failure conditions occurs it may be that the stronger consistency options are briefly not available while the software reorganizes itself to ensure that it can provide strong consistency. Under those conditions the default, eventually consistent read will remain available to use." [Vo10]

Generell können NoSQL Datenbanken in vier Klassen (mit einigen Beispielen) eingeteilt werden:

1. Document-oriented Databases: MongoDB, CouchDB
2. Graph Databases: Neo4j, AllegroGraph
3. Tabular/Column-oriented Databases: Googles BigTable
4. Key-Value Stores: Amazons Dynamo und SimpleDB, Cassandra (auch column-oriented), Voldemort

Außer dem weicheren Konsistenzmodell und dem Fehlen einer einheitlichen Datenbanksprache, gibt es für NoSQL Datenbanken auch keine einheitlichen Abstraktionsbibliotheken wie z.B. JDBC oder ODBC. In eine Entscheidungsmatrix für einen IaaS- oder PaaS-Provider muss demnach auch immer der Vendor Lock-in für die Datenschicht betrachtet werden. Dies ist ein weiterer Grund dafür, warum viele Unternehmensapplikationen nicht in eine kommerzielle Cloud ausgelagert werden.

Das Problem der fehlenden Clienttreiber hat sich Xkoto (aufgekauft von Teradata) mit dem Produkt Gridscale angenommen. Gridscale besteht aus einer Reihe von Clienttreibern (ODBC, JDBC, CLI), einem (oder aus Redundanzgründen mehreren) Gridscale Server, die direkt zwischen den Applikationsserver und den Datenbankserver platziert werden, sowie aus DB Connectoren zu jedem Datenbankserver. Somit stellt sich Gridscale den Anwendungen gegenüber als eine einzelne, voll funktionsfähige Datenbank dar. [Xk08]

2 Spree DatabaseHypervisor Prototype

Das Konzept vom Zimory Projekt „Spree“ geht noch ein Stück weiter – selbst die Clienttreiber werden nicht modifiziert, an der Applikation ändert sich somit überhaupt nichts, und auch an den Datenbankservern müssen keine Modifikationen vorgenommen werden. Die Kernkomponenten des Systems ist der „DatabaseHypervisor“, einer neue Schicht zwischen dem JDBC Treiber und des Datenbankservers, der als Interceptor dient. Zusätzlich zur ursprünglichen „Master Datenbank“ können eine beliebige Anzahl von „Satelliten“ an den DatabaseHypervisor angeschlossen werden. Die Satelliten übernehmen alle Leseabfragen, somit wird die Master-Datenbank nur bei Schreiboperationen belastet.

Um die Konsistenz des Systems auch über mehrere Satelliten zu garantieren, werden verschiedene Strategien der Replikation verwendet. Details hierzu und zu den unten abgebildeten Benchmarking-Ergebnissen bleiben künftigen Veröffentlichungen überlassen.

Der erste Prototyp fokussierte auf eine Oracle 11g Datenbank als Master sowie auf mehrere Oracle 11g Datenbanken als Satelliten. Zur Performancemessung wurde der TPC-W Benchmark verwendet. Abbildung 2 zeigt den Leistungsgewinn

(Transaktionen/s) beim Hinzufügen von Satelliten mit dem TPC-W Benchmark (konfiguriert mit 5% Updates, 95% Selects). In der ersten Phase wird die reine Leistung des Masters gemessen. Nach dem Hinzufügen des ersten Satelliten, steigt die Gesamtleistung (schwarze Linie) nur minimal. Jedoch ist zu sehen, dass der Master nur noch Schreiboperationen (rote Linie) erhält, während der Satellit die Leseoperationen (grüne Linie) erhält. Ein deutlicher Gewinn wird ab dem zweiten Satelliten erzielt.

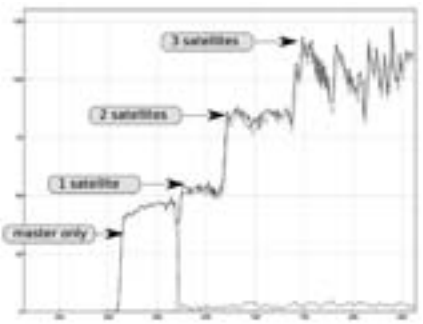


Abbildung 2: TPC-W Benchmark (5% Updates); Leistungsgewinn bei bis zu drei Satelliten

Da sowohl das Hinzufügen von Satelliten wie auch das Entfernen sehr schnell, transparent und on-the-fly passiert, lassen sich kurzfristige Lastspitzen auf der Datenschicht sehr gut bewältigen. Ein mögliches Szenario, wie es momentan auch bei Zimory entwickelt wird, ist der Einsatz der Satelliten auf einer pay-on-demand Basis als virtuelle Maschinen in der Cloud. Dabei können Resources eines externen Providers verwendet werden oder auch eigene Resourcen als „Enterprise Cloud“ bereitgestellt werden.

Im Gegensatz zur NoSQL Initiative, bietet Zimory Spree somit eine horizontale Skalierbarkeit für Applikationen, die relationale Datenbanken benötigen und nicht auf Konsistenz verzichten möchten. Entwickler können damit weiterhin SQL für neue Applikationen benützen und an existierenden Systemen müssen keinerlei Änderungen vorgenommen werden, um sie in der Cloud zu betreiben, weder auf der Applikationsseite noch auf der Datenbank.

Literaturverzeichnis

[Br00] Brewer, E. A.: Towards Robust Distributed Systems, Proceedings of the Annual ACM Symposium on Principles of Distributed Computing, 2000

[St10] Stonebraker, M.: Errors in Database Systems, Eventual Consistency, and the CAP Theorem, http://cacm.acm.org/blogs/blog-cacm/83396-errors-in-database-systems-eventual-consistency-and-the-cap-theorem

[Vo08] Vogels, W.: Eventually Consistent – Revisited, http://www.allthingsdistributed.com/2008/12/eventually_consistent.html

[Vo10] Vogels, W.: Choosing Consistency, http://www.allthingsdistributed.com/2010/02/strong_consistency_simpledb.html

[Xk08] Technical Whitepaper Gridscale ® Database Virtualization Software;

Die Schwachstellen virtueller Systeme

Holger Engelland

Manager Data Recovery Engineering
Kroll Ontrack GmbH
Hanns-Klemm-Str. 5
71034 Böblingen
hengelland@krollontrack.de

Abstract: Holger Engelland, Manager Data Recovery Engineering bei Kroll Ontrack, erklärt die Risiken der Virtualisierung und erläutert Best Practices bei Datenverlust.

1 Vorteile der Virtualisierung

Unter Virtualisierung versteht man eine Software-Technologie, die unsere IT-Landschaft verwandelt und einschneidende Veränderungen der Computing-Umgebung zur Folge hat. Während die gängige Computer-Hardware in den meisten Fällen mit nur einem Betriebssystem gleichzeitig arbeiten kann, verhilft die Virtualisierung den Unternehmen zu einer Überschreitung dieser Grenze. In einer virtuellen Computing-Umgebung können mit einer Maschine gleichzeitig mehrere Betriebssysteme betrieben werden. Damit werden Nutzen und Flexibilität der IT-Umgebung erhöht.

Die Virtualisierung der IT-Systeme hat viele Vorteile:

- Eine Konsolidierung der Ressourcen kann eine Zeit- und Geldersparnis bedeuten
- Reduzierung von physikalischen Serverkapazitäten
- Bereitstellung von Thin Clients für die Mitarbeiter
- Optimierung und Rationalisierung der IT-Infrastruktur

Diese Vorteile und Kosteneinsparungen erklären die zunehmende Virtualisierung innerhalb der Unternehmen. Mit der Implementierung dieser Technologien ist allerdings auch ein gewisses Risiko verbunden. CIOs und IT-Administratoren sollten alle Risiken berücksichtigen, die die Umstellung auf eine virtuelle IT-Umgebung mit sich bringt.

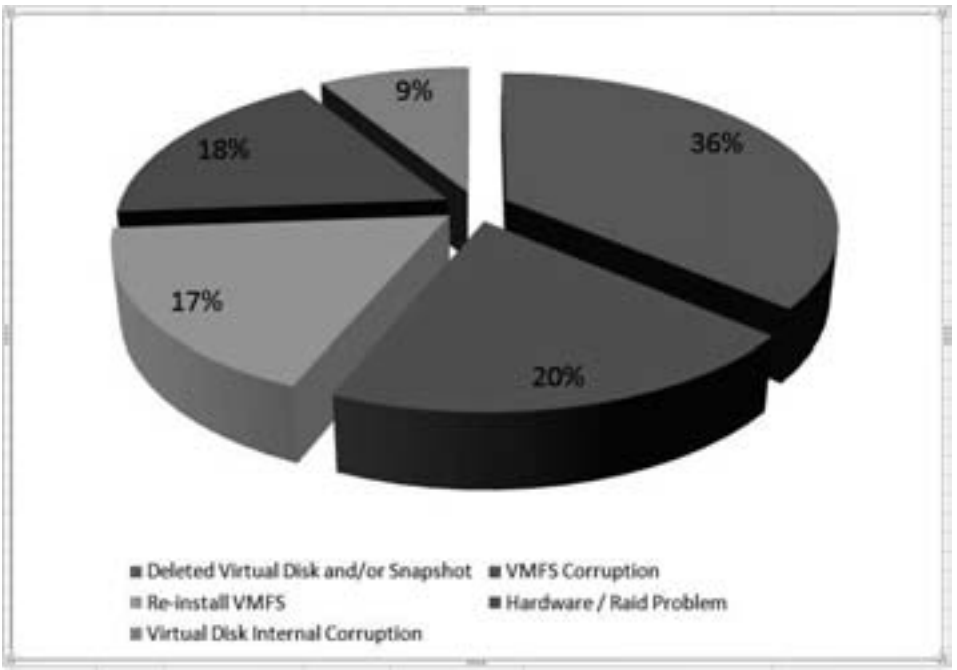


Abbildung 1: Die häufigsten Ursachen für virtuellen Datenverlust

2 Wo machen sich die Schwachstellen bemerkbar?

Wie bei jeder neuen IT-Implementierung müssen die Prozesse und Abläufe vor der Bereitstellung entsprechend verändert werden. Das gilt besonders dann, wenn Unternehmen virtuelle Systeme von der Testphase in die tatsächliche Arbeitsumgebung übertragen. Findet hier keine vernünftige Planung statt, könnte das einen verheerenden Datenverlust zur Folge haben.

Viele Fälle von Datenverlust sind bedingt durch die Zunahme virtueller Umgebungen. Dazu gehen oft Daten durch einen Systemausfall des Betriebssystems oder eines physikalischen Gerätes während der Konfiguration der virtuellen Umgebung verloren. Kroll Ontrack erhielt 2009 58 % mehr Anfragen zur Datenrettung in virtuellen Umgebungen als 2008. Das zeigt deutlich das Ausmaß potentieller Schwachstellen bei der Implementierung virtueller Infrastrukturen.

Eines der Hauptprobleme bei der Verwaltung virtueller Umgebungen hat seinen Ursprung in einer unzureichenden Speicherungs- und Serverorganisation. Ein verbreitetes Szenario ist die Erstellung von virtuellen Servern, ohne deren Bestimmung zu dokumentieren. Dadurch kann die Nachverfolgung wichtiger Daten und Anwendungen extrem erschwert werden. Hinzu kommt, dass das Frontend-Bedienfeld sehr einfach gestaltet ist. Beispielsweise kann ein unerfahrener Systemadministrator durch einen einfachen Klick mit der rechten Maustaste oder durch Betätigen der Entfernungstaste das komplette virtuelle Datenverwaltungssystem einer virtuellen Maschine (VMFS) verschwinden lassen.

Alte oder unnötige Datensicherungsprozesse oder eine Überkonsolidierung bergen ebenfalls gefährliche Risiken für die Daten in einem virtuellen System. Unternehmen, die unseren Service genutzt haben, sind z. B. fälschlicherweise davon ausgegangen, dass die Snapshot-Funktion eines VMware-Systems eine ausreichende Alternative zu einem zuverlässigen Datensicherungsprozess ist. Obwohl die Snapshot-Funktion zum Testen von Konfigurationen sinnvoll ist, ist sie auf keinen Fall ein Ersatz für ein traditionelles, umfangreiches Back-Up-System. Die Beschädigung einer VMFS oder der Ausfall eines physikalischen Servers kann durch die Nutzung eines Snapshots nicht behoben werden.

Eines der Hauptargumente für den Kauf von virtuellen Systemen ist die Möglichkeit der Datenkonsolidierung. Dieses kann sich allerdings, wenn keine umfangreiche Datensicherungsstrategie vorhanden ist, als größte Schwachstelle erweisen. Von einer übermäßigen Konsolidierung wichtiger Daten und Anwendungen ist unbedingt abzusehen. Bestimmte Anwendungen z. B. sind für eine physikalische Server-Umgebung am besten geeignet oder sie erfordern unabhängige Datensicherungsprozesse. Anwendungen mit einer hohen Input/Output-Rate gehören zu dieser Gruppe.

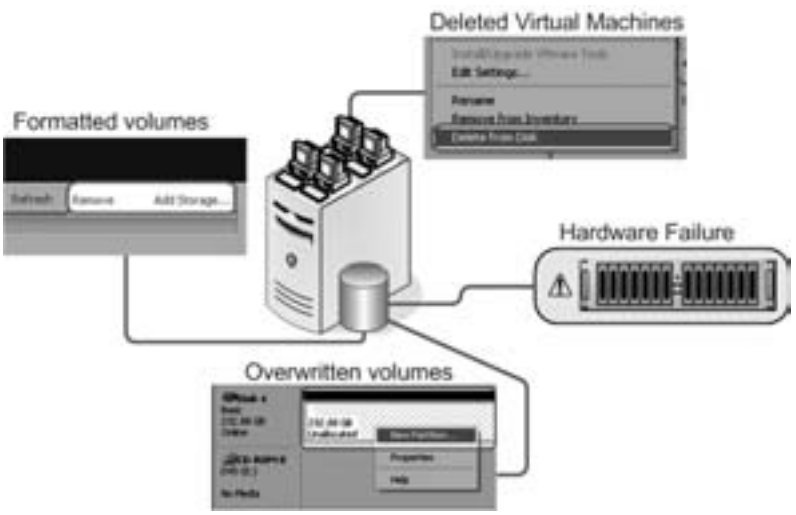


Abbildung 2: Typische Datenrettungs-Szenarien in virtuellen Umgebungen:
Formatierte Volumes, gelöschte Virtuelle Maschinen, Hardware-Fehler, überschriebene Volumes

2.1 Wie können Unternehmen ihre Risiken minimieren?

Um einem womöglich katastrophalen Datenverlust vorzubeugen, muss ein Unternehmen vor dem Rollout eines virtuellen Systems sorgfältig planen.

Wie alle umfangreichen Implementierungen muss eine teilweise oder vollständige Virtualisierung der IT-Infrastruktur eines Unternehmens sorgfältig überlegt und geplant werden. Für IT-Administratoren ist es ratsam, sich den ungünstigsten Fall auszumalen und von hinten nach vorne zu arbeiten. Dabei können an bestimmten Punkten Sicherungsmaßnahmen eingebaut werden.

Anstatt sich auf Snapshots zu verlassen, werden in regelmäßigen Abständen Back-Ups durchgeführt. Zusätzlich sollte ein firmenübergreifender Datensicherungsprozess für sämtliche Daten existieren.

Unternehmen sollten auch nicht vergessen, dass unabhängig davon, wie viele Daten im virtuellen System konsolidiert wurden, alle Daten immer noch physikalisch an irgendeinem Ort gespeichert werden müssen. Die kompletten Datenverwaltungsprozesse sollten grundsätzlich immer auf den umfassenden Schutz dieser physikalischen Speicher ausgerichtet sein. Dabei müssen vor allem die besonderen Eigenarten virtueller Systeme berücksichtigt und entsprechend eingeplant werden.

Wichtig ist es außerdem sicherzustellen, dass alle Mitarbeiter, die virtuelle Datensätze erstellen und bearbeiten können, sich an Abläufe halten. Damit wird das Risiko verringert, dass ein wuchernder Serverzuwachs entsteht und. gewährleistet, dass wichtige Daten und Anwendungen überwacht und gefunden werden können. Bei Bedarf müssen die Mitarbeiter Fortbildungsmaßnahmen über die Komplexität eines virtuellen Systems besuchen und eventuelle Wissenslücken geschlossen werden.

Eine Teilnehmerumfrage[1] bei einem kürzlich abgehaltenen Webinar zur Virtualisierung hat gezeigt, dass die Unternehmen hauptsächlich deshalb nicht virtualisieren, weil ihre IT-Teams nicht über die entsprechenden erforderlichen Fachkenntnisse verfügen. Die Unternehmen müssen erkennen, dass es keine Abkürzung zu einer erfolgreichen Implementierung gibt und sicherstellen, dass ihre Mitarbeiter ausreichend geschult sind. Tatsächlich waren 65 % aller Datenrettungen, die Ingenieure von Kroll Ontrack 2009 in virtuellen Umgebungen durchgeführt haben, auf einen Mitarbeiterfehler zurückzuführen. Dadurch wird die Komplexität einer Implementierung, der Verwaltung und/oder der Migration zu einer virtuellen IT-Umgebung klar erkennbar.

2.2 Was kann man bei Datenverlust tun?

Der beste Schutz sind fest vorgeschriebene Prozesse für die Verwaltung der virtuellen Infrastruktur, die von den Mitarbeitern eingehalten werden müssen. Die Kontrolle und Durchsetzung wird von Management-Teams und solchen Mitarbeitern durchgeführt, die eine Schlüsselfunktion in der Virtualisierungsinfrastruktur innehaben. Sollte es aber zum Schlimmsten kommen, muss man schnell handeln.

Es ist daher sinnvoll, einen Notfallplan zu haben, der die nächsten Schritte festlegt, wie z. B. die vorherige Absprache mit einem Datenrettungs- Anbieter, um den Rettungsprozess zu beschleunigen. Die erste Stunde nach dem Datenverlust ist oft die wichtigste und eine schnelle Reaktion unerlässlich.

Kommt es zu einem Datenverlust in der virtuellen Infrastruktur, ist eine Identifizierung der Ursache und die Datenwiederherstellung eine komplexe Prozedur. Dazu benötigt man große Fachkompetenz und ein umfassendes Wissen über virtuelle Umgebungen. Als erstes schlagen wir eine logisch aufgebaute Geschäfts- und Risikoanalyse durch die Beantwortung folgender Fragen vor:

- Ist eine Datenrettung der verlorenen Daten von der Back-Up- Datenbank möglich? (Hierbei kommt es auf das Alter dieser Sicherheitskopien an.)

- Was kostet es die Firma, wenn die Daten nicht wiederhergestellt werden? Was kostet im Vergleich dazu die Rettung dieser Daten?

In den meisten Fällen bieten die Kundenberater der Data Recovery Unternehmen den Firmen eine kostenlose Diagnose ihres Datenverlustes an – abhängig vom Datentyp kann Kroll Ontrack sogar durch Fernzugriff bei der Datenrettung helfen.

Abschließend ist noch zu sagen, dass man niemals auf eigene Faust versuchen sollte, die Daten zu retten - oft gibt es nur eine einzige Chance, um die Daten wiederherzustellen. Wenn Sie versuchen, die Daten selbst zu retten, kann das den Wiederherstellungsprozess verkomplizieren oder sogar die Daten komplett zerstören oder unbrauchbar machen.

Weitere Informationen finden Sie unter: http://www.ontrack.de/VMware-datenrettung/

Literaturverzeichnis

[ESG10] Enterprise Strategy Group: People May Be the Weakest Link in the Server Virtualization Chain, 2010, http://www.enterprisestrategygroup.com/2010/02/people-may-be-the-weakest-link-in-the-server-virtualization-chain/

Sichere Fernwartungszugriffe

Michael Sorg

Telekom Deutschland GmbH
PSSM Inhouse
Nauheimer Strasse 101
70372 Stuttgart
m.sorg@telekom.de

Abstract: Die folgende Arbeit befasst sich mit dem Thema der Fernwartungszugriffe. Welche Arten von Fernwartungszugängen sind in der Praxis zu finden und welche sicherheitstechnischen Gefahren ergeben sich daraus. Weiterhin wird ein Konzept vorgestellt, welches durch Einsatz eines Rendezvous-Servers eine Erhöhung des Security-Levels ermöglicht.

1 Einführung und Begriffserklärung zum Thema Fernwartung

1.1 Was ist Fernwartung?

Als Fernwartungen kann man alle Tätigkeiten bezeichnen, die ausgeführt werden, ohne das an dem zu wartenden System ein ausführender Techniker physisch vor Ort präsent sein muss. Eine Fernwartung kann „online“ oder „offline“ realisiert werden. Die Fernwartung eines Systems, die durch eine z.B. zugesande CD realisiert wird, welche ein selbstausführendes Updateprogramm enthält, entspricht einer „offline Fernwartung“. Eine „online Fernwartung“ entspricht dem Zugriff, eines aus LAN-Sicht externen Partners, welcher sich, im Regelfall, über das Internet und ein Zugangssystem auf das zu wartende System verbindet und die vorzunehmenden Wartungsarbeiten vornimmt. Die vorliegende Produktstudie befasst sich mit der Sicherung der „online Fernwartung“.

1.2 Wozu wird Fernwartung benötigt?

Gerade im Zeitalter der Globalisierung expandieren nationale Firmen in weit verstreute Standorte. Diese Globalisierung hat zur Folge, dass die Systeme vor Ort zu administrativen Zwecken nur unter erhöhten Aufwänden zur „direkten“ Konfiguration erreichbar sind. Auch Hersteller, welche ihren Abnehmern regelmäßige Updates und Wartungen, sowie Service- und Instandhaltungsarbeiten anbieten, benötigen eine Möglichkeit die dezentral aufgestellten Systeme zentral zu administrieren und auch zu steuern. Hierzu dient die Fernwartung. Durch einen Fernwartungszugriff ergeben sich folgende Vorteile:

- ermöglicht Administratoren und Herstellern einen zentralen Zugriff auf dezentral aufgestellte Systeme
- Zielsysteme sind 7x24h, rund um die Uhr, erreichbar
- Relevante und wichtige Daten können auf das System geladen werden
- Relevante und wichtige Daten können von dem System geladen werden
- Kein physikalischer Zugang zum System nötig

1.3 Wie realisiert man eine Fernwartung?

Wie bereits in 1.1 beschrieben werden im Regelfall die Fernwartungszugriffe über das Internet realisiert, denn über das Internet hat man die Möglichkeit eines weltumfassenden Netzes, welches bis auf die Einwahlkosten des Providers kostenlos zur Verfügung steht. Der Zugriff auf die zu wartenden Systeme wird in der Regel über ein von dem Hersteller vorgegebenes Verfahren bereitgestellt. Hierzu dient ein logisches Interface für den administrativen Zugang auf dem System. Der Zugriff auf dieses Interface ist dann von dem jeweiligen Netzbetreiber zu Wartungszwecken freizuschalten. Solche Zugriffe sind, aufgrund des Zugriffsweges über das Internet, meist verschlüsselte Verfahren, wie HTTPS-, SSH- oder VPN-Zugriffe. Durch diese Verfahren sind die in der Verbindung gesendeten Daten vor der Einsicht Dritter geschützt.

1.4 Welche Gefahren ergeben sich aus dem jetzigen Zugriffsverfahren?

Sollte ein Fernwartungszugriff nicht durch ein verschlüsseltes Zugriffsverfahren bereitgestellt werden, besteht die Möglichkeiten durch Man-in-the-Middle Attacks oder Snifferattacken, das sicherheitsrelevante Daten, wie z.B. Passwörter oder Ablaufprozesse, von Dritten im Internet, aber auch im LAN mitgelesen und ausgenutzt werden können. Bild 1.4.1 zeigt den unverschlüsselten Zugriff auf ein System

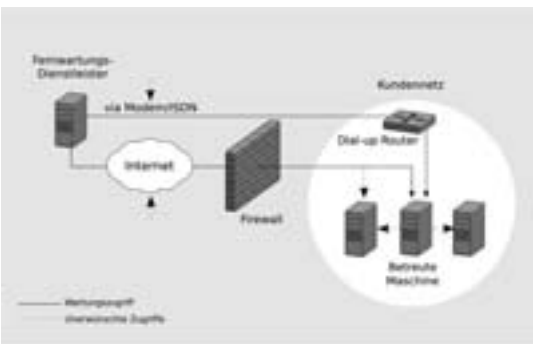


Bild 1.4.1

Der verschlüsselte Fernwartungszugriff verhindert zwar die Möglichkeiten Dritter im Netz, die Verbindung mitzulesen, der verschlüsselte Zugriff beschränkt aber auch den lokalen Administrator zu überprüfen, welche Verbindungen oder Handlungen durch die verschlüsselte Verbindung durchgeführt werden. Ein Fernwartungszugriff öffnet somit eine „Türe“ in das LAN der jeweiligen Firma und stellt somit ein Sicherheitsrisiko dar. Da sich zu wartende System oft in einem bestehenden Computernetzwerk befinden, besteht die Möglichkeit von einem zu wartenden System weiter auf andere Systeme zu „springen“ oder durch Spionagemaßnahmen weitere Informationen über die interne Netztopologie eines LANs zu erhalten, was zu weiteren Angriffsmaßnahmen genutzt werden kann. Die Möglichkeiten einen freigeschalteten Zugriff auf ein System boshaft auszunutzen sind groß. Es gilt somit den Zugriff so sicher wie möglich, aber auch so transparent wie möglich zu gestalten. *Bild 1.4.2* zeigt den verschlüsselten Fernwartungszugriff und noch bestehende Gefahren

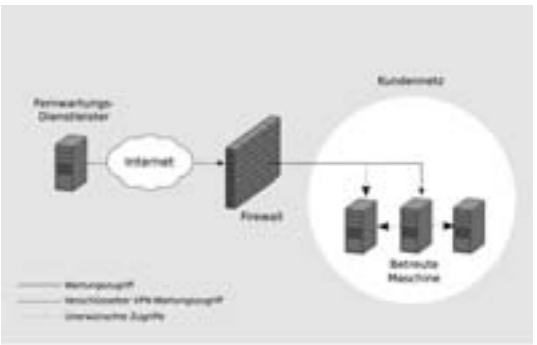


Bild 1.4.2

2 Realisierung einer sicheren Fernwartung

2.1 Allgemeine Konzeptbeschreibung

Das Konzept einer sicheren Fernwartung besteht aus dem logischen Separieren des zu wartenden Systems vom LAN. Bei einem Zugriff auf das System ist somit kein „Weiterspringen“ oder ein Sniffing auf andere bestehende Verbindungen möglich. Weiterhin lässt das vorliegende Konzept keine Verbindungen direkt vom Internet in das LAN, also auf das zu wartende System, zu. Man bringt durch das Zwischenschalten von sogenannten Rendezvous-Servern die internen Systeme in die Lage nach der Signalisierung eines Wartungswunsches die Verbindung auf das wartende System zu initialisieren. Durch dieses Konzept besteht die Möglichkeit eine verschlüsselte Verbindung über das Internet bereitzustellen, die Verbindungen aber auf den Rendezvous-Servern aber durch entschlüsseln, transparent und prüfbar zu machen (z.B auf Viren oder bekannte Angriffsverfahren). Weiterhin ist kein direkter Zugriff auf das System im LAN möglich, da dieses erst eine Verbindung zu Rendezvous-Server aufbauen muss, damit ein Fernwartungszugriff stattfinden kann. *Bild 2.1.1* zeigt eine Übersicht der Topologie mit integrierten Rendezvous-Servern.

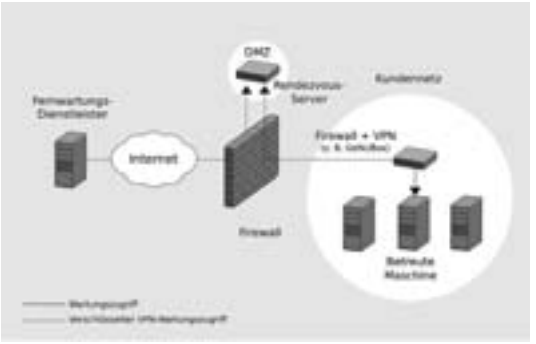


Bild 2.1.1

2.2 Technische Hintergründe

Der Rendezvous-Server wird, wie in Bild 2.2.1 zu sehen, in einer DMZ (demilitarisierte Zone) platziert. Das System ist somit physikalisch und logisch vom internen Netz getrennt. Eine weitere Firewall+VPN-Box wird im LAN platziert. An diesem System befindet sich auch ein Zugang zu dem zu wartenden System. Der externe Partner verbindet sich via SSH auf den Rendezvous-Server. Die Verbindung über das Internet ist somit verschlüsselt und nicht mitlesbar. Der Rendezvous-Server ist durch eine geeignet Authentifizierung und Authoriziering, vorzugsweise mittels Zertifikaten, in der Lage die Gegenstelle zu identifizieren. Der SSH-Tunnel endet auf dem Rendezvous-Server und wird hier auch entschlüsselt. Der Rendezvous-Server kann nun sehen, auf welches interne System eine Wartungsverbindung erstellt werden soll. Durch die Zuordnung des gesuchten Systems zu einer internen Firewall+VPN-Box wird hier ein Signal für einen Verbindungswünsch übermittelt. Dieses Signal ist ein administratives Protokoll, welches die vom Absender übermittelten Daten enthält. Dieses Protokoll wird über einen IPSEC-VPN verschlüsselt und ist somit auch nicht von Dritten im internen Netz mitlesbar. Wenn das zu wartende System das Signal für einen Verbindungswunsch erhält, überprüft es die Absenderdaten. Es besteht bis hierhin noch keine direkte Verbindung von dem Fernwarter auf das zu wartende System. Stimmen die gesendeten Parameter mit der internen Policy des Systems überein, baut das zu wartende System eine verschlüsselte Verbindung zu dem Rendezvous-Server auf. Gleichzeitig wird das System von den GenuBoxen durch die Zuweisung in ein eigens generiertes VLAN logisch aus dem internen Netz getrennt und in eine eigene Zone „geschoben“. Ist die geschehen, verbindet der Rendezvous-Server in der DMZ die externe Verbindung und die interne Verbindung miteinander. Der externe Techniker kann nun so verschlüsselt über das Internet und durch das LAN des jeweiligen Kunden auf ein zu wartendes System zugreifen. Die logische Umgebung gestaltet sich so, als ob sich das zu wartende System in einem eigenen Wartungsnetz befindet und so isoliert ist.

2.3 Vorteile des Zugriffsverfahrens

Diese Lösung eliminiert alle genannten Risiken und bietet folgende Vorteile:

- Der Einfluss des Fernwarters und alle damit verbundenen Gefährdungen beschränken sich auf den kleinstmöglichen Bereich um das Wartungsobjekt herum.
- Ein FernwartungsZugriff ist ohne Mitwirkung oder Zustimmung des KundennetzAdministrators unmöglich.
- Außer auf dem Wartungsobjekt selber können die Aktionen des Fernwarters auch auf der GeNUBox und dem RendezvousServer im Klartext protokolliert werden.
- Der externe VPN-Tunnel schützt den FernwartungsZugriff vor Abhören, Verändern oder Übernahme der Sitzung.
- Der interne VPN-Tunnel verhindert einen direkten Zugriff des Fernwarters auf Kundensysteme, die sich nicht im Bereich des Wartungsobjektes befinden.

- Die Filterfunktion der GeNUBox verhindert einen Zugriff des Fernwarters vom Wartungsobjekt aus auf Kundensysteme, die sich nicht im Bereich des Wartungsobjektes befinden.
- Diese Fernwartungslösung ist unabhängig vom Typ der bereits vorhandenen Firewall des Kunden.

Literaturverzeichnis

[Bilder] www.genua.de

Intelligentes Backup/Recovery für virtuelle Umgebungen

Mario Werner

Sales
Syncsort GmbH
Calor-Emag-Straße 3
40878 Ratingen
mwerner@syncsort.com

Abstract: In den letzten Jahren haben sich die Situation und die Anforderungen in modernen IT Organisationen grundlegend geändert. Zu den drei wichtigsten Aspekten, denen IT Manager heutzutage begegnen müssen, gehören: "Höhere Anforderungen an die Datenverfügbarkeit und Wiederherstellungszeiten nach einem Datenverlust bzw. Systemausfall", "Jährlich stetig steigende Datenvolumina" und "Zunehmender Einsatz von Servervirtualisierung". Ein innovatives und umfassendes Datenschutz-Modell (entwickelt von Syncsort) wird in diesem Beitrag erläutert, das diese drei größten Herausforderungen abdeckt.

1 Notwendigkeit für neue ganzheitliche Lösungen

In den letzten Jahren haben sich die Situation und die Anforderungen in modernen IT Organisationen grundlegend geändert. Zu den drei wichtigsten Aspekten, denen IT Manager heutzutage begegnen müssen, gehören:

- Höhere Anforderungen an die Datenverfügbarkeit und Wiederherstellungszeiten nach einem Datenverlust bzw. Systemausfall
- Jährlich stetig steigende Datenvolumina
- Zunehmender Einsatz von Servervirtualisierung

Die drei oben beschriebenen Aspekte haben nicht zuletzt auch eine massive Auswirkung auf die Bereiche Datensicherung und Wiederherstellung. Klassisch basierte Ansätze auf Datei-Ebene können diesen Anforderungen nicht mehr gerecht werden. Dabei spielt es keine Rolle, ob bei Anwendung eines solchen Verfahrens die Sicherungen auf Band, VTL oder Platte erfolgen und ob der Datentransport über LAN oder SAN geschieht. Insbesondere auch in virtualisierten Umgebungen sind derartige klassische Verfahren nicht sinnvoll praktikabel.

2 Oberstes Ziel: Optimierung der RPO & RTO

Aufgrund der großen und stetig wachsenden Datenvolumina gilt es, im Hinblick auf die Wiederherstellung von Daten und die situationsbedingte Ausfallzeit, zwei Ziele im Auge zu behalten: Recovery Time Objectives (RTO) und Recovery Point Objectives (RPO). RPO beschreibt den Zeitpunkt, auf den das System zurückgestellt werden kann. Im Idealfall würde dieser zeitlich direkt vor dem Zeitpunkt des Systemausfalls liegen. RTO beschreibt die Zeit vom Ausfall eines Systems bis zu dessen Wiederherstellung und der Wiederaufnahme des Produktivbetriebes; also die klassische Wiederherstellungszeit.

Die meisten Firmen achten allerdings kaum auf den RPO und nehmen in Kauf, dass die letzte Sicherung und damit der aktuellste wiederherstellbare Stand vom Vorabend ist. Wenn ein Server um 15 Uhr ausfällt, hat man einen ganzen Tag produktive Arbeit verloren. Mit steigendem Datenvolumen wird dies bei der klassischen Datensicherung auf Datei-Ebene immer problematischer. Andererseits ist eine zusätzliche Sicherung untertags bei den meisten Unternehmen völlig undenkbar, da dieses durch das hohe zu sichernde Datenvolumen und die entsprechende I/O Last der Server eine nicht akzeptable Belastung der vorhandenen Ressourcen bedeuten würde.

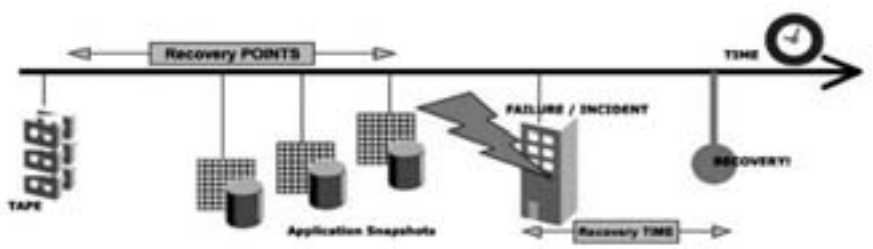


Abbildung 1: Beschreibung der Abbildung

3 Syncsort BEX Advanced Recovery Technologie optimiert RPO und RTO

Über die so genannte BEX Advanced Recovery Technologie bietet BEX eine Technologie zur Datensicherung, die auf dem Ansatz eines BLIB (Block-Level Incremental Backup) basiert und eine Datensicherung auf jedes beliebiges Disk System durchführt, das eine LUN an BEX herausgeben kann (eine spätere Auslagerung auf Tape – B2D2T - ist jederzeit möglich).

Bei der ersten Initialsicherung werden alle belegten Blöcke einer Platte gesichert. Bei jeder weiteren Datensicherung erkennt BEX redundante Daten und sichert als BLIB nur noch die seit der letzten Sicherung veränderten Blöcke einer Disk anstatt komplett geänderter Dateien und reduziert somit die transferierte und vorzuhaltende Datenmenge um 90% und mehr - analog einer Inline-DeDuplikation. Die Kosten für das Speichersystem für die Datensicherungen werden somit reduziert. Inkrementelle Sicherungen auf Block-Ebene ermöglichen täglich mehrfache Sicherungen und bieten zahlreiche Wiederherstellungszeitpunkte und bieten dadurch eine deutliche Verbesserung der Anforderungen hinsichtlich RPO.

Anhand einer ausgeklügelten Technologie erfolgt die Auswertung der zu sichernden Blöcke unter Umgehung der Dateisystem-Information. Im Ergebnis schützt dieses „Zero-Impact Backup“ die Daten ohne Beeinflussung der Performance laufender Applikationen. Datensicherungen erfolgen wesentlich schneller in einem Bruchteil der sonst üblichen Zeit und sind unabhängig von typischen Sicherungs-Zeitfenstern, Netzwerk-Engpässen und Ressourcen-Konflikten insbesondere bei virtuellen Maschinen.

Eine Vollsicherung ist nach der Erstsicherung nicht mehr notwendig. Durch intelligente Snapshot-Verfahren wird eine durchgeführte Sicherung auf dem sekundären Speichersystem immer mit der vorausgegangenen Sicherung verknüpft, so dass logisch betrachtet jede inkrementelle Block-Sicherung als Vollsicherung zur Verfügung steht.

Neben der Datensicherung von „normalen“ Dateien bietet Syncsort’s BEX Advanced Recovery Technologie auch entsprechenden konsistenten Datenschutz für Applikationen wie Exchange, SQL Server und Oracle.

Als wichtigste Unterschiede zur inkrementellen Sicherung auf Datei-Ebene ergeben sich eine erheblich schnellere Ausführungszeit und eine deutliche Entlastung der Ressourcen auf dem Server und dem Netzwerk während der Datensicherung.

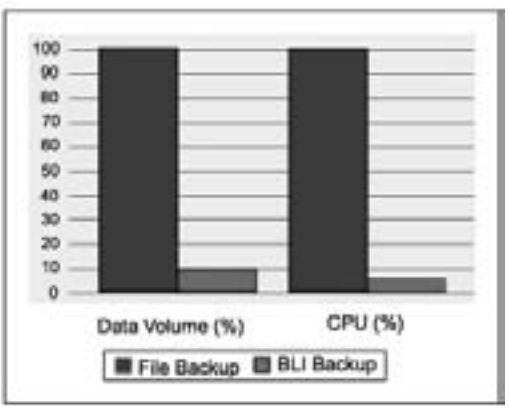


Abbildung 2: Vergleich von File- und Block-Level-Backup

Vor allem kann eine signifikante Reduzierung des vorzuhaltenden Sicherungsvolumens erreicht werden; insbesondere bei der Sicherung von Datenbanken oder Mailservern, da hier nur noch wenige Blöcke anstatt komplette Dateien gesichert werden müssen. Damit ist es für die Sicherung irrelevant, ob das gesamte Datenvolumen heute vielleicht 2 TB und in drei Jahren 20 TB beträgt; gesichert werden immer nur die Änderungen. Betrachtet man einen längeren Zeitraum von drei bis sechs Monaten, lässt sich bei Syncsort's BEX Advanced Recovery Technologie im Vergleich zu einer klassischen Sicherung auf Datei-Ebene eine Reduzierung der zu transportierenden Datenvolumina auf gerade einmal fünf bis zehn Prozent erreichen.

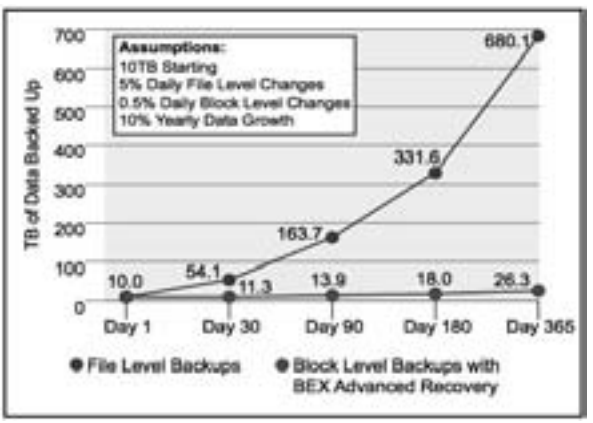


Abbildung 3: Vergleich von File- und Block-Level-Backup

Die sog. BEX Instant Availability Funktion geht im Wiederherstellungsprozess von Daten im Hinblick auf eine möglichst kurze Wiederherstellungszeit und einer damit verbundenen schnellen Wiederaufnahme des Betriebes noch einen Schritt weiter. Beliebige Datenmengen können umgehend wieder verfügbar gemacht werden für die Datenwiederherstellung, Tests, Entwicklung etc. und somit auch die strengsten RTO Anforderungen unterstützt werden.

Bei dem Ausfall eines gesamten Servers bietet Syncsort mit BEX Bare Metal Recovery ein Bare Metal Restore an. Mittels der boot-fähigen BEX Bare Metal Recovery CD wird der Server hochgefahren und gestartet. Am Ende des Startvorgangs werden ein paar Angaben zu dem herzustellenden Server und dem Advanced Recovery Server gemacht, der die wiederherzustellende Sicherung verwaltet. Anschließend wird der komplette Server inkl. Betriebssystem, Applikationen und Daten wiederhergestellt und kann nach einem abschließenden Neustart wieder in Betrieb genommen werden.

Syncsort's BEX Advanced Recovery Technologie ermöglicht es in einer einzigartigen Art und Weise den Kunden von dem exponentiellen Datenwachstum zu entkoppeln und somit bares Geld bei der Infrastruktur für die Datensicherung einzusparen. Verschiedene Abläufe wie Disaster Recovery, Dateisicherung, Applikationssicherung, Sicherung von Außenstellen, Auslagerung von Daten in andere Brandabschnitte, Sicherung von Notebooks, etc. lassen sich in einem einzigen Prozess zusammenfassen und damit deutlich vereinfachen.

4 Optimale Unterstützung virtueller Umgebungen: Virtuell ist nicht physisch

Servervirtualisierung hat sich für viele Rechenzentren als Segen erwiesen. Allerdings bringt sie auch weitere Herausforderungen mit sich, wie zum Beispiel die Sicherung einer virtuellen Maschine (VM). Zwar lassen sich sie meisten Anwendungen leicht virtualisieren, allerdings bedarf die Sicherung besonderer Beachtung, denn gerade Sicherungen auf Datei-Ebene benötigen die vollen Ressourcen, sowohl eine starke CPU als auch einen großen Speicher, der mit einem hohen Datendurchsatz zurechtkommt. Wird eine konventionelle Sicherung einer VM gestartet oder gar aller parallel laufenden VMs, kommt es zwangsläufig zu massiven Performance-Einbußen auf dem gesamten System und unter Umständen werden geforderte Antwortzeiten einzelner Applikationen nicht mehr gewährleistet.

VMware hat VMware Consolidated Backup (VCB) auf dem Markt gebracht, um Sicherungen von virtuellen Maschinen zu erleichtern und das I/O-Problem auf dem ESX Server zu beheben. Allerdings wurden VM-Sicherungen dadurch auch komplizierter, denn für diese brauchte man neben einer zweistufigen Sicherung für die Image-Backups auch eine zweistufige Wiederherstellung. So ist es nicht verwunderlich, dass nur wenige Anwender VCB verwenden.

5 BEX bietet optimale Unterstützung für Virtualisierung und Migrationsprojekte

Wie eingangs dargestellt, kann eine klassische Sicherung auf Datei-Ebene nicht der richtige Ansatz für eine sinnvolle Sicherung und Wiederherstellung in virtuellen Umgebungen sein. Und auch die von den Herstellern der Virtualisierungsprodukte gebotenen Ansätze versprechen hier nur beschränkte Mehrwerte.

Die BEX Advanced Recovery Technologie, die weiter oben vorgestellt wurde, bietet nicht nur eine in der Form einmalige Unterstützung für physikalische Umgebungen, sondern auch für virtuelle. Die Möglichkeit, Sicherungen mit nur minimaler Ressourcenbelastung durchzuführen, unter Umgehung einer CPU-intensiven Dateisystem-Auswertung und nur auf Basis der seit der letzten Sicherung modifizierten Blöcke ist ideal für virtuelle Umgebungen. Dieses gilt gleichermaßen für VMware, Hyper-V und andere Produkte - und genauso für physikalische Systeme.

BEX ermöglicht die schnelle Erstellung einer virtuellen Maschine von jedem Wiederherstellungszeitpunkt einer Datensicherung. Das Ursprungssystem kann hierbei ein physikalischer oder virtueller Server sein (P2V / V2V). BEX automatisiert die Erstellung der virtuellen Maschine zur Vereinfachung des erfolgreichen Ablaufs. Ein Backup Snapshot kann innerhalb von wenigen Minuten in eine komplett laufende virtuelle Maschine migriert werden. Damit bietet BEX eine einmalige, kosteneffektive Möglichkeit, Lasten, Systeme und Applikationen zwischen verschiedenen Hosts im Netzwerk zu migrieren inklusive Physical to Virtual (P2V), Virtual to Physical (V2P), Physical to Physical (P2P) und Virtual to Virtual (V2V).

SOA in der Cloud

Sabine Wieland

Institut für Telekommunikationsinformatik
Hochschule für Telekommunikation Leipzig
Gustav-Freytag-Str. 43-45
04277 Leipzig
wieland@hft-leipzig.de

Abstract: In der ICT-Branche kann man verschiedene interessante Entwicklungen beobachten. Auf der einen Seite wird SOA, die Service orientierte Architektur von Software Systemen vorangetrieben und auf der anderen Seite erfolgt eine Marktöffnung der Hardware Ressourcen für derartige Software Systeme durch das Cloud Computing Konzept. Hinter dem Begriff Cloud und SaaS, PaaS, IaaS,...stehen konkrete Technologien und Konzepte, die in der Wirtschaft aber auch in F/E-Teams an FHs und Unis diskutiert werden.

Im Beitrag wird auf technische und organisatorische Hindernisse eingegangen und Risiken aufgezeigt, die die ICT-Branche in den nächsten Jahren meistern muss, um den Anforderungen aus der Wirtschaft gerecht zu werden.

Der Beitrag diskutiert aber auch die Chancen, die die Verknüpfung der modernsten Technologien bieten.

1 Software Qualitätsanforderungen

Durch die Wirtschaftskrise haben sich die Bedingungen am Weltmarkt weiter verschärft und eine flexible Reaktion auf veränderte Situationen ist entscheidend für den Fortbestand der Firmen. Deshalb müssen die Geschäftsprozesse optimal von der IT und damit von den Software Systemen unterstützt werden. Das können die monolithischen Systeme nicht mehr leisten. Verschiedene Konzepte (CORBA, SOA, etc.) zielen darauf ab, Softwarekomponenten entsprechend der geschäftlichen Erfordernisse zur Laufzeit zu einem funktionierenden Softwaresystem zusammen zustellen. Abbildung 1 zeigt eindrucksvoll die Entwicklung der gestiegenen Qualitätsansprüche, die sich sowohl im Preis als auch im notwendigen Aufwand der Software Entwicklung widerspiegeln. Zusätzlich dazu muss Software den unterschiedlichen Interessen der Stakeholder Auftraggeber, Nutzer, Entwickler und Gesetzgeber gerecht werden. Diese Stakeholder Gruppen sind nicht homogen, da zum Beispiel jeder Nutzer subjektiv unterschiedliche Qualitätsansprüche an die von ihm genutzte Software stellt.

1.1 SOA Service oriented Architecture

SOA Service oriented Architecture muss eine Reihe von Qualitätsanforderungen erfüllen, wie hohe Flexibilität, Prozess Optimierung, kurze Markteinführung, hohe Innovationsrate, erhöhte Produktivität. Beispiele dafür sind on Demand Anlieferungen der Zulieferer in der Autoindustrie. Bauteile werden nicht mehr zwischengelagert sondern direkt nach der Anlieferung verbaut. Eine derartige Produktionsweise erfordert u.a. auch ein optimales Zusammenspiel der IT von Autohersteller und Zulieferer. Lose gekoppelte Services mit einfachen, standardisierten Schnittstellen sollen diese Anforderungen erfüllen. Für die Entwicklung derartiger Software Bausteine erhalten altbekannte SW-Entwicklungsregeln wie Geheimnis- und Lokalitätsprinzip eine neue Bedeutung. Die vier Kategorien Interaktion, Prozess, Funktion und Bestand [Hu] ermöglichen eine fachliche Trennung der Services. Abhängigkeiten sind nur top down zulässig, siehe Abbildung 2.

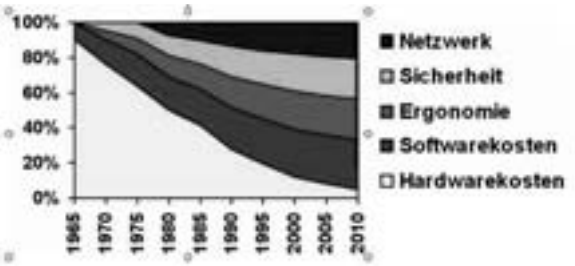


Abbildung 1: SW-Preisentwicklung im Vergleich zur Hardware

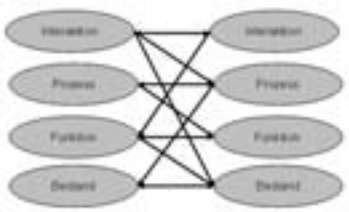


Abbildung 2: Service Abhängigkeiten

Daraus lassen sich weitere Forderungen ableiten, wie idempotent, kontextfrei, technikneutral, referenzfrei und grobgranular. Die Interoperability ist nur durch einfache, standardisierte Schnittstellen möglich, die durch eine lose Kopplung und enge Bindung der Services erreicht werden. Alle diese Anforderungen und Regeln stehen in Wechselwirkung zueinander. Durch Nichterfüllung einer Regel können alle anderen Anforderungen nicht ausreichend erfüllt werden und die Software Qualität leidet insgesamt. Diese Services sind in sich abgeschlossene fachliche Funktionalitäten die von verschiedenen Plattformen und Technologien bereitgestellt werden. [Jo]

Die Umstellung von monolithischen Software Systemen zur Service orientierten Architektur ist mit Aufwand verbunden und setzt bei den Entwicklern erhebliches Fachwissen voraus. Da die Services beliebig kombiniert werden können, wird der System Test entsprechend komplexer. Bereits vorhandene Software Komponenten, die die Anforderungen nicht erfüllen, können gekapselt weiter verwendet werden, wie das schon bei CORBA möglich ist. Das ist natürlich mit zusätzlichem Aufwand verbunden, spart aber zu Beginn der Umstellung auf SOA Zeit, das SOA- System ist schneller nutzbar. Trotzdem behindern solche „Legacy Systeme“ [He] den Technologiefortschritt und müssen zeitnah durch passende Services ersetzt werden.

Insgesamt steigt der Ressourcenverbrauch, es wird mehr Rechenleistung und mehr Bandbreite in den Netzwerken benötigt, da immer mehr Prozesse mit IT Systemen realisiert / unterstützt werden..

1.2 Cloud Computing

Deshalb müssen die Ressourcen (Hard- und Software) optimal ausgenutzt werden. Die Softwaresysteme beanspruchen die Ressourcen sehr unterschiedlich. Daraus resultiert, dass ein Teil der Computerleistung im idle Zustand ungenutzt bleibt, wobei andere Systeme die Ressourcen zu 100% beanspruchen. Um diese Schwankungen auszugleichen, wird die alt bekannte Technologie der Virtualisierung genutzt. Mehrere physikalische Hardware Systeme werden logisch zu einer Computer Cloud verbunden. Auf jedem physikalischen System können mehrere virtuelle Maschinen betrieben werden. Je nach Ressourcenbedarf können die virtuellen Maschinen konfiguriert werden. Eine virtuelle Maschine bleibt aber durch die physikalischen Ressourcen der jeweiligen Hardware beschränkt. Trotzdem entsteht durch das Cloud Computing ein Effektivitätsgewinn, da die idle Time des einen Prozesses durch einen anderen genutzt werden kann. Im Gegensatz zum Cluster sind die Hardware Systeme in der Cloud lose gekoppelt und heterogen. Hier eine Definition, die die wesentlichen Eigenschaften des Cloud Computing zusammenfasst:

„*Unter Ausnutzung virtualisierter Rechen- und Speicherressourcen und moderner Web-Technologien stellt Cloud Computing skalierbare, netzwerk-zentrierte, abstrahierte IT-Infrastrukturen, Plattformen und Anwendungen als on- demand Dienste zur Verfügung. Die Abrechnung dieser Dienste erfolgt nutzungsabhängig*.“ [Ba]

Der wesentliche Vorteil für den Nutzer besteht im Abrechnungsverfahren, er bezahlt nur die Dienste, die er genutzt hat. Der Begriff Computer Cloud umfasst demzufolge auch unterschiedliche Geschäftsmodelle, um Nutzern IT Dienste zur Verfügung zu stellen.

Die IaaS Infrastructure as a Service stellt Rechenleistung als Dienst zur Verfügung. Ein Nutzer mietet Hardwareressourcen und kann damit sein Wunschsystem betreiben. Damit liegen auch alle Risiken bezüglich des betriebenen Systems beim Nutzer. Er muss entsprechendes Fachpersonal beschäftigen, dass für Sicherheit, Verfügbarkeit und Skalierbarkeit sorgt. GRID Systeme nutzen im Gegensatz zur IaaS parallele Rechenleistung mehrerer administrativer Einheiten um komplexe Programme auszuführen.

Bei PaaS Plattform as a Service bietet zur Infrastruktur auch das Betriebssystem und die Laufzeit- sowie Entwicklungsumgebung an. Damit übernimmt der Provider auch die Verantwortung für Sicherheit, Verfügbarkeit und Skalierbarkeit. Für den Nutzer entsteht eine Abhängigkeit an den Betreiber und die Entwicklungsumgebung.

Standard Software wird als SaaS Software as a Service betrieben. Damit rückt die Entwicklungsumgebung für den Nutzer in den Hintergrund. Er erhält eine umfassende Dienstleistung, die er on- demand nutzen kann.

Backup- und Archivierungsdienste werden unter dem Begriff DaaS Data as a Service zusammengefasst. Zum Cloud Computing gehören auch LaaS Landscape as a Service, HuaaS Humans as a Service, HPCaas High Performance Computing as a Service, DICaaS Data Intensive Computing as a Service und in letzter Konsequenz XaaS oder EaaS Everything as a Service. Daraus ergeben sich verschiedene Vor- und Nachteile sowohl für den Betreiber, als auch für den Nutzer.

Als Vorteile für den Betreiber sind zu nennen: Auslastung der HW, Energieeinsparung, Automatisierung von Routineaufgaben, vereinfachte Wartung weil kein Wartungsfenster notwendig ist, höhere Sicherheit durch Sandbox, Migration auf neue Hardware einfach, Technologiewechsel ohne Betriebsunterbrechung möglich. Die Nachteile wirken dagegen unbedeutend: unflexible Lizenzierung, die komplexe Infrastruktur erfordert teure Managementwerkzeuge, die Virtualisierungsschicht verursacht zusätzliche Administration und ca. 10%Leistungsverlust, die Fehler sind schwer zu reproduzieren und das Backup wird unübersichtlich. Das Cloud Computing kann zu einem win2win System werden, denn auch der Nutzer profitiert von einer Reihe von Vorteilen: kein Administrationsaufwand, kaum Fachpersonal notwendig, Sicherheit und Verfügbarkeit liegt im Verantwortungsbereich des Betreibers und die Kosten sind überschaubar. Aber auch hier gibt es einige Nachteile: SLAs sind nicht ausgereift, es existieren noch keine Standards (ein Providerwechsel ist deshalb kaum möglich), hohes Vertrauen in den Betreiber in Bezug auf Sicherheit notwendig und der Nutzer wird abhängig von seinem Betreiber.

2 SOA in der Cloud

Was für ein Potential steckt nun in der Verbindung dieser beiden Technologien! Wenn nicht nur Standard Software as a Service in der Cloud angeboten wird, sondern geschäftsprozessabhängige Software Services wie Datenbanken, Ticketsysteme etc. frei kombiniert werden können. SOA wird in einigen Fachbeiträgen als Teil des Cloud Computing (SaaS) betrachtet. Um die Besonderheiten dieser beiden Technologien hervorzuheben und die Vorteile bei der Kombination aufzuzeigen, möchte ich diesem Beispiel nicht folgen. Cloud Computing hat zum Ziel in verschiedenen Geschäftsmodellen einem Nutzer IT Dienste bereitzustellen. SOA zielt darauf ab, IT Dienste nach den Geschäftsprozessen des Nutzers auszurichten. „*Web Services sind eine mögliche Art und Weise, die Infrastrukturaspekte von SOA zu implementieren*.“ [Jo]

2.1 Service Orchestrierung

„Die Orchestrierung beschreibt eine Form von Aggregation von Services zu höherwertigen Services und Geschäftsprozessen. Im Gegensatz zur Choreografie werden Services zu einem neuen Service komponiert, der dann die zentrale Steuerung des gesamten Ablaufs übernimmt.“ [Jo] Das heißt, die Kombination der Services durch den Geschäftsprozess bestimmt wird. Zusätzlich dazu müssen noch eine Reihe anderer Kriterien berücksichtigt werden, wie Performance, Sicherheit, Verfügbarkeit, Stabilität etc. In einer Cloud Umgebung stehen verschiedene Services zur Verfügung, die unterschiedlich gut den Geschäftsprozess unterstützen. Jeder Service muss dabei an den Anforderungen der vier Stakeholdergruppen Nutzer, Auftraggeber, Entwickler und Gesetzgeber gemessen werden. Da sich die Interessen dieser Stakeholder zum Teil ergänzen, aber auch zum Teil widersprechen müssen sie gewichtet werden. Je nach Geschäftsprozess wird die Bewertung durch die Stakeholder unterschiedlich bestimmt. Mit Hilfe von Linguistischen Variablen wird die Bewertung der Auswahlkriterien beschrieben. Auch die Qualitätsmerkmale der Services werden nach dieser Methode bestimmt. Danach kann der Service ermittelt werden, der die Anforderungen des Geschäftsprozesses am besten erfüllt. Dieses Mapping wird in Abbildung 6 dargestellt.

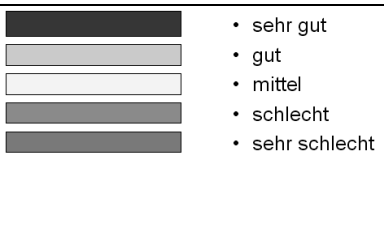


Abbildung 3: Farblegende



Abbildung 4: Abkürzungslegende



Abbildung 5: Linguistische Variable Performance



Abbildung 6: Verknüpfung der Kriterien

Die Verwendung der Fuzzy Logik ermöglicht eine Auswahl von Services unabhängig von subjektiven Entscheidungen und gewährleistet die angemessene Berücksichtigung der Interessen aller Stakeholdergruppen. Durch die grafische Darstellung der Verknüpfung bleibt die Orchestrierung auch in komplexen Systemen überschaubar und damit auch kontrollierbar.

In einem Verzeichnis ähnlich der UDDI können alle Services mit dieser Methode bewertet werden. Dabei ist es von Vorteil, wenn die Auswahlkriterien standardisiert sind. Werden Services mit falscher Bewertung im Verzeichnis beworben, werden sie bei der nächsten Service Auswahl nicht berücksichtigt.

Damit kann eine hohe Vertrauensbasis für die Verwendung von Open Source Services geschaffen werden und eine hohe Flexibilität in den IT Systemen erreicht werden.

Literaturverzeichnis

[Hu] Hess, A.; Humm, B.; Voß, M.: Regeln für serviceorientierte Architekturen hoher Qualität. Informatik Spectrum, Dezember 2006.

[He] Heilmann, H.; Sneed, H.; Wolf, E.: Softwaremigration in der Praxis, dpunkt Verlag, Heidelberg, 2010.

[Ba] Baun, C.; Kunze, M.; Nimis, J.,Tai, S: Cloud Computing: Web-Basierte dynamische IT-Services. Springer Verlag, Berlin, 2009.

[Jo] Josuttis, N.: SOA in der Praxis: System-Design für verteilte Geschäftsprozesse, Dpunkt Verlag; 2008.

Netzwerkvirtualisierung: Kosten senken im LAN & WAN

Pamela Krosta-Hartl, Roland Burlaga

LANCOM Systems GmbH
Adenauerstraße 20/B2
52146 Würselen
pamela.krosta-hartl@lancom.de
roland.burlaga@lancom.de

Abstract: Der vorliegende Beitrag führt in das Thema Netzwerkvirtualisierung ein, beleuchtet die unterschiedlichen Methoden in LAN, WLAN und WAN und zeigt anhand konkreter Praxisbeispiele, inwieweit die konsequente Virtualisierung von Netzen – wir sprechen von „Total Network Virtualization" – Unternehmen helfen kann, sowohl bei den Investitionen als auch bei Betriebskosten massiv Kosten zu reduzieren.

1 Netzwerkvirtualisierung und Einsparpotentiale

Grundsätzlich bezeichnet der Begriff Virtualisierung die Trennung einer IT-Anwendung von der verwendeten Hardware. Bei der Netzwerkvirtualisierung werden die Netzwerkdienste von der Netzwerkinfrastruktur abgelöst. Eine solche Virtualisierung kann im LAN über entsprechend intelligente Managed Switches realisiert werden, auf denen so genannte VLANs – Virtual Local Area Networks – konfiguriert werden. Ebenso stehen im Wireless LAN Virtualisierungsmöglichkeiten zur Verfügung, bekannt unter den Begriffen Multi-Service-WLANs oder Multi-SSIDs. Wird der Gedanke der Netzwerkvirtualisierung konsequent fortgesetzt, kann bei Einsatz geeigneter Hardware eine Virtualisierung bis ins WAN erfolgen – also auch des Routers und der dazugehörigen Internet-Anbindung. Das von LANCOM vor einigen Jahren eingeführte Advanced Routing & Forwarding (ARF) ermöglicht eine solche Router-Virtualisierung. Damit werden aus einem Router bis zu 64 logische – virtuelle – „Maschinen".

Die Gründe für den Griff zur Virtualisierung sind vielfältig. Während bei den Servern die zentrale Wartung eine große Rolle spielt, eröffnen virtualisierte Netzwerke völlig neue Anwendungen, die „normale" Netzwerke nicht bieten können. Beiden Bereichen ist jedoch eines gemein: durch die Virtualisierung sparen Unternehmen teure Ressourcen, was sich sowohl auf Investitionen als auch auf die laufenden Kosten nachhaltig positiv auswirkt.

Doch wodurch werden diese Einsparungen möglich? Der wirtschaftliche Hauptnutzen der Netzwerkvirtualisierung liegt in der Möglichkeit der Mehrfachnutzung der Netzwerkinfrastruktur. Konkret heißt das, dass auf Basis eines physikalischen Netzes mehrere logische Netze geschaffen werden, die zwar dieselbe Geräteinfrastruktur nutzen, abgesehen davon aber vollständig und sicher voneinander getrennt und füreinander unsichtbar sind. Diese Mehrfachnutzung kann sich einerseits auf Dienste beziehen – ein häufig zitiertes Beispiel ist der WLAN-Gastzugang, den Unternehmen ihren Besuchern gerne gewähren –, andererseits aber auch auf die Nutzung ein und desselben Netzwerks durch mehrere Firmen, zum Beispiel in einem Gründerzentrum.

Dank Virtualisierung reicht also ein physikalisches Netz aus Switches, Kabeln, Access Points und Routern aus, um mehrere logische Netze zu betreiben. Das Einsparpotential hinsichtlich der benötigten Geräte ist enorm, ebenso die Senkung der laufenden Kosten beispielsweise beim Energiebedarf.

2 Methoden der Virtualisierung

Technisch gesehen kommen bei der Netzwerkvirtualisierung zwei unterschiedliche Methoden zum Einsatz.

2.1 Statische Virtualisierung auf Layer 2

Bei VLANs und Multi-SSIDs bezieht sich die Virtualisierung auf das Übertragungsmedium, das zu einem Shared Medium umfunktioniert wird. Ein Access Point mit mehreren SSIDs spannt beispielsweise einfach mehrere voneinander getrennte Funkzellen „nebeneinander“ auf. VLANs und Multi-SSIDs werden also auf der physikalischen Netzwerkebene, dem Layer 2 des OSI-Modells, realisiert. Diese Art der Virtualisierung ist auf das verkabelte oder drahtlose Unternehmensnetz, das LAN, begrenzt.

Fakt ist jedoch, dass die IP-basierte Zusammenarbeit zunehmend über die Grenzen einer Organisation und damit des LAN hinausgeht, sie verlagert sich immer mehr ins WAN. Zudem orientiert sie sich immer stärker an den Aufgaben der Mitarbeiter oder Kommunikationsteilnehmern. Das einfachste Beispiel für das Überschreiten der Grenzen des LAN ist der Netzwerkzugang für Gäste in den eigenen Räumen, in komplexen Szenarien erhalten externe Dienstleister über das Internet Zugriff auf bestimmte Anwendungen im lokalen Netzwerk. Eine rein statische Virtualisierung wie sie bei VLANs und Multi-SSIDs auf Layer 2 geschieht, reicht hier nicht mehr aus.

2.2 Dynamische Virtualisierung auf Layer 3

Der nächste Schritt in der Virtualisierung von Netzwerken ist daher die dynamische Virtualisierung auf Layer 3, also die Trennung der Anwendung selbst von den physikalischen Übertragungsmedien: die IP-Netzwerke und das Routing der Datenpakete zwischen diesen IP-Netzwerken. Ähnlich wie bei der Virtualisierung von Servern wird dabei eine Hardware – ein Router – genutzt, um mehrere virtuelle Router einzurichten. Jeder dieser virtuellen Router kann speziell für sein Netzwerk konfiguriert werden.

Mit einer solchen höheren Ebene der Virtualisierung können auf vorhandenen Infrastrukturen parallel völlig unterschiedliche Anwendungen mit dedizierten Einstellungen für das Routing und die Zugriffsberechtigungen realisiert werden. Ein Mechanismus, mit dessen Hilfe solch komplexe Virtualisierungsszenarien realisiert werden können, ist das eingangs erwähnte Advanced Routing & Forwarding.

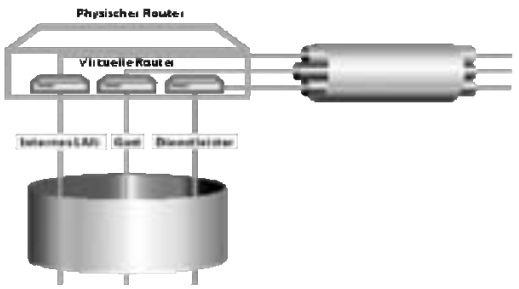


Abbildung 1: Router-Virtualisierung mit ARF

Der Kernpunkt dieser Technologie ist die Möglichkeit, für jede Anwendung ein eigenes IP-Netzwerk auf dem zentralen physikalischen Router einzurichten. Für jedes dieser Netzwerke können grundlegende Funktionen wie die Firewall oder der DHCP-Server separat konfiguriert werden. Besonders wichtig ist jedoch die Möglichkeit, über ein spezielles Tag die Einträge in der Routing-Tabelle einem IP-Netzwerk zuzuordnen: so entstehen in einem physikalischen Router mehrere virtuelle Router, deren Verhalten speziell auf das jeweilige IP-Netzwerk abgestimmt wird. Dabei werden die Tags zur Unterscheidung der Datenpakete anhand verschiedener Kriterien zugewiesen.

Bei der Einrichtung eines WLAN-Zugangs für die Besucher eines Unternehmens können die Gäste nach der Authentifizierung automatisch mit beschränkten Rechten, z. B. nur mit Internetzugang und Zugriff auf einen Netzwerkdrucker ausgestattet werden. Neben den internen Teilnehmern können aber auch externe Unternehmen in die virtuelle Netzwerkstruktur aufgenommen werden. Wird z. B. einem Dienstleister ein VPN-Zugang zur Überwachung der Heizungsanlage eingerichtet, kann dieser Zugang gezielt einem anderen IP-Netz zugeordnet werden.

Um diese Form der Virtualisierung zu erreichen, müssen die verwendeten Router ARF beherrschen und die Zuordnung der IP-Netzwerke zu einer LAN-oder WLAN-Schnittstelle unterstützen, gegebenenfalls genauer spezifiziert über eine VLAN-ID. Um die Tags zur Unterscheidung der IP-Netzwerke richtig zu behandeln, müssen alle Switches und Access Points im LAN VLAN- beziehungsweise Multi-SSID-fähig sein.

Sollen überlappende IP-Netze dann getrennt über eine WAN-Verbindung weiter übertragen werden, kommt ein zusätzliches Tunnel-Protokoll zum Einsatz: das PPTP (Point-to-Point Tunneling Protocol). Dadurch bleibt die LAN-seitige Trennung durch VLAN oder Multi-SSID auch bei der Übertragung über das Internet komplett erhalten. Für maximale Sicherheit sorgt dabei die verschlüsselte Übertragung durch einen IPSec-VPN-Tunnel.

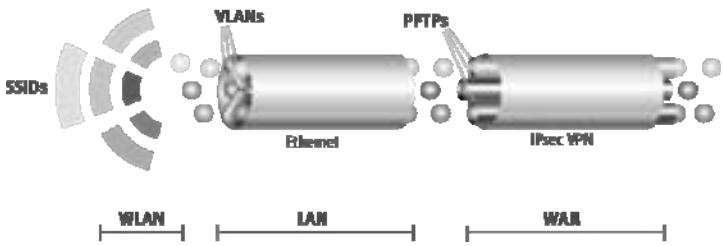


Abbildung 2: Multi-VPN: Tunnel im Tunnel

3 Praxisbeispiele

3.1 Bürogemeinschaft

Virtualisierte Netzwerkstrukturen bieten schon für kleine Unternehmen deutliche Vorteile. Auch Arztpraxen, Steuerkanzleien oder Ingenieur-Büros können heute nicht mehr auf die Vernetzung mit Geschäftspartnern verzichten. In vielen Gebäuden reicht die vorhandene Verkabelung aber nicht aus, um für jeden Mieter ein komplett eigenes Netzwerk einzurichten.

In diesem Fall kann für z. B. für die Arztpraxis und das Ingenieurbüro auf Basis desselben physikalischen Netzwerks jeweils ein separates virtuelles IP-Netzwerk eingerichtet werden. Beide Netzwerke sind intern völlig voneinander getrennt, so dass kein unbefugter Zugriff auf Patientendaten oder Konstruktionspläne möglich ist. Das Ingenieur-Büro kann zusätzlich noch einen WLAN-Zugang für Gäste einrichten, die nur Zugriff auf das Internet haben.

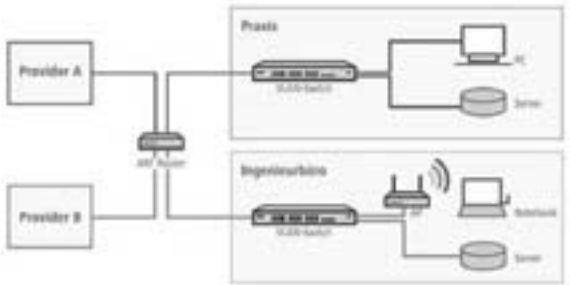


Abbildung 3: Gemeinschaftliche Nutzung durch mehrere Teilnehmer

Durch die gemeinschaftliche Nutzung sparen unsere Arztpraxis und das Ingenieurbüro gleich mehrfach: 30 - 40 % der Investition in Hardware (Access Points, Switches und Router – je nach Räumlichkeiten); 50 % bei der Internet-Anbindung (Leitungskosten); 30 - 50 % beim Energiebedarf; Einsparungen bei Verkabelung und Verlegung je nach Räumlichkeiten.

3.2 Supermarkt

In dem ersten Beispiel geht es vor allem um die Trennung von internen Datenströmen. Ein entscheidender Vorteil von Netzwerkvirtualisierung ist jedoch, auch Anwendungen mit externen Teilnehmern sauber in das eigene Netzwerk zu integrieren – ohne in zusätzliche, separate Infrastruktur investieren zu müssen.

Der Blick hinter die Kulissen einer modernen Supermarktfiliale zeigt, welche umfangreichen Sparpotentiale „Total Network Virtualization“ mittels VLAN, ARF und PPTP eröffnet.

Die Supermarktfiliale ist mit der Zentrale der Handelskette per VPN vernetzt, die PCs in der Filiale sind direkt in das ERP-System der Handelskette eingebunden, um den Warenfluss optimal zu steuern. Eine weitere VPN-Verbindung zwischen der Filiale und einem Geldinstitut sichert den elektronischen Zahlungsverkehr ab, wenn Kunden sicher und bequem per EC-Karte und PIN bezahlen möchten. Innerhalb des Supermarkts wird die gesamte Inventur mittels WLAN-fähiger Handscanner durchgeführt, die die Ergebnisse automatisch und in Echtzeit in das ERP System übermitteln und Nachbestellungen auslösen.

Zusätzlich zu diesen unternehmensinternen Verbindungen müssen zahlreiche externe Dienstleister in das Netzwerk des Supermarkts gelangen, denn viele Filialisten haben bedeutende Teile der Arbeiten in ihren Filialen bereits ausgelagert: so werden beispielsweise Kassensysteme, Kühltruhen und Getränkerückgabeautomaten schon heute vielerorts per IP-Verbindung durch externe Dienstleister in Echtzeit überwacht und gewartet. Die Inhalte auf den digitalen Werbedisplays werden von einer externen Werbeagentur direkt über eine Internet-Verbindung eingespeist, die Heizungssteuerung bezieht die Klima-Prognose von einem Wetter-Server, die VoIP-Telefone nutzen eine externe SIP-Telefon-Anlage.

Während es beim ersten Beispiel um die reine Trennung von Datenströmen ging, ist die Gemengelage im Supermarkt ungleich komplexer. Neben der Trennung unterschiedlicher Datenströme muss sichergestellt sein, dass die zahlreichen externen Dienstleister jeweils nur auf die Anwendung und Ressource zugreifen können, für die sie verantwortlich sind.

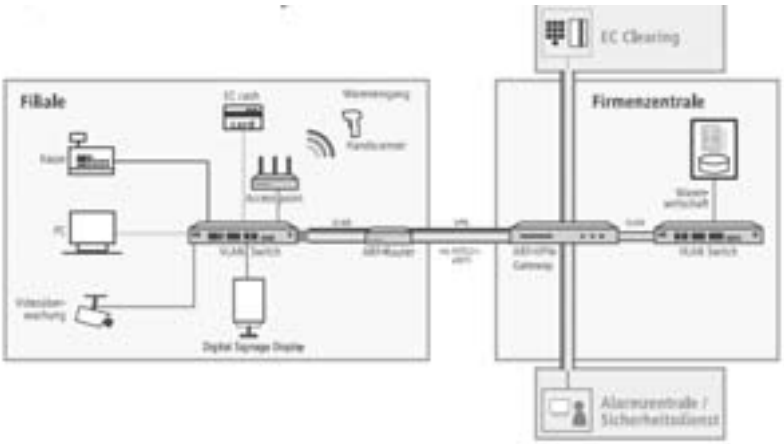


Abbildung 4: Virtualisierung im Supermarkt

Nur über die konsequente, dynamische Virtualisierung des gesamten Netzes inklusive der WAN-Verbindungen der Filiale kann dies wirtschaftlich realisiert werden. Andernfalls müsste aus Sicherheitsgründen eine Vielzahl von Netzen aufgebaut und eine Vielzahl an Internet-Anbindungen betrieben werden. Die Kosten für Geräte, Wartung, Energieverbrauch und Leitungen wären auf Dauer nicht tragbar.

Im Rahmen der Virtualisierung des Supermarkt-Netzes wird für jede Anwendung bzw. jeden Dienstleister ein eigenes virtuelles IP-Netzwerk eingerichtet, für das ein spezieller IP-Adresskreis und separate Routing-Einstellungen definiert werden. Der Netzabschnitt für die Kassenabrechnung kann so z. B. an die IP-Adressen angepasst werden, die der Betreiber in seiner VPN-Struktur verwendet. Im hausinternen LAN werden die IP-Netze zusätzlich über entsprechende VLAN-Tags markiert, die über einen VLAN-fähigen Switch getrennt werden. Alle anderen Teilnehmer können nicht auf dieses Netzwerk zugreifen.

Die Einsparpotentiale sind enorm, lassen sich jedoch aufgrund der sehr unterschiedlichen Szenarien nur für Einzelfälle genauer beziffern: Auslagerung arbeitsintensiver Dienstleistungen an günstige, externe Anbieter; Minimierung der Hardware-Ausstattung (Access Points, Switches und Router); Wegfall von Leitungskosten durch Mehrfachnutzung der Internet-Verbindung z. B. für EC-Zahlung, Warenflusslogistik und die Anbindung externer Dienstleister; Minimierung des Verkabelungsaufwands durch konsequente Kommunikation über IP-Netze (ein Netz für alles); massive Einsparungen beim Energiebedarf.

4 Fazit

Die hier skizzierten Beispiele lassen sich auf viele andere Bereiche übertragen. Unternehmen und Institutionen profitieren massiv von den Möglichkeiten der konsequenten Netzwerkvirtualisierung. Einerseits durch massive Einsparungen bei Hardware, Installation, Betrieb und Nebenkosten, andererseits aber auch durch ganz neue Interaktionsmöglichkeiten z. B. mit externen Partnern wie Kunden und Lieferanten und durch ein deutlich höheres Maß an Sicherheit, das sich durch die sehr flexible Lenkung von Datenströmen und Zugriffsberechtigungen erzielen lässt.

Voraussetzungen hierfür ist der konsequente Einsatz weithin bekannter Virtualisierungskonzepte wie VPN, VLAN und Multi-SSID in Kombination mit der Virtualisierung der IP-Netze und Router mittels ARF und der Weiterführung der getrennten Übertragung übers WAN mittels PPTP.

Moderne Kommunikationskomponenten integrieren diese Funktionalitäten standardmäßig, so dass für eine kostensparende Virtualisierung keine zusätzliche Software-Lösung erforderlich ist.

Workshop

Applications of Semantic Technologies

-

AST2010

Andreas Abecker, Stephan Bloehdorn, Stephan Grimm

Applications of Semantic Technologies

AST2010

5th International Workshop

Organisers: Andreas Abecker[1], Stephan Bloehdorn[2], Stephan Grimm[1]

[1]FZI Forschungszentrum Informatik, 76131 Karlsruhe, abecker@fzi.de / grimm@fzi.de
[2] Karlsruhe Service Research Institute (KSRI),
Karlsruhe Institute of Technology (KIT), 76131 Karlsruhe stephan.bloehdorn@kit.edu

Organised in cooperation with
Semantic Technologies Institute Germany, STI Germany.

http://km.aifb.kit.edu/ws/ast2010

Semantic Web is a major international research effort with the goal to make web content available for intelligent knowledge processing. It draws on standard and novel techniques from various disciplines within Computer Science, including Knowledge Management, Artificial Intelligence, Databases, Internet Technology, Software Agents, eCommerce, etc. The methods and tools developed and integrated for this purpose – often called Semantic Technologies – are generic and have a very large application potential outside the domain of the Semantic Web. Such applications are currently being investigated in various disciplines within Computer Science, including Ambient Intelligence, Software Engineering, Cognitive Systems, Corporate Intranets, Knowledge Management, Bioinformatics, etc. We believe that Semantic Technologies provide methods and tools that will persist in these and other application areas for the foreseeable future.

This workshop brings together researchers and practitioners who work on applications of Semantic Technologies. It furthers the cross-fertilization between application areas and aids the technology transfer from foundational research into practice. The AST workshop series covers diverse application areas of Semantic Technologies, including:

- Ambient Intelligence
- Cognitive Systems
- Data Integration
- Software Engineering
- Multimedia Data Management
- Service-Oriented Computing
- Machine Learning
- eScience
- Information Extraction
- Grid Computing
- Peer-to-Peer Systems
- eCommerce
- eGovernment
- Bioinformatics

For AST2010, we selected eight submitted papers for presentation. Four of them address topics in the area of the INFORMATIK 2010 guiding theme “Service Science” , namely Web service composition, Web service discovery, matching of business and electronic services as well as business-oriented CAx integration. The other four papers cover various aspects of Semantic Web applications and technologies, namely using semantic Wiki and semantic metadata for IT Service Management and e-Learning, respectively, as well as ontology engineering and ontology versioning as important base technologies. The workshop program will be complemented by two speakers from industry: Christian Seitz (Siemens AG, Munich) who elaborates on semantic technologies in indusry, energy and healthcare, and Jens Lemcke (SAP AG, Karlsruhe) who discusses the SAP view on Semantic Business Process Management.

This year’s program committee included *Sören Auer (University of Leipzig, Germany), Franz Baader (TU Dresden, Germany), Bernhard Bauer (University of Augsburg, Germany), Christoph Beierle (University of Hagen, Germany), François Bry (University of Munich, Germany), Johannes Busse (ontoprise GmbH, Germany), Andreas Friesen (SAP AG, Germany), Thomas Fuhr (Georg Simon Ohm University of Applied Sciences Nuremberg, Germany), Jens Hartmann (EnBW AG, Germany), Mark Hefke (CAS Software AG, Germany), Wolfgang Hesse (University of Marburg, Germany), Gabriele Kern-Isberner (University of Dortmund, Germany), Harald Kosch (University of Passau, Germany), Daniel Krause (L3S Hannover, Germany), Thorsten Liebig (University of Ulm, Germany), Meenakshi Nagarajan (Wright State University Dayton Ohio, USA), Daniel Oberle (SAP AG, Germany), Simon Schenk (University of Koblenz, Germany), Stefan Schlobach (VU Amsterdam, The Netherlands, Germany), Michael Sintek (DFKI Kaiserslautern, Germany), Robert Tolksdorf (FU Berlin, Germany), Valentin Zacharias (FZI Karlsruhe, Germany),* and *Jürgen Ziegler (University of Duisburg-Essen, Germany)*. We thank all program committee members for their help in reviewing the contributions.

Karlsruhe, July 2010

Andreas Abecker

Stephan Bloehdorn

Stephan Grimm

Semantic MediaWiki as an Integration Platform for IT Service Management

Frank Kleiner, Andreas Abecker

FZI Forschungszentrum Informatik
an der Universität Karlsruhe (TH)
{kleiner, abecker}@fzi.de

Abstract: We describe our work on using Semantic MediaWiki as a central platform for managing IT services and the underlying technical components within an organization. We present our toolkit of MediaWiki extensions which add ITSM-specific functionalities to Semantic MediaWiki and describe their interactions. We also outline the benefits gained by using Semantic Technologies in IT Service Management.

1 Introduction

IT Service Management (ITSM) deals with providing reliable IT services. The perspective of ITSM is customer-centric which distinguishes it from more technology-oriented IT management approaches of the past. This customer focus helps to ensure the alignment of services provided by the IT department with the business goals of an organization [Add07]. The *IT Infrastructure Library (ITIL)* is the most widely used ITSM framework. ITIL consists of five volumes which form a lifecycle encompassing all aspects of an IT service. Our work focusses on the aspects of *Configuration Management* and *Change Management* [LM07] as well as *Problem and Event Management* [CW07]. Configuration Management deals with providing a system in which all relevant items for providing services are stored, together with their interactions with each other. Items represented in the *Configuration Management System (CMS)* are referred to as *Configuration Items (CIs)*. *Change Management* focusses on providing processes and procedures for planning changes in order to minimize the associated risks of failed changes or unwanted side effects. *Problem Management* deals with finding and fixing the causes of malfunctions in components which impact the working of services. *Event Management* is concerned with monitoring the working of IT services and the underlying technical components as well as reporting of failures or potential problems.

Wikis [EGHW07] are Web-based platforms which enable users to edit articles from within their Web browsers. This makes Wikis a perfect platform for the collaborative generation of knowledge, as exemplified in Wikipedia[1], the world's largest encyclopedia. Within organizations, Wikis are more and more often used for collaborative knowledge manage-

[1] http://www.wikipedia.org

ment. **Semantic Wikis** [SBBK09] extend Wikis by semantic features which enable them to use ontologies [SS09] as the underlying data model; this makes available additional mechanisms for representing, organizing and retrieving information. Semantic MediaWiki (SMW) is an extension for the popular MediaWiki [Bar08] platform. It allows to give meaning to links between articles by enabling users to explicitly describe their semantics. Attributes can be used to state fact knowledge in Wiki articles which can be processed within SMW or by external reasoners. Class hierarchies can be built by using MediaWiki categories. [KVV+07] describes SMW in detail and gives examples.

Problem description and approach: Professional ITSM requires processes for the management of configuration items, for the structured application of changes, as well as for diagnosing problems and for monitoring services and hardware components for events that indicate malfunctions. Software tools help system administration personnel to run these processes. While very small IT environments usually are fine with a few text or spreadsheet documents describing the functions of components and common problems, large IT environments use specialized commercial software which in most cases is expensive and whose maintenance is labour-intensive. This paper gives an overview of our Semantic ITSM Wiki which was introduced in [KA09]. It addresses the following challenges: (1) provide an extensible platform, built on top of freely availably software, for managing medium-sized IT environments; (2) enable computer-savvy non-administrative users to collaboratively contribute to the ITSM Wiki; (3) provide extensions for automatically adding information from managed components, integrating a systems monitoring tool and a network intrusion detection system into the ITSM Wiki, as well as providing a mechanism for supporting administrative personnel in tracking down hardware and software problems.

The **structure of the paper** is as follows: Section 2 gives an overview of our extensions which extend the Semantic MediaWiki platform by ITSM-specific components. Section 3 surveys some related work. Section 4 concludes and sketches some future work.

2 ITSM Extensions for Semantic MediaWiki

The goal of the authors' ongoing work is to extend the Semantic MediaWiki platform by extensions for the use in the ITSM area. Semantic Wikis provide a platform which combines the ability to store structured information in the form of relations and attributes with the ability to store free text. In the context of ITSM, this enables the storage of, for example, dependencies between services and computers in the form of relations, together with documentation in the form of free text. The following subsections present four extensions which address four of the most commonly implemented ITSM processes. The most implemented processes are Incident Management including the Service Desk function, as well as Configuration Management, Problem Management and Change Management (cp. [BT05]). The authors have decided not to implement the Service Desk function at the moment because of the use of the OTRS tool in their environment[2].

[2] http://www.otrs.org

Population with Static and Dynamic Information: In order to have available always up-to-date information about the status of IT components (e.g., computer configurations, including the installed hardware and software), it is imperative to have a mechanism for automatically gathering the needed information, because manually gathering information is both time consuming and error-prone. The component described in more detail in [KAL09] implements a mechanism for remotely collecting information over the network from Windows computers via the Windows Management Instrumentation (WMI) mechanism [Jon07]. This enables ITSM Wiki users to always access the current version of configurations, including information about hardware and software (e.g., installed applications). Furthermore, a component for reading and writing configuration information via the Simple Network Management Protocol (SNMP) [Sch05] was implemented which extends the range of devices from which information can be gathered to network-enabled hardware components, e.g., printers and network switches.

Generating Systems Monitoring Information: Monitoring IT components for their correct behavior is important for providing a high level of service quality. If a service stops working, correct measures must be initiated, ranging from automatically restarting the troubled service to alarming the administrator who is responsible for the correct working of the service. In ITIL, systems monitoring is part of the Event Management process which is described in the Service Operation volume [CW07]. The Systems Monitoring component, previously described in more detail in [KAB09b, KAB09a], builds on top of the Nagios [Bar05] systems monitoring tool which can be configured to monitor networks of almost all sizes and complexities. The Systems Monitoring extension integrates Nagios into the ITSM Wiki by allowing to register Wiki articles which represent Configuration Items (i.e., computers, or networking equipment) to be monitored by the external monitoring tool. Semantic relations which are used to represent relations between CIs, are converted into service dependencies within Nagios. This frees the administrative staff from separately maintaining a systems monitoring configuration and enables them to access systems monitoring information from a unified user interface.

Integration of Intrusion Detection Information: Monitoring systems for security incidents is accomplished by implementing an intrusion detection system infrastructure. Depending on their location, intrusion detection systems can be differentiated between network intrusion detection systems and host-based intrusion detection systems. Our work builds on the signature-based network intrusion detection system Snort [Roe99]. Signature-based network intrusion detection systems capture network traffic and use signatures for checking if it contains malicious data (cp. [Koh06]). In [Axe00], it is shown that intrusion detection systems have an inherently high rate of false positives because of the high amount of non-malicious traffic in contrast to the amount of malicious traffic. The integration of the Semantic ITSM Wiki with an external intrusion detection system helps to minimize the number of false positives by taking into account semantic facts stored in the Wiki. By using the ITSM Wiki as the user interface for displaying potential intrusions, semantic features are made available (e.g., the dynamic creation of a customized filter for finding certain network traffic). A filter between the Snort database and the Wiki only imports information which is classified as potentially malicious to the target system. An example for a potential attack which is recorded in the Snort database due to the lack of

background knowledge, but not imported into the ITSM Wiki, is the following: an attack is detected which tries to exploit a security hole in the Microsoft Internet Information Server (IIS), but on a server running Apache under the Linux operating system. The attack is logged into the the Snort database. While it would be presented to the administrator when using a standard program (e.g., BASE[3]) for displaying events, the Intrusion Detection extension acts more intelligently. It checks whether the system against which the attack is targeted, is indeed susceptible to the attack. Because the facts stored in the Wiki show that the system is running Linux, it is deduced that it cannot be running IIS and thus cannot be attacked by an exploit targeted at IIS. By reducing the number of false positives, administrative personnel can concentrate on real attacks, which leads to an improvement in overall systems security.

Semantic Problem Finder: Problems in ITIL are defined "as the unknown cause of one or more incidents" [CW07]. In order to deliver reliable IT services, efficient processes and tools for tracking down problems have to be present. The Semantic Problem Finder, which is implemented as a MediaWiki extension, provides tool support for administrators to help locate the underlying cause of problems. By using the SMW-based semantic Configuration Management Database as a foundation, possible common causes for a set of given problems are determined and given to the administrators for review. The process for finding the cause of a problem is as follows: (1) a problem is reported to the administrative staff; (2) an administrator enters the affected configuration items into the Semantic Problem Finder extension; (3) a list of possible causes is determined by building trees which represent connections and dependencies between CIs. The following (trivial) example helps in clarifying the process: (1) a problem is found which prevents multiple computers from connecting to the network; (2) the names of some or all of the affected computers are entered into the Problem Finder; (3) it can be seen that all computers are connected to the same network switch which with high probability is the cause of the problem.

3 Related Work

There exists a number of tools supporting administrative personnel in providing IT services. Specialized tools for managing Configuration Management Databases are Peregrine, i-doit, and OTRS::ITSM. Being built for managing structured data, the flexibility of these tools does not reach the flexibility provided by a Semantic Wiki as the technical foundation. Tools for automatically gathering information from computers over the network are available from commercial vendors and as Open Source. The software Spiceworks[4] is an example for a specialized tool for asset and configuration management which automatically reads configuration data over the network. By transferring automatically gathered information into semantic relations and attributes, the extension created by the authors makes possible features which exceed the functional range of classic tools, e.g., much more flexible, complex queries. Tools for managing systems monitoring configurations for Nagios are Lilac Configurator, Monarch and NCPL. A tool for displaying and

[3] http://base.secureideas.net/

[4] http://www.spiceworks.com/

filtering intrusion detection data is BASE, which is a Web-based front-end for accessing Snort databases. In [HO09], the benefits of using a Semantic Wiki for managing metadata are quantified, which is related to the approach presented in this paper, which uses a Semantic Wiki for managing metadata about IT infrastructures. In [Lan10b, Lan10a], a Semantic Wiki-based helpdesk and data center inventory system is described, which in some aspects can be compared to the work presented in this paper.

4 Summary and Future Work

We summarized our ongoing work on building an integrated platform for the management of IT services and infrastructures on top of a Semantic Wiki. In order to enable administrative staff to gain productivity from the use of the Semantic Wiki, four specialized tools were presented which are implemented as MediaWiki extensions. The first tool collects data from computers and other components over the network and integrates the information in the form of semantic facts into the Wiki. The second tool integrates an external systems monitoring tool, while the third one builds an interface for importing information from an external intrusion detection system. Finally, the fourth one helps in tracking down problems by looking for common causes in tree data structures generated from semantic facts. While parts of the system described in this paper are running in a productive environment consisting of 500 computers, a formal evaluation of the benefits gained from the use of a Semantic Wiki as a hub for all information relating to IT Service Management within an SME company, will be done in the future. Furthermore, a component for integrating an external Service Desk tool (e.g., OTRS) is planned.

References

[Add07] Rob Addy. *Effective IT Service Management: To ITIL and Beyond!* Springer, Berlin, 1st edition, September 2007.

[Axe00] Stefan Axelsson. The Base-Rate Fallacy and the Difficulty of Intrusion Detection. *ACM Trans. Inf. Syst. Secur.*, 3(3):186–205, 2000.

[Bar05] Wolfgang Barth. *Nagios: System and Network Monitoring*. No Starch Press, 1st edition, December 2005.

[Bar08] Daniel J. Barrett. *MediaWiki (Wikipedia and Beyond)*. O'Reilly Media, Sebastopol, CA, US, 1st edition, October 2008.

[BT05] Ulrike Buhl and Jürgen Töns. *ITIL Praxisbuch. Beispiele und Tipps für die erfolgreiche Prozessoptimierung*. Mitp-Verlag, 1 edition, 2005.

[CW07] David Cannon and David Wheeldon. *Service Operation ITIL, Version 3 (ITIL)*. Stationery Office Books, Norwich, UK, 2007.

[EGHW07] Anja Ebersbach, Markus Glaser, Richard Heigl, and Alexander Warta. *Wiki: Web Collaboration*. Springer, Berlin, 2nd completely rev. and enl. edition, November 2007.

[HO09] Kai M. Hüner and Boris Otto. The Effect of Using a Semantic Wiki for Metadata Management: A Controlled Experiment. In *HICSS '09: Proc. 42nd Hawaii Int. Conf. on System Sciences*, pages 1–9. IEEE Computer Society, 2009.

[Jon07] Don Jones. *VBScript, WMI, and ADSI Unleashed: Using VBScript, WMI, and ADSI to Automate Windows Administration (Unleashed)*. Addison-Wesley Longman, Amsterdam, 2nd edition, June 2007.

[KA09] Frank Kleiner and Andreas Abecker. Towards a Collaborative Semantic Wiki-based Approach to IT Service Management. In Adrian Paschke, Hans Weigand, Wernher Behrendt, Klaus Tochtermann, and Tassilo Pellegrini, editors, *Proceedings of I-SEMANTICS '09, 5th International Conference on Semantic Systems*, 2009.

[KAB09a] Frank Kleiner, Andreas Abecker, and Sven F. Brinkmann. WiSyMon - Managing Systems Monitoring Information in Semantic Wikis. In *Advances in Semantic Processing, 2009. SEMAPRO '09. Third International Conference on*, pages 77–85, October 2009.

[KAB09b] Frank Kleiner, Andreas Abecker, and Sven F. Brinkmann. WiSyMon: Managing Systems Monitoring Information in Semantic Wikis. In *WikiSym '09: Proceedings of the 5th International Symposium on Wikis and Open Collaboration*, New York, NY, USA, 2009. ACM.

[KAL09] Frank Kleiner, Andreas Abecker, and Ning Liu. Automatic Population and Updating of a Semantic Wiki-based Configuration Management Database. In Stefan Fischer, Erik Maehle, and Rüdiger Reischuk, editors, *Informatik 2009 – Im Focus das Leben*, volume P-154. Bonner Köllen Verlag, 2009.

[Koh06] Toby Kohlenberg, editor. *Snort Intrusion Detection and Prevention Toolkit (Jay Beale's Open Source Security)*. Syngress Media, U.S., pap/cdr edition, December 2006.

[KVV+07] Markus Krötzsch, Denny Vrandečić, Max Völkel, Heiko Haller, and Rudi Studer. Semantic Wikipedia. *Journal of Web Semantics*, 5:251–261, December 2007.

[Lan10a] Ryan Lane. Creating a simple ticketing system with Semantic MediaWiki. Online: `http://ryandlane.com/blog/2010/04/01/creating-a-simple-ticketing-system-with-semantic-mediawiki/`, April 2010.

[Lan10b] Ryan Lane. Helpdesk system and datacenter inventory Semantic MediaWiki prototypes added to my prototype wiki. Online: `http://ryandlane.com/blog/2010/03/29/helpdesk-system-and-datacenter-inventory-semantic-mediawiki-prototypes-added-to-my-prototype-wiki/`, March 2010.

[LM07] Shirley Lacy and Ivor Macfarlane. *Service Transition, ITIL, Version 3 (ITIL)*. Stationery Office Books, Norwich, UK, 2007.

[Roe99] Martin Roesch. Snort - Lightweight Intrusion Detection for Networks. In *LISA '99: Proceedings of the 13th USENIX Conference on System Administration*, pages 229–238, Berkeley, CA, USA, 1999. USENIX Association.

[SBBK09] Sebastian Schaffert, François Bry, Joachim Baumeister, and Malte Kiesel. Semantische Wikis. In Andreas Blumauer and Tassilo Pellegrini, editors, *Social Semantic Web*, X.media.press, pages 245–258. Springer, Berlin, 1st edition, 2009. In German.

[Sch05] Thomas Schwenkler. *Sicheres Netzwerkmanagement: Konzepte, Protokolle, Tools (X.Systems.Press)*. Springer, Berlin, 1 edition, December 2005.

[SS09] Steffen Staab and Rudi Studer. *Handbook on Ontologies. (International Handbooks on Information Systems)*. Springer, Berlin, 2nd edition, 2009.

SVoNt – Version Control of OWL Ontologies on the Concept Level

Markus Luczak-Rösch, Gökhan Coskun, Adrian Paschke, Mario Rothe, Robert Tolksdorf
{luczak,coskun,paschke,mrothe,tolk}@inf.fu-berlin.de

Abstract: Like in Software Engineering the development of ontologies in a team requires efficient support for the management of ontology versions. In this paper we introduce an SVN-based approach for versioning of W3C OWL ontologies called SVoNt – a subversion system for ontologies. Our major goal was to enable views on the revision history of an ontology on the concept level in the same fashion as it is known from the classical SVN approach for the document level.

1 Introduction

Different versions of ontologies can arise in the possibly iterative and collaborative ontology engineering process, as well as during runtime for reusing, extending and refactoring ontologies to incorporate, e.g. knowledge changes and optimizations. Efficient version management is an important functionality in the ontology life cycle management.

Several works on ontology versioning exist which capture especially the key problems of (1) how to detect and track ontology changes, (2) how to merge different ontology versions, and (3) how to release and distribute new versions of ontologies such as the works of Klein[Kle02] and Noy[NM02] amongst others. However, an widely applied engineering tool for ontology version management is still missing.

In this paper we introduce an Apache Subversion (SVN)-based approach for versioning OWL ontologies that conform to the lightweight description logics $\mathcal{EL}$. The system is called SVoNt – subversion for ontologies. Our major goal was to enable a view to the revision history of an ontology on the concept level in the same fashion as it is well-known from the general SVN approach on the file system level. The major difference to other ontology versioning approaches is that we keep the backward compatibility to classical SVN clients. The prototype is implemented as a plug-in for the Eclipse IDE. The adoption and integration into mature ontology development tools which also set up on the Eclipse IDE, such as the NeOn toolkit, TopBraid Composer, and Protégé, is possible in an easy way.

The paper is structured as follows. In the next section we report on existing and related work in the field of ontology versioning. In Section 3, we present the system design of our approach and the current state of the prototype implementation of SVoNt. This paper closes with the conclusions from our recent research and development activities related to SVoNt and a road map of our future steps in Section 4.

2 Related Work

In the past years several approaches to Ontology versioning were pursued. Prompt [NM03] is a collection of tools for ontology versioning realized as a plugin for the Ontology IDE Protégé. The versioning works on a local basis, so this is no classical versioning system. Its feature list contains the calculation of the semantic difference (ontology diff) and the semi-automatical merge of ontologies. Collaborative Protégé [TNTM08] offers another approach to develope ontologies in a group, through an annotation and voting system for changes. Another tool with enabled versioning support is TopBraid Composer[Top] , with usual subversion versioning support and local change history and rollback ability. SemVersion [VG06] is an RDF-centered approach for an ontology versioning system. Model-based versioning is used, so that every version of the ontology is represented by a Triple model. Each version contains Metainformation triple in its model. The Ontology Engineering Platform NeonToolKit (NTK) offers some possibilities for versioning ontologies. The change capturing plugin[NeOa] logs local changes on an ontology and synchronizes them with an external registry. The OWLDiff plugin[NeOb] for NTK enables difference detection and merging of ontologies, two main tasks for a versioning system. Timothy Redmond et al. [RSDT08] shows the design of an ontology version control system that is build up from the scratch. Roman Kontchakov[Kon08] developed a theoretical framework for comparing different versions of DL-Lite ontologies. The problem of the semantic difference of ontologies is handled by an work of Enrico Franconi et al. [EF10].

3 Design and Implementation of SVoNt

The SVoNt system consists of an extended Subversion server and a special SVoNt client. The extensions of the server reuse existing functionalities like logging, authentification and versioning features of Apache Subversion. This allows integrating the SVoNt server into existing Subversion environments, such as Eclipse, and enables classical SVN clients to address the server for OWL document file management - but, without the ontology specific versioning functionalities which can be used only by an extended SVoNt client. In the SVoNt system the W3C Web Ontology Language (OWL) is used as basis for versioning. Each SVN repository represents the evolution of an ontology. To keep the backward compatibility to classical SVN clients we decided that each ontology is located in a specific OWL file in the repository and is not partitioned into several files representing single ontology entities.

To express changes of the structure or the semantics of an ontology an additional versioning mechanism is used, which combines the Subversion system to version the ontology source file and a separate system for concept-based versioning of the ontology. This means that every concept owns a specific revision number, which works equivalent to the Subversion revision system for single files. Therefore conceptual changes of the ontology must be created on the SVoNt server, because Subversion only submits file line deltas to the server. These semantic changes are generated on the server by performing a diff between the updated ontology and the base ontology. This way of generating a difference is not trivial

and only works if the ontologies are syntactically and semantically consistent, thus this consistence has to be checked. The changes are deposited persistently in the conceptual change log and served via an additional external interface used by a SVoNt client.

Because Subversion uses a client-server-architecture and SVoNt directly extends this system, SVoNt is using such an architecture as well. The server stays connectable with a regular SVN client, because the extended SVN interface of the SVoNt system is fully compatible with Subversion. The overall architecture of the SVoNt system is depicted in Figure 1.

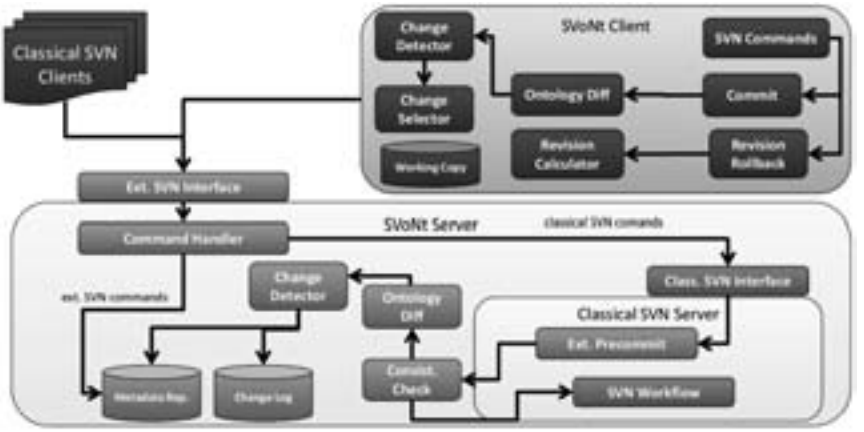


Figure 1: Design of the SVoNt System Architecture

The SVoNt server is an extended SVN server with additional functionality for versioning ontologies. E.g. there is a check done through a precommit-hook to see, if an ontology was changed and needs to be handled in the special ontology versioning process. This process follows a specific procedure running through the modules (1) consitency check, (2) change detection/ontology diff and (3) change log/metadata repository. To allow access to the additional versioning information generated by this process, the access interface of the server was extended. That way an external access apart the classic SVN interface is granted.

As mentioned before the classical SVN clients can use the regular SVN information provided by the SVoNt server. An extended SVoNt client on is a specialized software tailored for versioning ontologies. In addition to the classic SVN revision information it uses ontology-based concept changes from the change log of the SVoNt server. This information is used to display versioning specific metadata in the concept view. A rollback module checks, whether a particular revision of concepts can be undone without generating incompatibilities. Furthermore the client keeps a local change log that stores changes since the last update of the ontology. These, together with the SVoNt server, are detected by a local Change-Detection module. The Change-Selector module allows for the user to choose specific changes to either undo or commit them to the versioning server.

The server-sided extension of Subversion consists of an ontology specific processing of the commit of an ontology, which runs through the three mentioned modules sequentially. On a commit to the SVoNt server a precommit hook is executed on the server to check, if the commit contains an ontology. In such a case the versioning process is triggered. For that the ontology is passed on to an OWL validator, which scans the ontology for specific properties. If that is not the case the execution of the commit is cancelled and the client receives a specific error. In a positive case the Change-Detection module is activated, which generates the difference between the basis ontology and the updated ontology. If no changes between the ontologies are detected, the precommit is succeeded and the ontology is being versioned in Subversion. This is possible if textual comments in the ontology or the order of concepts are changed, leaving the semantics of the ontology untouched. In the case of detecting changes with the ontology diff, these changes are passed to the Change log module, which writes them to the change log of that revision. With that the ontology specific processing is completed and the ontology can be added to the SVN repository and being versioned.

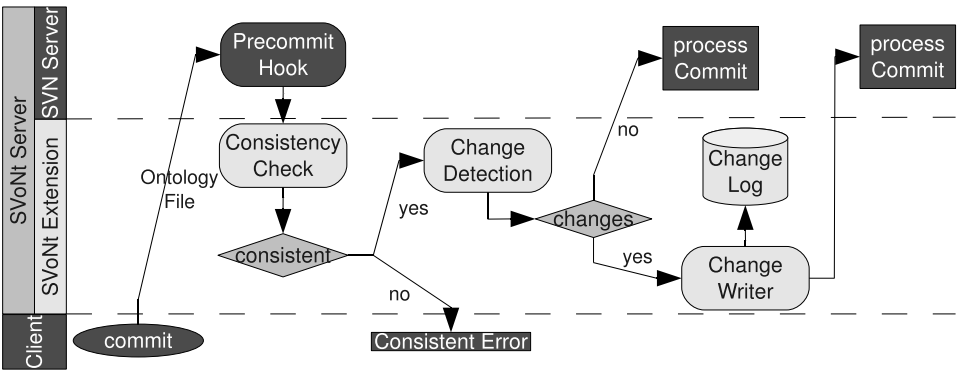


Figure 2: Workflow of the SVoNt Commit Operation

The connection between the additional SVoNt specific components to the SVN server is done via precommit hook scripts provided by Subversion, what enables to interfere with the commit process. The SVoNtRunner, whose `run()` method triggers the ontology specific versioning process, is the heart of the implementation. For the standard implementation the validation module the OWL reasoner Pellet with the OWL-API is used to check the syntactic and semantic validity of the ontology.

The Change-Detection (Diff) module uses algorithms of the OWLdiff (*http://krizik.felk.cvut.cz/km/owldiff/*) library, which can be integrated in adapted form. It provides (1) the basic ontology comparison algorithm and (2) the CEX logical diff. The first algorithm is an implementation of an axiomatic difference generation, which detects the structural changes between the ontologies and returns those axioms that were added or removed. With an entailment checking with Pellet it is possible to determine the semantic equivalence of both ontologies. The OWLdiff implementation of the CEX Logical Diff Algorithm is based on the work of Konev et al.[KWW08] and is able to detect pervasive changes

in the semantics of two ontologies which comply to the $\mathcal{EL}$ description logics. The algorithm returns two sets of concepts which represent the difference between two ontologies which cannot be detected in the class hierarchies. In detail this solution bases on deciding the Σ-entailment for the lightweight description logics $\mathcal{EL}$ which means that it decides whether two compared terminologies imply the same concept implications with reference to a common signature. This implementation is used to identify concepts as being changed in marking all concepts as semantically changed, which are returned by CEX-Diff. Ontologies besides the $\mathcal{EL}$ language profile have to use the basic diff algorithm. The semantic changes of the ontology being recognized are brought into a persistent state by generating a file in the Ontology Metadata Vocabulary (OMV) format for each commit, which represents a revision.

4 Conclusions and Future Work

In this paper we presented the SVN-based ontology versioning system SVoNt. Our next step is to run a case study with some real life data of one of the industrial partners of the German research project Corporate Semantic Web[1] – the Ontonym GmbH – which is a provider of ontology-based services which supply the appropriate background knowledge for the client's applications and services. Ontonym needs an efficient ontology versioning strategy since the company has to handle several often and rapidly changing ontologies which are accessed and maintained by several people internally. On the one hand, these people are the ontology engineers but on the other hand software engineers have to access the ontology as well from time to time. The former prune and refine the ontology depending on new knowledge they observed by evaluating queries and responses of their services as well as time consuming research in the customer's domain. The latter have to align the service functionalities depending on the capabilities of the ontology. In summary, Ontonym needs an ontology version control system for users with disparate viewpoints and skills. SVoNt is designed to fulfill this requirement because it adopts principles from the well-known SVN system and may be used with classical SVN clients.

We also started to implement and study some simple graph analyses which allow representing the content and the evolution of the ontologies through appropriate visualization techniques and envision to simplify comprehensibility. To achieve that, we investigated in ontology visualization techniques (based upon the SONIVIS:Tool (*www.sonivis.org*)) which simplify the reuse of ontologies by reducing the costs for analyzing ontologies.

In future work, we will extend these visualization techniques with versioning information from SVoNt, so that information about the change between different versions can be visualized. By combining the information about the structure of the ontology and its evolution it is possible to identify frequently changing parts within the ontology and to mark them as unreliable and unstable sections that need further attention. On the other hand, parts of the ontology model, which do not change over several versions can be marked as stable and reliable.

[1] *www.corporate-semantic-web.de*

Acknowledgement

This work has been partially supported by the "InnoProfile-Corporate Semantic Web" project funded by the German Federal Ministry of Education and Research (BMBF).

References

[EF10] I. Varzinczak E. Franconi, T. Meyer. Semantic Diff as the Basis for Knowledge Base Versioning. In *13th Workshop on Non-Monotonic Reasoning (NMR)*, 2010.

[Kle02] Michel A. C. Klein. Versioning of Distributed Ontologies. Technical Report Del 20, Vrije Universiteit Amsterdam, december 2002.

[Kon08] Roman Kontchakov. Can you tell the difference between DL-Lite ontologies. In *In Proceedings of KR'08*, pages 285–295. AAAI Press, 2008.

[KWW08] Boris Konev, Dirk Walther, and Frank Wolter. The Logical Difference Problem for Description Logic Terminologies. In *IJCAR*, volume 5195 of *Lecture Notes in Computer Science*, pages 259–274. Springer, 2008.

[NeOa] NeOn Toolkit. Change Capturing Plugin. http://www.neon-toolkit.org/wiki/Change_Capturing. Visited on June 17th, 2010.

[NeOb] NeOn Toolkit. OWLDiff Plugin. http://www.neon-toolkit.org/wiki/OWLDiff. Visited on June 17th, 2010.

[NM02] Natalya F. Noy and Mark A. Musen. PROMPTDIFF: A Fixed-Point Algorithm for Comparing Ontology Versions. In *National Conference on Artificial Intelligence and Fourteenth Conference on Innovative Applications of Artificial Intelligence*, pages 744–750, Edmonton, Alberta, Canada, July 2002.

[NM03] Natalya F. Noy and Mark A. Musen. The PROMPT suite: Interactive tools for ontology merging and mapping. *International Journal of Human-Computer Studies*, 59(6):983–1024, 2003.

[RSDT08] Timothy Redmond, Michael Smith, Nick Drummond, and Tania Tudorache. Managing Change: An Ontology Version Control System. In *OWLED*, volume 432 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2008.

[TNTM08] Tania Tudorache, Natalya Fridman Noy, Samson Tu, and Mark A. Musen. Supporting Collaborative Ontology Development in Protégé. In *International Semantic Web Conference*, volume 5318 of *Lecture Notes in Computer Science*, pages 17–32. Springer, 2008.

[Top] TopQuadrant. TopBraid Composer. http://www.topquadrant.com/products/TB_Composer.html. Visited on June 17th, 2010.

[VG06] Max Völkel and Tudor Groza. SemVersion: RDF-based Ontology Versioning System. In *Proceedings of the IADIS International Conference WWW / Internet 2006 (ICWI 2006)*, 2006.

Commonsense Ontologies and the Use of Words in Natural Language

Ruth Janning

FernUniversität in Hagen
Fakultät für Mathematik und Informatik
ruth.janning@fernuni-hagen.de

Abstract: Since the appearance of the 'Semantic Web' and the development of 'RDF' and 'OWL', ontologies gained new importance in computer science. Ontological structures can be used to make knowledge available to artificial intelligent systems. But such systems need commonsense knowledge to simulate human reasoning beyond the boundaries given by specific domains. For this purpose commonsense ontologies are employed. However, existing commonsense ontologies, e.g. Cyc [Cy10], were constructed over a lengthy period of time. An interesting proposal to reach this in shorter time with less effort is to expose the structure of commonsense knowledge by analyzing the use of words in natural language. Based on this, a method to automatically gain commonsense ontologies with less effort will be presented. The main point of this method is the automatization. For this purpose data mining techniques are applied and an algorithm to generate the resulting ontology out of the gained data is introduced.

1 Exposing the structure of commonsense knowledge

To avert the immense effort[1] for the construction of commonsense ontologies, there is need for new methods to gain these ontological structures. Saba [Sa07] proposes to expose – and not newly develop – the structure of commonsense knowledge by analyzing the use of words in natural language. According to this, natural language should guide the process of gaining a commonsense ontology.

2 Four steps toward an ontological structure

Asking a child, if it makes sense to say 'a dog barks' or 'a book barks', one gets an explicit positive or negative answer. On this simple base – the view of a child – the analysis of words is executed.

[1] Knowledge was fed over 20 years into the knowledge base of Cyc [Cy10] until the system was able to learn by itself [Wi05].

Accordingly, quantitative or philosophic assessments, which an adult would maybe have in mind (e.g. there are dogs that do not bark or such dogs who bark louder or lower), are not regarded. So it is a binary decision if it in general makes sense to say 'a dog barks' or not.

Introducing a predicate *App(p,c)* [Sa07], one can describe a process which guides within four steps toward an ontological structure (see Figure 1). In this expression p denotes a property (adjective) or action (verb) and c a concept (noun)[2]. *App(p,c)* receives the value *true* if it makes sense to speak of the property or action p of c.[3]

1. Assume a set of concepts (nouns) $C = \{c_1, \ldots, c_m\}$ and a set of properties (adjectives) or actions (verbs) $P = \{p_1, \ldots, p_n\}$ to be already known.

2. Furthermore a predicate *App(p,c)*, $c \in C$ and $p \in P$, is given. *App(p,c)* becomes *true* if the property or action p is reasonably applicable to objects of type c (i.e. if it makes sense to speak of the property or action p of c).

3. For every property or action $p \in P$ a set $C_p = \{c \mid App(p,c)\}$, which includes all concepts c for which *App(p,c)* is *true*, is generated.

4. As a result, the desired hierarchy is gained through an analysis of the subset relationship between the sets generated in step 3.

Figure 1: Four steps toward an ontological structure.

3 Realizing and automatizing the four steps

This process of four steps proposed by Saba [Sa07] is an interesting beginning and should be further developed. Now a suggestion for realizing and especially for automatizing the four steps will follow.

3.1 Step 1 and 2: Choice of the sets and evaluation of the combinations

To automatize step 1 and 2, i.e. the choice of the sets of nouns and adjectives or verbs and the decision, whether a specific combination makes sense, data mining techniques should be used. That is, a computer searches in given texts (e.g. from the numerous sources of the Internet like Wikipedia [Wi10a]) for nouns with a preceding adjective or following verb.[4] To detect the nouns, adjectives and verbs, the computer uses a computerized dictionary (e.g. WordNet [Pr10] or Wiktionary [Wi10b]).

[2] Nouns are supposed to be plain base concepts.
[3] If a child answered 'yes' to the question whether this combination makes sense.
[4] Auxiliary verbs should be excluded. For words with the same character string, which can be nouns on the one hand and adjectives on the other hand (e.g. 'human'), the meaning of the first discovered word is used. The other meaning, respectively combinations with it, will be ignored in the following.

The discovered words (e.g. as given in Figure 2) are inserted in alphabetical order into a table (see Figure 3). The nouns occur in the first column (down from the second row) and the adjectives and verbs in the first row (except the first cell of the first column). The combinations of nouns and adjectives or verbs are explored with data mining by checking how often some combinations occur (checking the *support* of *associations* [BKI06]). Often occurring combinations (with higher support than other combinations)[5] get a '+' in the corresponding table cell (that means *App(p,c)* is *true* for the combination of *p* and *c*). Otherwise – if the support has a value below the determined threshold – the table cell gets the symbol '-' (see Figure 3).

C	=	{bird, book, dog, machine, man, woman}
P	=	{bark, exist, defective, fly, live, pregnant, read}

Figure 2: Example sets.

	bark	**exist**	**defective**	**fly**	**live**	**pregnant**	**read**
bird	-	+	-	+	+	-	-
book	-	+	-	-	-	-	-
dog	+	+	-	-	+	-	-
machine	-	+	+	-	-	-	-
man	-	+	-	-	+	-	+
woman	-	+	-	-	+	+	+

Figure 3: Example table.

3.2 Step 3 and 4: Obtaining concept sets and generating a concept hierarchy

The sets C_p from step 3 can be obtained by picking out the table column corresponding to *p*. A set C_p contains all nouns of which the relevant table cell holds the symbol '+' (see Figure 4). Step 4 can be realized by an algorithm (see Figure 5). The root node of the desired hierarchy is the set which contains all considered concepts (nouns). The algorithm first chooses from the sets C_p the largest proper subset of the root node set and inserts this subset as left son of the root into the hierarchy. The right son node is the complementary set of the left son. Subsequently the algorithm chooses the largest subsets of both gained son nodes and their son nodes and so on until there are no more subsets other than the empty set. The reason for choosing the largest subset is that the sets, which represent nodes, should become smaller downward. That is, they should become more and more specific and represent more specific concepts. The result is an ontological structure like the one shown in Figure 6. In such a hierarchy all son nodes and descendants are sub concepts of their father node.

The obtained hierarchy causes relations between the different concepts and allows to derive rules, e.g. in first-order logic[6], like

$$\forall c(App(bark,c) \Rightarrow \neg App(read,c))$$

[5] Of course a reasonable threshold has to be chosen.
[6] In this paper no specific ontology representation language is mentioned since the method is yet a conception.

which means that for objects, which can bark, it does not make sense to say they read (in short: dogs cannot read).

C_{exist}	=	{bird, book, dog, machine, man, woman}	(complementary set: {})
C_{live}	=	{bird, dog, man, woman}	(compl. set: {book, machine})
C_{read}	=	{man, woman}	(compl. set: {bird, dog})
$C_{defective}$	=	{machine}	(compl. set: {book})
$C_{pregnant}$	=	{woman}	(compl. set: {man})
C_{bark}	=	{dog}	(compl. set: {bird})
C_{fly}	=	{bird}	(compl. set: {dog})

Figure 4: Sets C_p.

Method: generateHierarchy()
Given: Set which contains all concepts.
Output: Concept hierarchy.
Algorithm:

1. Initialization of the root node of the concept hierarchy with the set which contains all concepts.
2. Call of the recursive method getSubSets(<*set*>) with the set of the root node as its argument:
   getSubSets(*root node set*)
3. **return** concept hierarchy

Method: getSubSets(<*set*>)
Given: Discovered sets C_p of concepts (see above step 3).
Input: Set of which the largest subset and its complement are to be determined and inserted as left and right son.
Algorithm:

1. Find all proper subsets of <*set*>.
2. Determine the largest set of these subsets (if there are more than one, choose those, of which the related property or action is alphabetically smaller than all the others).
3. *LeftSonNode* = determined largest set.
4. *RightSonNode* = <*set*> \ *LeftSonNode*.
5. **if** *LeftSonNode* $\neq \emptyset$
   then
   - Insert into the hierarchy *LeftSonNode* as left son of the node which is represented by <*set*>.
   - Insert into the hierarchy *RightSonNode* as right son of the node which is represented by <*set*>.
   - Determine the subsets of the left son: getSubSets(*LeftSonNode*).
   - Determine the subsets of the right son: getSubSets(*RightSonNode*).

Figure 5: Algorithm for the concept hierarchy generation (step 4).

The presented algorithm seems to be similar to the *Formal Concept Analysis* (FCA) [GW96] but the procedure as well as the result is different in both methods. Contrary to FCA, in this algorithm a concept is composed of only one attribute (adjective or verb) and a set of objects (nouns). Additionally this algorithm produces exclusively binary trees and in the given example the hierarchy generated by FCA is much shallower.

An efficient *approach to learning taxonomies or concept hierarchies from text*, which uses FCA and creates domain ontologies, is presented in [CHS04].

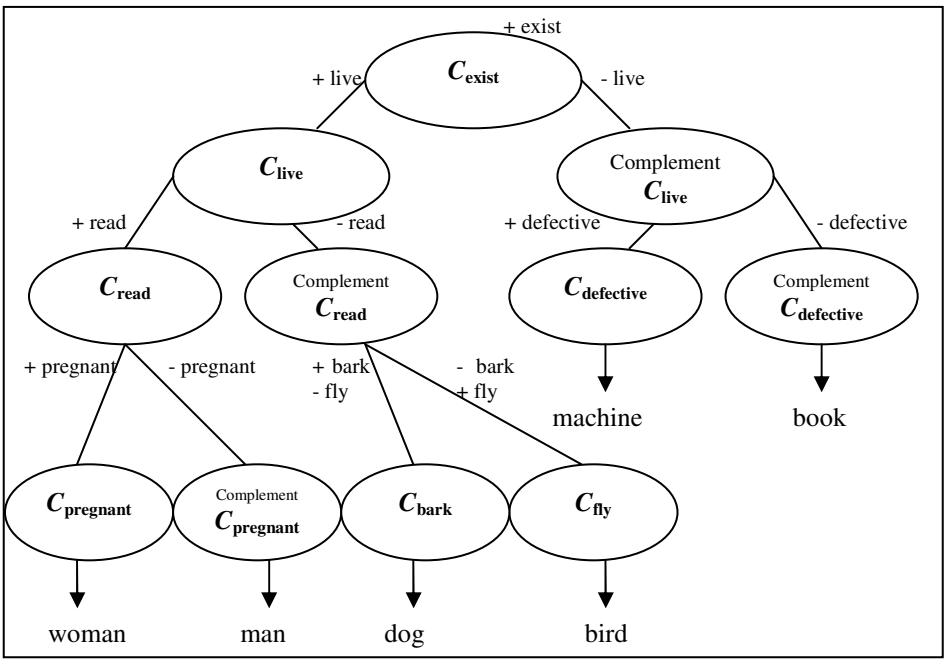


Figure 6: Generated ontological structure.

An oval stands for a set of nouns, respectively for a concept, which includes all the sub concepts represented by the nouns. An edge symbolizes the relationship between the two linked concepts: The concept below includes all nouns, to which the considered property or action is applicable. So C_{read} contains all concepts from C_{live} (existing living objects), which possess the capability to read. The arrows indicate concepts from sets, which possess just one element.

4 Related work

The proposed method is also similar to the underlying ideas of the platform OLE [NS05] since OLE deals with *automatic acquisition of semantic relations from texts* [NS05]. But there are some differences between the two methods. OLE is a *platform for bottom-up generation and merging of ontologies* and uses a *pattern-driven concept extraction process* based on proposals by M. A. Hearst [He92]. The method presented here is a kind of top-down clustering. Moreover it searches directly for nouns, adjectives and verbs in the given texts and uses data mining techniques. Also the OntoLearn system [NVM04] aims at *extracting knowledge from electronic documents to support the rapid construction of a domain ontology*. But its machine concept learning is an iterative process and is initially *based on the use of external, generic knowledge sources*. So it uses WordNet [Pr10] *as a start-up ontology*. Furthermore, OntoLearn uses statistical techniques.

5 Conclusions

The presented method is a good beginning, but there are still some problems. To get a useable ontological structure for commonsense knowledge, a very large number of nouns and adjectives or verbs need to be cumulated. In addition to this, these words must be sufficiently different from each other to gain a full commonsense ontology. Every concept of the resulting hierarchy must possess a unique characteristic to be different from other concepts. A further problem results from the need to separate original and metaphorical meanings of words, which must appear at different places in the hierarchy. This is not avoidable, if one wants to get an unambiguous and reasonable ontological structure. But in normal texts words are used in their original meaning as well as in metaphorical meanings. Possibly the usage of data mining techniques is already a solution. These methods give a '+' only to combinations of substantives and verbs or adjectives, which reach a high support. So if we suppose that, compared to the original meaning, metaphorical meanings occur scarce in texts from lexicons like Wikipedia [Wi10a], they will be ignored. Finally, the presented method to automatically gain commonsense ontologies is yet a conception. The next steps will be: completely implementing the method and executing empirical tests.

Bibliography

[BKI06] Beierle, C.; Kern-Isberner, G.: Data Mining und Wissensfindung in Daten. In: Methoden wissensbasierter Systeme. 3. Vieweg Verlag, 2006; chapter 5.5, pp. 141-154.

[CHS04] Cimiano P.; Hotho A.; Staab S.: Clustering Ontologies from Text. In: Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC). 2004; pp. 1721-1724.

[Cy10] Cycorp Inc.: Cycorp, Inc.. http://www.cyc.com/.

[GW96] Ganter, B.; Wille R.: Formale Begriffsanalyse: mathematische Grundlagen. Springer-Verlag, 1996.

[He92] Hearst, M. A.: Automatic acquisition of hyponyms from large text corpora. In: Proceedings of the 14th conference on Computational linguistics. Association for Computational Linguistics, Morristown, NJ, USA, 1992; pp. 539-545.

[NS05] Nováček, V.; Smrž, P.: OLE - A New Ontology Learning Platform. In: Proceedings of International Workshop on Text Mining Research, Practice and Opportunities. Incoma Ltd., 2005; pp. 12-16.

[NVM04] Navigli, R.; Velardi, P.; Missikoff, M.: Web Ontology Learning and Engineering: An Integrated Approach. In (Zhong N.; Liu J. Eds.): Intelligent technologies for information analysis. Springer-Verlag, 2004; chapter 10, pp. 223-242.

[Pr10] Princeton University: About WordNet. http://wordnet.princeton.edu/.

[Sa07] Saba, Walid S.: Language, logic and ontology: Uncovering the structure of commonsense knowledge. In: International Journal of Human-Computer Studies 65, March 2007; pp. 610-623.

[Wi05] Witbrock, M.; Matuszek, C.; Brusseau, A.; Kahlert, R.C.; Fraser, C.B.; Lenat, D.B.: Knowledge Begets Knowledge: Steps towards Assisted Knowledge Acquisition in Cyc. In: Papers from the 2005 AAAI Spring Symposium on Knowledge Collection from Volunteer Contributors (KCVC). Stanford, California, 2005; pp. 99-105.

[Wi10a] Wikimedia Foundation, Inc.: Wikipedia. http://www.wikipedia.org/.

[Wi10b] Wikimedia Foundation, Inc.: Wiktionary. http://wiktionary.org/.

Utilization of Semantic Networks for Education: On the Enhancement of Existing Learning Objects with Topic Maps in <ML>³

Ulrike Lucke, Alke Martens

University of Rostock, Institute for Computer Science
Albert-Einstein-Str. 21, 18059 Rostock, Germany
firstname.lastname@uni-rostock.de

Abstract: Semantic networks represent dependencies between concepts or topics, and thus are well-suited not only for processing by computers but also for knowledge transfer to learners. A challenge is to combine knowledge networks (for self-directed learning) with traditional, hierarchical courseware. The paper presents an approach for integration of both, semantic networks and hierarchical structures, within a single learning object based on the eLearning markup language <ML>³.

1 Educational Benefits of Semantic Networks

Semantic networks are also described as knowledge networks, since they combine isolated facts with associations and meanings to applicable knowledge. That's why an utilization for knowledge transfer and description of learning objects (LOs) stands to reason. Learning and knowledge construction (in the constructivist theory) are related to the idea that each individual has to *re-construct knowledge structures* based on given information in interaction and in relation with the individual's own mental model(s) and prior knowledge. The mental model of a person can best be compared with a knowledge network – i. e. in computer science terms a mental model consists of associations and relations between information parts (e. g. facts). When learners re-construct these facts and relations for other learners, the result is comparable to a semantic network.

Coming from the perspective of instructional design, the traditional structure of courses and training material usually is purely hierarchical (e. g. based on categorizations, top-level/sub-level structures, easy-to-difficult structures and the like). But a closer look reveals that often sub-structures are realized. For example, in the preamble of a book, the author might explain how to use the book – e. g. a beginner should read chapter 1 first and then continue with chapter 3, whereas the advanced learner could start at chapter 2. The index structure, read on a meta-level, often reveals a knowledge network.

When instructional material is developed for computer-based teaching and training settings, the advantages of semantic networks or task models can easily be used in the instructional design: instead of a hierarchical composition of the training material, the

content can be developed a) for different levels of expertise (i. e. adaptable), b) in networked relations. The result is a navigation structure where not only the information hidden behind the "information nodes" is relevant and part of the learning, but also the nodes' interrelation [Ma04]. The relation of facts which currently *should* be learned with those facts the learner *might* additionally be interested in support the construction of knowledge and thus has the potential to facilitate learning. Moreover, the network-like information composition is one of the advantages of computer-based training.

The paper is structured as follows: A short overview on prior work in this area is given in section 2. Next, section 3 presents the concept of combining several didactical models for given content within a single learning object, and the implementation of this concept in the description language <ML>[3]. Finally, the pro's and con's as well as general conditions for the use of semantic networks in eLearning are discussed.

2 Related Work

There are many facets of using semantic technologies in eLearning. Early and widely accepted mechanisms include the description of *metadata* for educational material in order to make it easily accessible. The most common standard for the educational sector is Learning Object Metadata (LOM). It consists of a variety of fields grouped in 9 categories explaining the characteristics of a resource and the conditions of its use. A more complex problem is content interoperability, to be targeted with interchange standards like the Sharable Content Object Reference Model (SCORM). Metadata and content packaging formats are a strong requirement for identification and re-use of LOs.

A more challenging approach is to use metadata in combination with subject ontologies for automated *course generation*. Considering the learner's specific pre-knowledge, competencies, and needs, the LOs in given repositories can be selected, configured, and combined to individual course material. Evaluation of inter-dependencies between LOs leads to a network structure of the educational content. After a selection of the best suited LOs (regarding topic, didactical approach, technical issues, and so on) and the serialization of the resulting sub-graph, individual material can be generated [Bun03]. Additionally, missing parts of the subject that are not covered by any LO can be identified with the help of a domain ontology [APB03, Bal04, DWC05] and then added manually. These mechanisms make use of semantic technologies *across a set of LOs*.

Another aspect is the utilization of semantic data *inside a learning object*, in terms of a *knowledge network*. The benefits of using relationships between content fragments for visualization of and navigation through a course were shown in several works [DDA04, Hie05]: non-linear knowledge creation, intuitive and memorable visualization of dependencies, synergies by collaboration with other knowledge networks and/or other learners, and starting point for additional semantic and intelligent technologies. Nevertheless, learning objects have to demonstrate their usefulness in traditional educational settings, too. That's why a hierarchical arrangement of content cannot be missed. This leads to the question how to combine content arrangements in hierarchies and in networks without redundant description of LOs.

3 Flexible Didactical Descriptions with <ML>³

The XML based eLearning Language <ML>³ (Multidimensional LearningObjects and Modular Lectures Markup Language) was developed in 2001 [LTV03]. Recent works include a commercial adaptation and development of authoring tools [GLT09]. This also brought up an advanced didactical conception that makes use of semantic networks.

3.1 General Model of <ML>³

Beyond the *separation of content and layout* (as associated with XML in general) <ML>³ implements a *separation of content and didactics*, too. These aspects are described separately, with references from basic didactical units to main structural elements on content-side. This is depicted in Figure 1.

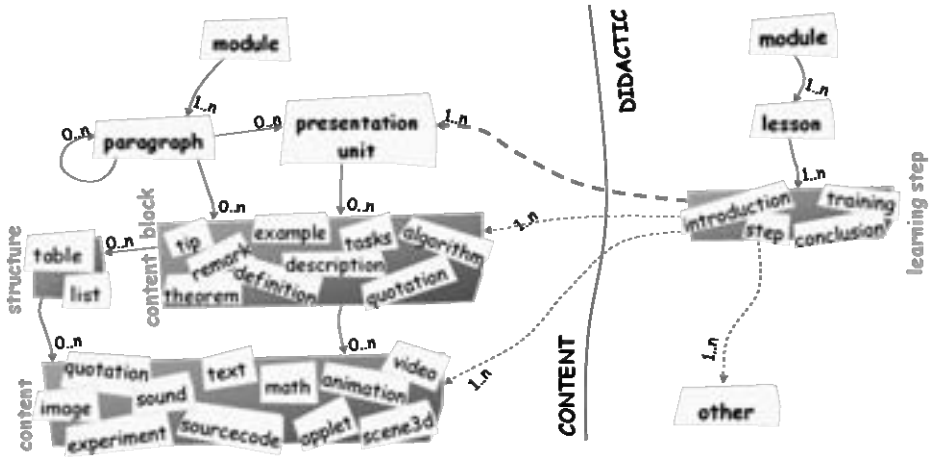


Figure 1: Structural Model of <ML>³ with References from Didactics to Content

Thus, there are the following possibilities to add a didactical description for another educational setting to an existing module:

- additional description based on the same didactical model:
  - re-sorting of existing presentation units
  - re-arrangement of single content elements
  - inclusion of content from external sources
- additional description using a different didactical model

Existing and currently introduced mechanisms for didactical descriptions in <ML>³ are explained in the following sections.

3.2 Traditional Educational Material based on Web Didactics

Most LOs are structured in a hierarchy, applying Meder's concept of Web Didactics [Med06]. Every topic to be communicated is considered as a unit of learning, which

again consists of units of knowledge that can be captured as a whole (like a screen). Media units (like text or images) build the basic level.

<ML>³ implements this concept by a didactical structure divided into lessons and typed learning steps, as depicted in the right-hand part of Figure 1. A lesson contains all content that closely belongs together, like a section or chapter. A step equals the amount of content to be placed together, like a screen or slide. Every step keeps a reference to a presentation unit in the content-side of a module.

The similarity to the structure of existing material (books, slides) makes this model easy to author for the teacher. But, the paths of learners through this material are limited (differences appear mainly from omitting a step or lesson), and individual creation of inter-dependencies between knowledge elements is not forced.

3.3 Topic-Maps for Self-directed Learning with Traditional Material

The use of spatial content arrangements in eLearning is not new, e.g. for combining achieved knowledge to a portfolio or for visualization of content dependencies. Semantic networks are a mature mechanism for this idea, since they include several starting points and typed associations between any elements in the network. That's why we extended the specification of <ML>³ by the possibility to describe Topic Maps according to the ISO standard XTM 2.0 (XML topic map syntax) [GM08]. Figure 2 depicts the resulting vocabulary, including topics as core elements and typed associations between topics. Referencing to presentation units of the content-side of a module is realized from topics.

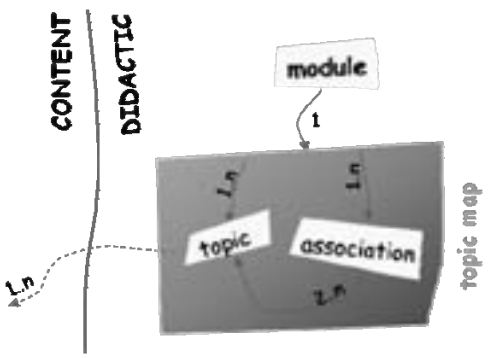


Figure 2: Extended Didactical Model of <ML>³ for the Use of Topic-Maps

An example for a Topic Map is shown in Figure 3, as used in the editors of the XMLeditools [GLT09]. In the project editor, the authors selects between hierarchical or networked structure. Afterwards, the content editor provides mechanisms to edit the topic map as well as the content fragments behind each topic. By clicking on a topic, the corresponding presentation unit is opened. Navigation takes place completely in the network structure. Independently, a classical LO with hierarchical structure may exist that consists of (i. e., references to) the same content fragments. Thus, redundant descriptions during re-use of LOs for different settings are avoided. This offers a higher

degree of autonomy, and a self-directed learning experience to the learner. The intentional navigation and arrangement of topics encourage an inter-connection of knowledge, thus achieving a higher cognitive level and learning outcome.

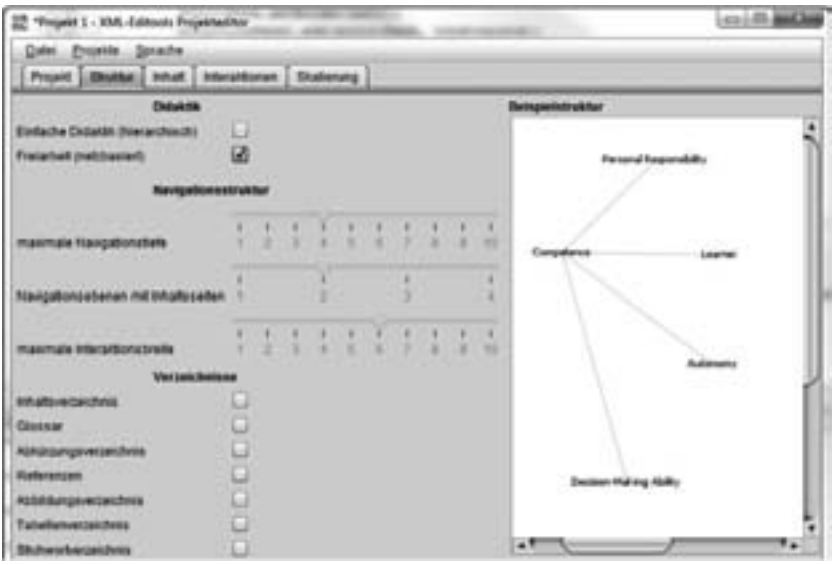


Figure 3: Exemplary Topic-Map from a LO on „Individualized Instruction“ in the XMLeditools

3.4 Discussion

The benefits of the presented approach can be seen in the application of self-directed learning, and of a constructivist learning theory, without having to change existing content. Moreover, hierarchical and networked descriptions of a LO co-exist without the need to change the learners infrastructure (i.e., web browser or learning platform). But, there are also some problems or challenges related to this approach. Basic pre-knowledge on the topic is required from the student in order to navigate through the network. Moreover, development and practical use of such descriptions are not trivial for teachers, and acceptance in conventional educational environments might be reduced. This leads to a set of conditions for this approach. It is useful only for experienced teachers and learners. The most benefit is gained in scenarios with the need to select individually required knowledge out of larger repositories (like learning at work place), best in combination with traditional educational settings.

4 Conclusion and Further Work

Some educational settings may require the use of courseware without pre-defined learning paths. The benefits of such a self-directed learning can be achieved with the help of semantic networks that visualize the structure of the learning content and allow for individual navigation. The eLearning language <ML>³ can be used to consistently

describe different didactical models (i. e., hierarchies and networks) on the same content. An integration of Topic Maps into <ML>³ and LOs was presented. It leads to a degree of didactical flexibility that cannot be achieved by existing approaches without redundant descriptions of content.

An open question in the existing language specification is how to deal with different amounts of knowledge associated to a topic. Currently, a topic is linked to a single presentation unit, but the capacity of a slide or screen is limited. Complex topics might require an author to break down the content into a set of presentation units, or even into a hierarchical sub-structure of presentation units (and sections). This is not yet covered by the <ML>³ specification and will be subject of further studies.

Moreover, first experiences with teachers and students brought up the need for a more guided navigation through the semantic network, e. g. pre-defined learning paths. This can be achieved by labelling associations between topics as mandatory, recommended, or optional. Thus, a multi-level description of possible paths through the learning object can be achieved. Though this would require a schema modification away from the existing XTM standard, it might be of certain benefit to Learning.

References

[APB03] Aroyo, L; Pokraev, S; Brussee, R.: "Preparing SCORM for the Semantic Web", in *On The Move to Meaningful Internet Systems*, LNCS 2888, pp. 621-634, Springer 2003.

[Bal04] Baldoni, M.; Baroglio, C.; Patti, V.; Torasso, L.: "Reasoning about learning object metadata for adapting scorm courseware", in Proc. of *Int. Workshop on Engineering the Adaptive Web* (EAW), pp. 4-13, 2004.

[Bun03] Bunschkowski, M.; Lucke, U.; Röser, M.; Tavangarian, D.: "The role of metadata for the automated generation of educational material", in *Proc. of 2nd European Conference on E-Learning (ECEL)*, Glasgow/UK, November 2003.

[DDA04] Dichev, C.; Dicheva, D.; Aroyo, L.: "Topic Maps for E-Learning", in *International Journal on Advanced Technology for Learning*, 1(1), 2004, pp. 1-7.

[DWC05] Dagger, D.; Wade, V.; Conlan, O.: "Personalisation for All: Making Adaptive Course Composition Easy", in *Educational Technology & Society*, 8 (3), pp. 9-25, 2005.

[GLT09] Gries, V.; Lucke, U.; Tavangarian, D.: „Werkzeuge zur Spezialisierung von XML-Sprachen für die vereinfachte, didaktisch unterstützte Erstellung von eLearning-Inhalten", in Proc. of *Die 7. e-Learning Fachtagung Informatik (DeLFI)*, Berlin, September 2009.

[GM08] Garshol, L. M.; Moore, G.: „*ISO 13250-2: Topic Maps — Data Model*", http://www.isotopicmaps.org/sam/ (published in June 2008)

[Hie05] Hierl, S.: „*Die Eignung des Einsatzes von Topic Maps für e-Learning*", Churer Schriften zur Informationswissenschaft, Nr. 4, Hochschule Chur, 2005.

[LTV03] Lucke, U.; Tavangarian, D.; Voigt, D.: „Multidimensional Educational Multimedia with <ML>³", in Proc. of *Conference on E-Learning in Corporate, Government, Healthcare & Higher Education* (E-Learn), Phoenix/AZ/USA, November 2003.

[Ma04] Martens, A.: „*Ein Tutoring Prozess Modell für fallbasierte Intelligente Tutoring-systeme*". DISKI 281. Akademische Verlagsgesellschaft AKA, infix, Berlin 2004.

[Med06] N. Meder (Hrsg.): „*Web-Didaktik. Eine neue Didaktik webbasierten, vernetzten Lernens*", wbv – Wissen und Bildung im Internet, Band 2, Bertelsmann 2006.

Towards Semantic Matching of Business Services and Electronic Services

Rolf Kluge[1,2], André Ludwig[1], Roman Belter[1,2]

[1]Information Systems Institute
University of Leipzig
04109 Leipzig, Germany
{rkluge,ludwig}@wifa.uni-leipzig.de

[2]Department of Computing
Macquarie University
Sydney NSW 2109, Australia
{rkluge,rbelter}@science.mq.edu.au

Abstract: Service oriented environments consist of business and electronic services. Business services encapsulate core business logic and activities whereas electronic services support the operation of business services by means of software applications. In environments with a large number of business and electronic services an approach for selecting electronic services that satisfy certain business services concerns is needed. This paper presents an approach for matching business and electronic services based on semantic entity similarity.

1 Introduction

In the context of environments which are defined according to the service orientation paradigm such as service-oriented enterprises [He08] and service-oriented computing platforms [PG03] the term service is defined as a self-contained, loosely coupled entity that encapsulates a limited piece of functionality, is reusable, able to be composed, and provides a well defined external interface. Services can be differentiated into business- and electronic services. A business service (BS) is an auxiliary artifact that structures a certain business concern according to the service-orientation paradigm. Electronic services (ES) describe technical aspects (also structured according to the service-orientation paradigm) with a special focus on enterprise software applications. Service orientation as a design paradigm for enterprise and computing environments promises a number of benefits – among them flexible re-configuration, dynamic binding, easy access to heterogeneous resources and processes, transparency across implementation details and last but not least a relatively stable set of standards for aspects such as interface, orchestration, choreography or contracting. Research on service orientation has been conducted in various directions. An area which has received insufficient support is the alignment of BS and ES. BS and ES are interrelated in a way that ESs support the operation of BSs. The spectrum of support ranges from partial to entire. In infrastructures with a large number of ES and BS implementations and which are subject to regular changes the decision on which ES provides the most suitable support for a certain BS becomes quite complex.

This paper presents an approach for matching business and electronic services based on semantic entity similarity. Matching addresses predominantly the evaluation and

decision making process for selecting ES for certain BS. The objective is to filter the large number of available ESs to a distinct set of high potential ES candidates that support a certain BS. The approach is based on the hypothesis that the higher the similarity between BS and ES semantics, i.e. the similarity between semantic entities, the higher the match between a BS and an ES. Based on this hypothesis, an automatic quantifiable similarity measurement of all BS-ES combinations is possible and would reveal suitable BS-ES candidates. The paper is structured as follows: as a starting point assumptions and prerequisites are examined in detail in chapter 2. Chapter 3 presents the approach in detail, chapter 4 examines related work, and chapter 5 summarizes and presents next steps.

2 Hypothesis and Prerequisites

The approach is based on the hypothesis that a certain ES supports the operation of a BS if an ES and a BS are similar in their semantics[1]. Hence, the degree of similarity of the semantic entities of BS and ES determines the degree of support. This assumption includes that if an ES supports the operation of a BS, the ES-description represents (a part of) the BS-description. Further, an ES provides functionality that is implicitly stated in the BS description and, thus, required by the BS. The justification for that hypothesis is given by an analogy to the "SOA Common Information Model" (CIM) [Pa07]. A CIM contains a set of information objects and relations with business semantics (so called business objects). The definition of a CIM is domain specific (focused on a certain business area) and strictly independent from software implementations. According to [Pa07] in "Real SOA" the business objects have to be used on business level and on software level as well. Consequently, a proper software support is only accomplished, if the software contains semantics of the business domain. Adapted to ES and BS, this means that BS and ES should use similar semantics. As a reverse conclusion, BS and ES fit to each other (in a sense that an ES supports the operation of a BS) if both have similar semantic entities. Applied to the presented problem, a measurement of the similarity of semantic entities between a certain BS and a certain ES would reveal the fit-level.

In order to define a meaningful starting point the following prerequisites (italic) are defined. *There is a large number of already defined BSs and ESs.* The approach does not examine how to structure and describe BS and ES. *BS and ES are described on the basis of a common model.* The description of BS and ES is based on a common model in order to ensure basic comparability between BS and ES. The model defines the elements and the syntax of the BS and ES description. *BS and ES have different semantics.* Although the common description model defines basic elements of a service (syntax), the instantiations of this model are different. The differences are derived from the semantics. *BS and ES contain business semantics***.** BSs contain business semantics per se. However, there are ES which exist on technical level only (e.g. messaging service). Thus, a

[1] A service is represented by its service description. This paper does not consider any difference between a service and its description.

restriction is that ESs with a technical focus are not under consideration. Instead, ESs with business semantics are of interest.

3 Approach

Based on the hypothesis and the mentioned prerequisites, the approach has the objective to measure similarities between BSs and ESs. As a result each BS-ES combination gets a quantifiable and comparable similarity value. With this similarity value all possible BS-ES combinations can be ranked. Taking the highest similarity values for a certain BS, the large number of ESs are reduced to a significant small number of high potential candidates. These candidates can be taken for further examination. Hence, the approach will identify potential ES candidates and provide decision support.

In order to calculate the similarity value of a single BS-ES combination, first, the elements under consideration (i.e. semantic entities) have to be identified. Thus, semantic entities have to be extracted from BS and ES descriptions. Second, the semantic entities of BS have to be compared to the semantic entities of ES. A similarity value is measured and calculated for each BS-ES semantic entity combination. These values have to be aggregated in a third step in order to conclude about the similarity value for the BS-ES combination. Calculating the similarity values for each BS-ES combination requires an adequate visualization for decision support. The approach including the key steps extraction, measurement, aggregation and visualisation is illustrated in figure 1 and will be described in detail below.

The starting point is an environment with a large number of BSs (x) and a large number of ESs (y). All services are described by a service description which is based on a common meta-model (as depicted on the left-hand side of figure 1). To illustrate the approach an example BS x_i and an example ES y_j is chosen from the environment and their matching level is calculated. As the first step, semantic entities, i.e. concepts, relations, instances, and constraints, are extracted from the descriptions (cf. [FGJ97]). This leads to K semantic entities of the BS and L semantic entities of the ES respectively. Each BS semantic entity $n_{i,k}$ and each ES semantic entity $m_{j,l}$ form a BS-ES semantic entity combination. There are K*L BS-ES semantic entity combinations. In the next step, a similarity value $e_{i,j,k,l}$ is calculated for each BS-ES semantic entity combination. Since semantic entities for a certain domain form ontologies, ontology matching mechanisms [ES07] are applied in order measure the similarity for each combination $e_{i,j,k,l}$. Ontology matching comprises several matching techniques which can be divided into element- and structure-level techniques at first stage and into syntactic, external and semantic techniques at second stage. However, single matching techniques are not sufficient ([ES07] pp.115). Usually, matching techniques are combined in order to get a proper matching result. The success of matching techniques and strategies for similarity measurement depend on different circumstances, such as the application domain, the use case etc. Hence, providing a concrete matching strategy (including matching techniques) is out of scope of this paper. Nevertheless, after matching on an entity-level there are quantifiable similarity values $e_{i,j,k,l}$ for each BS-ES semantic entity combination. The similarity measurement based on ontology matching is represented by

the function $f(n_{i,k}, m_{j,l})$ in figure 1. After having similarity measures on semantic entity-level, the values have to be aggregated for a certain BS-ES combination in order to conclude about similarity on service-level. The aggregation is embodied by the function $f(e_{i,j,1,1}, e_{i,j,1,2}, \ldots, e_{i,j,k,l}, \ldots e_{i,j,K,L})$ taking into account each $e_{i,j,k,l}$ for a certain BS-ES combination (i,j). The first draft of the approach contains a basic mean aggregation for that purpose (cf. figure 1). Future examinations will elaborate more sophisticated aggregation algorithms (including weighting and normalization for instance). There is a quantifiable similarity value for each BS-ES combination $s_{i,j}$ as a result of the aggregation. For achieving the goal of supporting decision making the pure values are not sufficient. They have to be presented and visualized. The approach provides a ranked list of high potential ESs for each BS as shown on the right-hand side of figure 1.

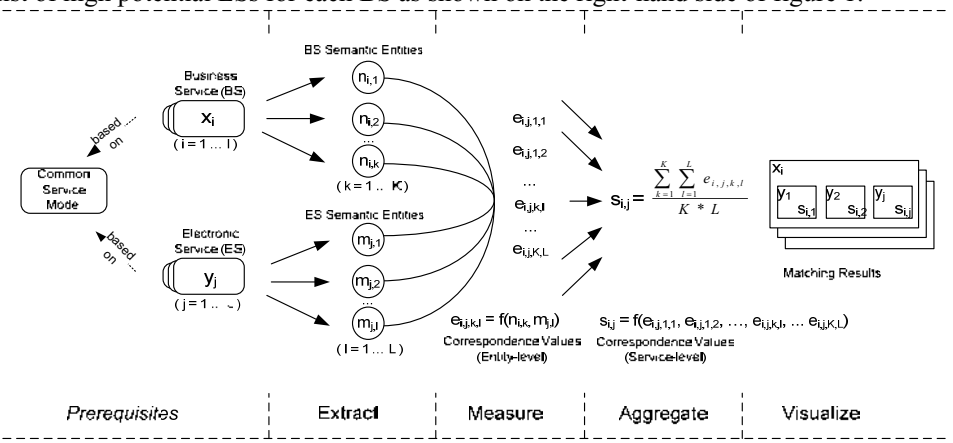


Figure 1. Four step approach for matching of business and electronic services

4 Related Work

The selection of services that fit consumer concerns is an integral part of the service oriented paradigm. The basic SOA model contains the service registry role (also known as service broker) for that purpose. A service consumer discovers a registry in order to find a service that fits its concerns (cf. [Pa07]). The selection of ES that fit a certain BS relates to this topic. The process of locating existing service is called service discovery [KK09]. The service description plays an important role in service discovery. Related work for semantic service description is examined first[2]. There are several semantic service description frameworks. WSMO [RLK06] and OWL-S [Ma04] are the most referenced ones in literature. Both provide a model for the description of Web services with functional and non-functional semantics. The description of Web service might be useful for the ES description. However, since ES are defined as software services, Web services are just a subset of ES. Further, there is no model with a distinct business focus, i.e. a semantic description of BS is not provided. Thus, the utility of WSMO and OWL-S

[2] Traditional service description mechanisms such as WSDL and UDDI are skipped due to their limited expressiveness for service discovery (cf. signature vs. specification matching in [KK04]).

is limited. Concerning service discovery, WSMO provides goals that define expectations for a (sought-after) service in terms of functional and non-functional requirements. Similar to the Web service description, a goal description is based on predefined ontologies. However, the definition of goals might be problematic, since "service requesters are not expected to have the required background to formalize their goals" [Ke05]. In contrast the proposed approach does not define goals (i.e. requirements) explicitly. It is assumed that the BS bears requirements for an ES support implicitly. Based on the semantic service description models there are some research efforts in the area of automatic service discovery (also Semantic Web Service Matching). One of them is [Pa02]. [Pa02] defines service capabilities as functional building blocks of a service and describes them on the basis of DAML-S (predecessor of the OWL-S). They suggest the description of requirements similar to the capabilities. However, it is assumed that requirements and capabilities are described on the basis of a common ontology. This is not realistic in case of BS and ES, since both are described by different stakeholders from different perspectives. Further, it is hard to specify the requirements in detail (cf. [Ke05]). Moreover, the correspondence values are ordinal scaled in [Pa02]. In order to get a ranked list of similarity values an interval scale of measurement is necessary. Furthermore, [Pa02] capabilities contain in- and output parameters only. Thus, [Pa02] match provided in- and outputs with required in- an outputs. This is not realistic since equality between in- and output parameters of services does not guarantee equal functionality (cf. [KK04]). Concerning service matching related work examines the matching of BS-BS and ES-ES [ZW97] but often it does not differentiate between ES and BS explicitly. Furthermore, WSMO-MX [KK09] is a Web service matchmaker which applies different matching filters on semantic Web services based on the language WSML-Rule [Br05]. As a result an ordered set of services including their matching value is generated. However, at the moment it cannot be judged whether WSML-Rule is a proper mechanism for the description of BS and ES.

5 Summary and Future Work

Business services describe business concerns structured as service according the service-orientation paradigm. Electronic services describe technical aspects as services (i.e. according to the service-orientation paradigm) with a special focus on enterprise software applications. Electronic service support operations of business service. There is a large number of business and electronic service available. The evaluation of a certain electronic service that supports a business service is time and cost consuming. This paper provides an approach which supports the evaluation process. Based on the hypothesis that a certain electronic service matches a certain business service if electronic service and business service are similar in their semantics, the approach provides a four step procedure for measuring similarity between business and electronic service. First, semantic entities are extracted from service descriptions, second similarity values are measured for each semantic entity combination, third aggregation is applied in order to gain similarity values on service-level, and fourth results are visualized for further examination, i.e. evaluation and decision making. Future work will address the application of the approach by a scenario of the logistics domain (e.g. transportation service as a BS and a transportation management software as an ES). In course of that,

proper ontology matching techniques and strategies as well as aggregation algorithms are evaluated. Furthermore, a tool will be developed that supports the approach.

Acknowledgment

The work presented in this paper was partly funded by the German Federal Ministry of Education and Research under the projects InterLogGrid (BMBF 01IG09010F) and Logistics Service Bus (BMBF 03IP504).

References

[Br05] Bruijn, J.d.; Lausen, H.; Polleres, A.; Fensel, D.: The WSML Rule Languages for the Semantic Web. Proc.: W3C Workshop on Rule Languages for Interoperability, Washington DC, USA 2005.

[ES07] Euzenat, J.; Shvaiko, P.: Ontology Matching. Springer, Berlin Heidelberg, 2007.

[FGJ97] Fernández-López, M.; Gómez-Pérez, A.; Juristo, N.: METHONTOLOGY: From Ontological Art Towards Ontological Engineering. Proc.: AAAI97 Spring Symposium Series on Ontological Engineering, Stanford University, Stanford, California, USA 1997; pp. 33-40.

[He08] Henneberger, M.: Towards the service-oriented enterprise: innovative approaches for flexible information systems. Sierke, Göttingen, 2008.

[Ke05] Keller, U.; Lara, R.; Lausen, H.; Polleres, A.; Fensel, D.: Automatic Location of Services. In: Lecture Notes in Computer Science (LNCS) - The Semantic Web: Research and Applications, 3532/2005 (2005); pp. 1-16.

[KK04] Klein, M.; König-Ries, B.: Coupled Signature and Specification Matching for Automatic Service Binding. Proc.: European Conference on Web Services (ECOWS 2004), Erfurt, Germany 2004.

[KK09] Klusch, M.; Kaufer, F.: WSMO-MX: A hybrid Semantic Web service matchmaker. In: Web Intelli. and Agent Sys., 7 (2009) 1; pp. 23-42.

[Ma04] Martin, D.; Burstein, M.; Hobbs, J.; Lassila, O.; McDermott, D.; McIlraith, S.; Narayanan, S.; Paolucci, M.; Parsia, B.; Payne, T.; Sirin, E.; Srinivasan, N.; Sycara, K.: W3.org (Publ.): OWL-S: Semantic Markup for Web Services, W3C Member Submission. http://www.w3.org/Submission/OWL-S/, accessed on 2009-05-18.

[Pa02] Paolucci, M.; Kawamura, T.; Payne, T.R.; Sycara, K.P.: Semantic Matching of Web Services Capabilities. Proc.: First International Semantic Web Conference on The Semantic Web, 2002.

[Pa07] Papazoglou, M.P.: Web Services: Principles and Technology. Prentice Hall, Essex, 2007.

[PG03] Papazoglou, M.P.; Georgakopoulos, D.: Service-oriented computing - Introduction. In: Communications of the ACM, 46 (2003) 10; pp. 24-28.

[Pa07] Paruchuri, S.: Toolbox for IT (Publ.): SOA - Common Information Model (CIM), Part 1, 2, 3. http://it.toolbox.com/blogs/the-real-soa/soa-common-information-model-cim-part-1-18292, accessed on 2010/03/29.

[RLK06] Roman, D.; Lausen, H.; Keller, U.: Web Service Modeling Ontology - WSMO Final Draft (D2v1.3). DERI, Innsbruck 2006.

[ZW97] Zaremski, A.M.; Wing, J.M.: Specification matching of software components. Proc.: 3rd ACM SIGSOFT symposium on Foundations of software engineering, Washington, D.C., United States 1995.

Elicitation of Preferences for Web Service Compositions

Carolin Michels and Sudhir Agarwal

Karlsruhe Institute of Technology
{michels,agarwal}@kit.edu

Abstract: Preference elicitation is often used in e-services to create product recommendations for their customers. We present an approach for applying preference elicitation techniques for a semi-automatic query generation system for Web service compositions. These techniques are used for identifying the user's demands on the desired Web service composition and furthermore the trade-offs between interdependent or conflicting preferences. In our iterative process the query is automatically (re-)formulated to present results to the user, so that he can detect undesired characteristics and revise his specifications until he is presented a satisfiable result.

1 Introduction

For years, online shops have provided functionalities to make their customers feel comfortable even though they cannot – as opposed to real world shops – offer advice and support in person.[1] These functionalities are attempts to personalize online shops, like for example a personalized greeting when entering a shop: If a salesperson knows us, he greets us with our name, an online shop does in most cases the same, when we log in. If a salesperson knows, which products we bought in the past, what we like in general or for which purpose we buy at his shop, he can suggest products, in case we are indecisive. These suggestions are not only helpful for the customer, but also for the salesperson as far as he sells more products and creates customer loyalty, which is why many online shops provide product recommendations, too.

In contrast, if a user looks for a Web service composition he is dependent on himself alone. Most Web service composition approaches assume that the user is capable of identifying his demands and expressing them formally in a Web service composition query. For those cases, in which he is not, we want to provide a Web service composition recommendation system similar to a product recommendation system.

There are various reasons why a user may need support for creating a Web service composition query. First of all, he actually may not know how to formulate his request or he may not be capable of identifying his concrete preferences. Also, he may be oblivious of the granularity of possible compositions, of the difference between his goal (*fundamental objectives*) and the *means* he wants to use for achieving his goal (cf. [SPM06]) or even

[1] Except for hotlines and email support, which pose a safety net, if something goes wrong and the user is not able to solve the problem with the online functionalities.

of the information needed to distinguish those services which may be suited best for his needs. Furthermore, dependencies or contradictions between his requirements may not be obvious but crucial and a system pointing them out and/or inquiring could help expressing the actual concrete preferences. Sometimes, the user is actually not aware of the fact that his needs and the possible solutions differ. A system explaining to him why and how a solution deviates from his requirements may be useful.

Secondly, a (successful) Web service composition process requires more information about the user and his requirements and preferences than an atomic Web service discovery (which is of course a part of the aforementioned): not only information about *what* the user wants (and therefore the solution) is required but also about the *how* (cf. [SPM06]). As an advantage, a Web service offers more possibilities to derive information about a user and his requirements and preferences and thus provides more means to help the user. One can think of various steps in the Web service composition process in which a preference elicitation could be helpful: (i) query generation for Web service composition, (ii) presenting the Web service compositions to the users and (iii) revising the proposed solutions. We will consider all these aspects in the process presented in Section 2.

In this paper, we present a process for preference elicitation for Web service compositions. First, in Section 2 we introduce the information used in our approach and the specifications for the preferences which have to be considered and handled for the query generation. We continue with an outline of our approach and the factors used for ranking Web services. In Section 3 we compare former research for preference elicitation. Finally, in Section 4 we give an outlook based on the proposed process model.

2 Request generation and handling

Before we describe the details of our approach, we define some fundamental terms: **Hard constraints** are requirements that must be fulfilled by the composition, i.e. the user would reject every solution that did not cover all given hard constraints. Hard constraints are fulfilled or not, e.g. the constraint “air mail delivery”. In contrast to hard constraints **preferences** need not necessarily be included in the solution, but if a user is confronted with two compositions, one fulfilling a preference to a larger degree than the other, he would select the former. Preferences can be fulfilled partially and weighted by the user (e.g. “I’d rather have a fast delivery than a cheap one”). **Longterm** preferences and hard constraints are conditions that the user wants to be fulfilled every time he composes Web services, e.g. “I always prefer those services by company X”. In contrast **ephemeral** preferences and hard constraints apply only temporarily (due to changed/temporal circumstances) [PFT03]. **Dependencies** can occur between services of a composition or one or more services of the composition and the overall composition (see [WS09]), i.e. those parts of a composition that interact. For example a service may be depending on the output or successful execution of another service. Additionally we regard dependencies between the preferences, if e.g. one preference (e.g. “fast shipment”) is influenced by the fulfillment of another preference(e.g. “cheap shipment”).

The query for Web service compositions is generated by considering three different information sources: (i) The user **profile/history**, e.g. his domain, expertise, former preferences, (ii) the **context** of the user, e.g. field of work, current environment, search context and (iii) the user's **goal definition** and further specifications stated in a form or questionnaire, e.g. input and output parameters and functional properties. While the user profile contains longterm information, i.e. general information that in the majority of the cases holds or aggregated information from previous instantiations, the goal definitions and the questionnaire comprise ephemeral preferences and constraints. The user context can cover longterm properties, e.g. the domain of the user, or properties that change from time to time, e.g. the location context of the user. We use ontologies to represent the user's expertise, his context and the services' inputs, outputs, preconditions and effects (IOPEs, cf. [SPH04]), which should include the terms used in the goal definitions. The inclusion of IOPE ontologies allows furthermore the representation of general dependencies by including a corresponding relation. Consider for example the dependencies "shipping costs depends on mode of shipment" and "total costs depends on shipping costs". Further deductions are possible with reasoning on the ontologies, e.g. if a user has a location context that induces time constraints (e.g. at the airport), he probably wants to have a fast service. Furthermore, problems arising due to the user's expertise, e.g. different wording or lack of knowledge of precise terms become treatable. The ontologies can be used for adapting the questionnaire with different wordings to the user's expertise and making suggestions for relaxing and adjusting the query (see steps 4 and 7 below), e.g. by advising a concept term that is more general or by aggregating preferences.

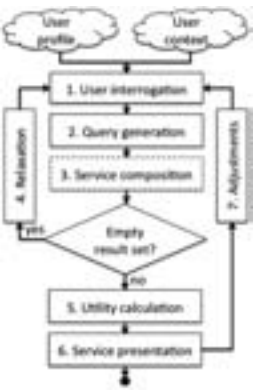


Figure 1: The process starting with the initial information (the user profile and context) ending in the final composition

Figure 1 illustrates our preference elicitation process. It starts with a questionnaire adapted to the user's expertise and context (step 1), which covers the following items: (i) the final goal of the Web service composition (a document, an action or an item), (ii) further specifications depending on the first answer, e.g. the type of the document, (iii) a demand for possible inputs by the user depending on the previous answers, (iv) properties to be regarded (e.g. price, time, quality) and (v) a menu to express the preferences over the chosen properties.

Items 1 to 4 can be realized with interdependent drop-down-menus or free text annotations depending on the user's expertise. Item 5 can be realized by letting the user construct a ranking of the properties or a numeral valuation of the properties. In step 1 not only an ordinal preference is evaluated but also the relation between the preferences. The further use of these results in step 5 is described below.

The questionnaire is used to gather new information, which (in combination with the previous information) is used for the query generation (step 2). The generated query is passed on to a service composition component (step 3) using only the hard constraints and the

goal to find matching service compositions. The preferences are used for the ranking in step 6. In this paper, we consider a scenario with an implemented Web service composition component. This component is used for the first and for further iterations in the preference elicitation process. If the service composition returns an empty set, i.e. no services are found that match the query, the process repeats the information elicitation process with the aim of relaxing some requirements and preferences (step 4) like e.g. described in [Jan04]. To achieve this, the user may revise his former questionnaire or get proposals for relaxed preferences that he can choose from. Those proposals are computed in a way that they guarantee a non-empty result set if the user accepts them.

If at least one service is found, the utility of each Web service composition is calculated in step 5 based upon the information gathered in step 1. We propose two different approaches for the utility calculation: (i) using an actual utility function of the user or (ii) using fuzzy sets for calculating the overall acceptance (which is comparable to the utility of the service) like done in [ALS09]. For the former, a utility function u for the different attribute values v_{ij} of the according attribute a_i is recorded in the user profile and complemented in the preference elicitation process. Like in [TGL04] we consider a weighting for each attribute.The overall utility of one specific Web service composition c_j is calculated by summarizing all utility values of the Web service composition (cf. [ALS09]) adjusted with the weighting w_i of the respective attribute a_i: $U(c_j) = \sum_i w_i u(v_{ij})$.

If fuzzy sets are applied like e.g. in [AL05], the user has to specify categories for each attribute or choose according pre-defined categories for the attributes.[2] In [AL05] it was assumed, that the user could express his needs as fuzzy IF-THEN rules. When using fuzzy sets we cover the generation of these fuzzy IF-THEN rules and therefore the elicitation of the fuzzy sets, the (fuzzy) membership function of each attribute value to the according attribute category (both in step 1) and the calculation of the overall acceptance of a Web service composition (step 5).

The resulting values of step 5 may be used in the ranking of the results to be presented in step 6 where the composition with the highest utility value/acceptance is ranked first. The user history and context is used in this step, again, to adjust the result representation and explanations for the services. If the user is not satisfied with the result set presented to him, he may adjust or revise his previous statements (step 7). In this case a new questionnaire is presented depending on his desired changes or he may revise the previous questionnaire.

3 Related Work

Much research work (e.g. [ABK+02, CP04]) dealt with automated eliciting preferences for product recommendations. Some approaches (e.g. [SPM06]) coped with applying preference customizing techniques for Web services and Web service discovery, but to our knowledge none actually used preference elicitation techniques, even though they bear many advantages to support Web service composition.

[2] Like [AL05] we consider only attribute values "which can be mapped to an interval scale". Nominal attributes, that are either fulfilled or not like e.g. a sepecific mode of shipping, need not be expressed fuzzy.

In [JK05] an application for a user preference elicitation system for e-services is described. In contrast to our approach the focus is not on direct service recommendations but on product recommendations in e-services . Nonetheless we adapt their idea to relax requirements if the result set is empty and their questionnaire adaption: The questions, answers and explanations are personalized in content, language and number according to the user's expertise. The defaults are set according to the user's expertise and previous answers. The final representation of the results and the explanation is adapted to his expertise.

In [PFT03] different aspects for decision support systems based on findings from behavior decision theory are stated. The authors claim that even though users are not capable of stating "other preferences beyond the very fundamental ones ... it is easier for them to critique examples, especially those showing violations" [PFT03]. We therefore conclude, that a revision of the preferences after a first result representation is crucial, so that the user can specify why some of the services presented to him are not fulfilling his needs.

The user model in [TGL04] covers user preferences that are derived by his tendencies, which are not solely indicated by his direct answers, but also by his decisions and a frequent provision of certain attributes or attribute values. It stores, for each item, the ratio of its acceptance to the number of times it was recommended to the user. In addition, with the user model the weighting for all attributes and the probability of the associated attribute values are stored. For initialization, the weighting is set according to the typical preferences in the use-case, e.g. in the case of a restaurant recommendation system the *price* has more importance and therefore a higher weighting than *parking*. The probabilities for the different attribute values start with a uniform distribution, while the accept/present-ratio begins with a large value, e.g. 90%. We anticipate to consider context information not only for the on-the-fly calculation but for the initial values as well.

In contrast to our approach, in [LHL97] the user preferences are handled as a set of constraints, which can also be weighted differently. The sum over all constraints C_i multiplied with their corresponding weights w_i results in the overall *error* of a proposed solution: $E((v_1, v_2, ..., v_n)) = \sum_{i=1}^{n} C_i(v_i) * w_i$ [LHL97].

4 Conclusion and Outlook

We presented a process model to elicit user information for Web service composition. Our approach uses information stored in a user profile as well as information derived from the user's context. Furthermore, preferences and hard constraints are surveyed via a questionnaire. Using a Web service composition component, all Web service compositions that satisfy the hard constraints build the result set. The collected information about the user's preferences allows implications on the user's utility function, which can be used to rank and present the services.

Future work will include concrete implementation details and evaluations. These will especially focus on the preference elicitation process itself, i.e. means for building a questionnaire to derive information about the preferences, the preference rankings, the accord-

ing utility function and the implementation of the query generation.

Acknowledgements

This work is supported by the German Federal Ministry of Education and Research (BMBF) under the WisNetGrid-Project. We thank Sebastian Rudolph (AIFB, KIT) and René Jäkel (ZIH, TU Dresden) for their help.

References

[ABK+02] J.L. Ambite, G. Barish, C.A. Knoblock, M. Muslea, J. Oh, and S. Minton. Getting from here to there: interactive planning and agent execution for optimizing travel. In *18th national conference on Artificial intelligence*, pages 862–869. AAAI Press, 2002.

[AL05] S. Agarwal and S. Lamparter. SMART – A Semantic Matchmaking Portal for Electronic Markets. *Seventh IEEE International Conference on E-Commerce Technology (CEC'05)*, pages 405–408, 2005.

[ALS09] S. Agarwal, S. Lamparter, and R. Studer. Making Web services tradable: A policy-based approach for specifying preferences on Web service properties. *Web Semantics: Science, Services and Agents on the World Wide Web*, 7(1):11–20, 2009.

[CP04] L. Chen and P. Pu. Survey of preference elicitation methods. Technical report, Swiss Federal Institute of Technology in Lausanne, Technical Report No. IC/200467, 2004.

[Jan04] D. Jannach. Preference-based Treatment of Empty Result Sets in Product Finders and Knowledge-based Recommenders. In *Poster Proceedings of the 27th Annual German Conference on Artificial Intelligence, KI 2004*, pages 145–159, 2004.

[JK05] D. Jannach and G. Kreutler. Personalized User Preference Elicitation for e-Service. In *IEEE International conference on e-Technology, e-Commerce and e-Service, Hong Kong, IEEE Computer Society*, pages 604–611, 2005.

[LHL97] G. Linden, S. Hanks, and N. Lesh. Interactive assessment of user preference models: The automated travel assistant. *Courses and Lectures – International Centre for Mechanical Sciences*, pages 67–78, 1997.

[PFT03] P. Pu, B. Faltings, and M. Torrens. User-involved preference elicitation. In *IJCAI Workshop on Configuration*, pages 56–63. Citeseer, 2003.

[SPH04] E. Sirin, B. Parsia, and J. Hendler. Composition-driven Filtering and Selection of Semantic Web Services. In *In AAAI Spring Symposium on Semantic Web Services*, 2004.

[SPM06] S. Sohrabi, N. Prokoshyna, and S.A. McIlraith. Web service composition via generic procedures and customizing user preferences. *LNCS*, 4273:597–611, 2006.

[TGL04] C.A. Thompson, M.H. Göker, and P. Langley. A personalized system for conversational recommendations. *Journal of Artificial Intelligence Research*, 21(1):393–428, 2004.

[WS09] M. Winkler and A. Schill. Towards Dependency Management in Service Compositions. In *Proceedings of the International Conference on e-Business, Milan*, 2009.

Query Languages for Semantic Web Services

Stefan Schulte, Melanie Siebenhaar, Julian Eckert, Ralf Steinmetz

schulte@kom.tu-darmstadt.de

Abstract: Even though service discovery is one of the vital steps in Web service invocation, service requests are usually expressed by rather simple means. While syntax-based service descriptions are usually addressed using keyword-based queries, Semantic Web Services are queried using "perfect" service descriptions. Hence, a service requester is not able to control the outcome of the discovery process in-depth by specifying similarity ranges or to explicitly state which syntax- and semantic-based parts of a service description should be addressed.

In this paper, we present an analysis of current approaches to service request formulation and identify requirements a query formalism for Semantic Web Services should fulfill. Furthermore, we give a brief preview on SWS2QL, a SPARQL-based query language for Semantic Web Services.

1 Introduction

Today, Semantic Web Services (SWS) are a prominent field of research and have resulted in a number of different approaches and standards such as *OWL-S*, *WSMO*, or *SAWSDL*, i.e., formalisms which explicitly make use of semantics in different parts of a service description.

One of the primary application areas of SWS is service discovery, which is essentially affected by three aspects: (i) The ability of service providers to describe their services, (ii) the ability of requesters to describe their requirements towards services, and (iii) the effectiveness of the service matchmaker, i.e., an algorithm that takes into account a request and finds the best fitting services from a set of service offers. The discovery of services is one of the core success factors for the invocation of services and is hence deemed one of the grand challenges in Web service research [PvdH07].

Service discovery research focuses mostly on matchmaking [Klu08], and the integration of SWS into service registries (e.g., [KP08, PKPS02]), two areas which have been examined from the very beginning of SWS research. In contrast, surprisingly little effort has been put into the deployment of alternatives to keyword- and service description-based query formulation. In the following, we will examine different ways to express a service request and identify their individual shortcomings (cp. Section 2). Based on this analysis, we specify requirements a service query formalism should fulfill (cp. Section 3). The paper closes with a description of SWS2QL, a query language based on SPARQL we developed in order to meet these requirements, and a conclusion.

2 Query Formulation for Semantic Web Services

In the following, we will give a short overview of different ways to express requirements towards a service:

Keyword-based service requests are the typical query formulation for syntactic Web services; this includes wildcard- and table-based queries. Keyword-based matchmaking is based on standard Information Retrieval (IR) methods, e.g., term matching: The terms from the request are matched with keywords found in a service advertisement. Explicit semantic information is not regarded.

Service description-based requests are widely adopted in research approaches which primarily aim on the deployment of new matchmaking algorithms (e.g., [KKZ09, PP09]). The basic idea is that the service request is formulated as a model instance, e.g., a WSDL-based service description. This model instance is deemed to be a "perfect" service, i.e., the perfect answer to the service request. While this is a very rich request format which can incorporate both syntactic and semantic information as well as the relationships between the single service components, there are a couple of shortcomings elaborated on in the following.

Query language-based service requests make use of a structured query syntax, e.g., a query language like SQL or SPARQL. Query language-based service requests do not directly represent a service as a model instance. Instead, parts of a virtual service model instance are *described* in terms of a well-defined query language. Service models provide the information and relationships required for creating a query language-based request. Query-based service requests can refer to both syntactic and semantic information.

Today, Web service portals like `www.seekda.com` primarily rely on keyword-based search which is not surprising as the listed services are mostly syntactically described [LS08]. The same applies to state-of-the-art service registry standards like UDDI and ebXML. As a result of the shortcomings of keyword-based retrieval, several approaches have already integrated semantic technologies into service registries, primarily UDDI (e.g., [KP08, PKPS02]). The most common approaches are built on top of syntactic Web service descriptions, require different execution environments and do not make use of potentially already existing (syntax-based) query formats. Alternatively, it is often possible to describe a service model instance, i.e., to follow the service description-based approach. However, this approach possesses several drawbacks:

Integration of syntactic search: Service descriptions do not allow for the explicit definition of syntactic information. It might be the case that a matchmaking algorithm takes syntactic information into account (e.g., [KKZ09, PP09]). However, this cannot be directly controlled by the requester if using a service description-based query.

Missing range: A service description does not recognize the definition of ranges for complete services and single elements. Thus, it is not possible to define, e.g., that only

those services which *exactly* meet the request should be returned by a matchmaker. Again, a matchmaking algorithm could define that only certain ranges are included in the result set of a query. However, it is not possible to set according parameters in the service query if using the service description-based approach.

Incompatible query formats: If a service request is formulated in a certain Web service formalism, it is not possible to discover service descriptions following another Web service standard. This forces users to formulate different queries if a service registry is able to manage more than one Web service formalism.

Lack of knowledge: Service requesters are not necessarily familiar with one or more service standards. However, requesters might be familiar with a query formalism which is, e.g., already used in a particular infrastructure or domain.

Usability: Neither the fact that a customization or parameterization of queries has to be done at the matchmaker-level nor the need to formulate queries for different Web service standards in different formats is very user-friendly. It is quite likely that the usage of semantic information in service requests will only be accepted by users, if it is possible to standardize Web service queries.

3 Specification of Requirements

In order to search for services based on semantic information, the requester needs to be able to define semantics in an intuitive and, if possible, uniform way, i.e., the request formulation should be independent from the service formalism(s) supported by the registry. Based on the shortcomings of the keyword- and service description-based approaches highlighted in the last section, the following generic requirements towards a query language for SWS have been identified:

Combination of syntactic and semantic search: A query language should integrate semantic as well as syntactic information. Furthermore, it should be possible to combine syntactic and semantic information within a single query.

Reuse: If a query language is already common to potential users, the chance to get it accepted will be likely higher. If multiple registry standards have to be supported, an appropriate mapping of the global query language to the specific registry query syntax should be provided.

Ranges: It should be possible to define a threshold of similarity a service offer needs to fulfill in order to be included in the result set.

As it was mentioned above, some of these requirements are currently reckoned in matchmaking algorithms, e.g., the usage of syntactic data or the definition of ranges or similarity thresholds. However, the actual handling of syntactic data or ranges in matchmaking is usually not explicitly regarded but used as a complement for semantic-based matchmaking. Hence, the usage of syntactic data and thresholds is not visible to the service requester

and cannot be directly controlled. In fact, such a control would make it necessary to offer the requester different setting properties of a matchmaking algorithm, which need to be manipulated apart from the actual query. The definition of further search constraints within a query is a much more intuitive approach.

Apart from the more generic requirements mentioned above, there are some requirements, which aim on the independence from concrete technologies and need to be regarded, too:

Flexible handling of different matchmakers: It should be possible to "manually" choose a matchmaker. If this is not done, the query processor needs to recognize appropriate matchmakers based on the actual query content.

Uniform query syntax: The query syntax should be the same for all Web service standards addressed by the corresponding service registry.

Handling of different registry standards: A uniform query syntax should be applicable to different registry standards.

4 SWS2QL – A Query Language for Semantic Web Services Based on SPARQL

Taking the requirements defined above into account, we have conceptualized and implemented the *SWS Structured Query Language* (SWS2QL). SWS2QL enhances SPARQL by new features aiming at SWS discovery. This is done based on the requirements described in Section 3. In addition, an abstract data model for services has been created. The actual query language, SWS2QL, is an enhancement of SPARQL and refers to this abstract model. The major benefit of making use of an abstract service model is the applicability of SWS2QL to different service standards which share common features with the abstract model. In the work at hand, these features are inputs, outputs, operations, interfaces, preconditions, and effects, which make it possible to map OWL-S- and SAWSDL-based services to the model and therefore apply SWS2QL to these SWS formalisms. As SAWSDL does not define preconditions and effects, we make use of *WSMO-Lite* in order to add these constructs to SAWSDL [VKVF08].

SPARQL has been selected as foundation for SWS2QL for several reasons. First of all, it is the major query language of the Semantic Web [KBM08]. Second, within SPARQL, query statements are expressed in terms of triples made up from a subject, object, and predicate. Thus, SPARQL represents a very intuitive approach, since the query statements are very close to shortened natural language-based expressions. Several enhancements of SPARQL for specific purposes have already been proposed as presented by [BFL$^+$09], showing the flexibility and adaptability of SPARQL towards different application areas. Although SPARQL is intended to be executed against RDF-based data, it may also be applied as a query *description* language only, as it is done in SWS2QL.

SWS2QL is compatible with the existing SPARQL grammar, makes use of existing extension principles, and minimizes the amount and complexity of further rules. SWS2QL

allows a service requester to address different service abstraction levels, incorporate and parameterize matchmakers, define similarity thresholds, etc. Due to the abstract service model, SWS2QL may be easily transferred to different service description and registry standards (through a mapping).

5 Conclusion

If comparing the six requirements defined in Section 3 with the work presented in this section, we can make the following statements: First of all, SWS2QL allows to combine syntactic and semantic search. Second, SWS2QL reuses and enhances an already known query syntax. Third, ranges can be defined by the definition of thresholds. Fourth, different matchmakers may be applied using SWS2QL as long as they implement an according interface. Fifth, SWS2QL is not restricted to one particular SWS formalism as it makes use of an abstract service model. Last, SWS2QL might be integrated into different service registry standards – however, a first proof of concept implementation has been done using ebXML.

Finally, we want to compare SWS2QL with the most relevant related work: Even though there are some approaches to establish query languages for SWS, they lack particular attributes regarding the requirements defined in the last section. To start with, Iqbal et al. present the usage of SPARQL ASK and CONSTRUCT queries to retrieve SAWSDL-based services [ISPG08]. More precisely, user goals are specified in the form of SPARQL ASK queries, which have to be fulfilled by a matching service, and the SPARQL CONSTRUCT query form is used for the representation of a service result including its pre- and post-conditions. Basically, this approach focuses on the definition of pre- and postconditions and sophisticated functionalities as, e.g., similarity ranges are not regarded. Kiefer and Bernstein also make use of SPARQL queries as SWS query language [KB08]. In their solution, *iSPARQL*, the authors propose the integration of imprecise matching capabilities into SPARQL. As the approach aims on imprecise matching, matchmaking is conducted using IR-based similarity values, which are applied to OWL-S service attributes (e.g., name, description) or sets of concept names resulting from the unfolded input and output concepts of a given service. In general, the approach proposed by Pantazoglou and Tsalgatidou, *USQL*, is closest to the work at hand [PT09]. Even though USQL provides a rich query model which allows to define syntax- and semantic-based query statements, thresholds, and may be applied to different service registry standards, it is only applicable to the proprietary PS-WSDL format, which is a major drawback of this approach.

In summary, SWS2QL represents a unified query language for SWS, which is not bound to specific service description formalisms, registry standards, or matchmakers like other service retrieval approaches. Concerning the query facilities, a holistic analysis of the requirements for a unified querying language for SWS has been performed. In doing so, the resulting query language SWS2QL is capable to overcome the shortcomings of related service retrieval approaches, while utilizing and enhancing the major query format of the Semantic Web, namely SPARQL, to facilitate user acceptance.

References

[BFL+09] François Bry, Tim Furche, Benedikt Linse, Alexander Pohl, Antonius Weinzierl, and Olga Yestekhina. Four Lessons in Versatility or How Query Languages Adapt to the Web. In *Semantic Techniques for the Web, The REWERSE Perspective*, volume 5500 of *Lecture Notes in Computer Science*, chapter 2, pages 50–160. Springer, 2009.

[ISPG08] Kashif Iqbal, Marco Luca Sbodio, Vassilios Peristeras, and Giovanni Giuliani. Semantic Service Discovery using SAWSDL and SPARQL. In *Fourth International Conference on Semantics, Knowledge and Grid (SKG 2008)*, pages 205–212. IEEE Computer Society, 2008.

[KB08] Christoph Kiefer and Abraham Bernstein. The Creation and Evaluation of iSPARQL Strategies for Matchmaking. In *5th European Semantic Web Conference (ESWC 2008)*, volume 5021 of *Lecture Notes in Computer Science*, pages 463–477. Springer, 2008.

[KBM08] Vipul Kashyap, Christoph Bussler, and Matthew Moran. *The Semantic Web, Semantics for Data and Services on the Web*. Springer-Verlag, Berlin, Heidelberg, 2008.

[KKZ09] Matthias Klusch, Patrick Kapahnke, and Ingo Zinnikus. SAWSDL-MX2: A Machine-Learning Approach for Integrating Semantic Web Service Matchmaking Variants. In *IEEE 7th International Conference on Web Services (ICWS 2009)*, pages 335–342. IEEE Computer Society, 2009.

[Klu08] Matthias Klusch. Semantic Web Service Coordination. In *CASCOM: Intelligent Service Coordination in the Semantic Web*, chapter 4, pages 59–104. Birkhäuser Verlag, 2008.

[KP08] Dimitrios Kourtesis and Iraklis Paraskakis. Combining SAWSDL, OWL-DL and UDDI for Semantically Enhanced Web Service Discovery. In *5th European Semantic Web Conference (ESWC 2008)*, volume 5021 of *Lecture Notes in Computer Science*, pages 614–628. Springer, 2008.

[LS08] Holger Lausen and Nathalie Steinmetz. Survey of Current Means to Discover Web Services. Technical report, STI Innsbruck, Austria, August 2008.

[PKPS02] Massimo Paolucci, Takahiro Kawamura, Terry R. Payne, and Katia P. Sycara. Importing the Semantic Web in UDDI. In *International Workshop on Web Services, E-Business, and the Semantic Web (WES 2002)*, volume 2512 of *Lecture Notes in Computer Science*, pages 225–236. Springer, 2002.

[PP09] Pierluigi Plebani and Barbara Pernici. URBE: Web Service Retrieval Based on Similarity Evaluation. *IEEE Transactions on Knowledge and Data Engineering*, 21(11):1629–1642, 2009.

[PT09] Michael Pantazoglou and Aphrodite Tsalgatidou. The Unified Service Query Language. Technical report, National and Kapodistrian University of Athens, Greece, 2009.

[PvdH07] Mike P. Papazoglou and Willem-Jan van den Heuvel. Service oriented architectures: approaches, technologies and research issues. *The VLDB Journal*, 16(3):389–415, 2007.

[VKVF08] Tomas Vitvar, Jacek Kopecký, Jana Viskova, and Dieter Fensel. WSMO-Lite Annotations for Web Services. In *5th European Semantic Web Conference (ESWC 2008)*, volume 5021 of *Lecture Notes in Computer Science*, pages 674–689. Springer, 2008.

Business-oriented CAx Integration with Semantic Technologies Revisited

Claudia Hess[1], Willy Chen[1], Thomas Syldatke[2]
[1]SysTec-CAx GmbH, München, {claudia.hess,willy.chen}@systec-cax.de
[2]AUDI AG, Ingolstadt, thomas.syldatke@audi.de

Abstract: Product development is one of the core business processes for car manufacturers. Today it is heavily support by information systems enabling computer-aided technologies for design (CAD), engineering (CAE), and testing (CAT), which we refer to as CAx systems. In order to improve the alignment between business and IT we have created an enterprise architecture for the research and development department at AUDI AG. The description of the various relevant architecture domains is thereby based on ontologies. A web-based tool enables collaborative editing of these contents as well as complex analysis on various aspects relevant to IT management.

1 Introduction

Core challenges for today's automotive industry are manifold. Market requirements are constantly changing, e. g. due to new regulations or customer demands. The product complexity increases, especially due to an increasing number of electronics. Moreover, the model range grows – 50% more models can be expected in the next years – while the development cycle time is reduced. Research and development departments (R&D) of automotive manufacturers nowadays have a number of specialized IT applications in use. They implement especially the different computer-aided methodologies, such as design (CAD), engineering (CAE) and testing (CAT). These specialized systems are highly dependent and constantly evolve in order to better adapt to the changing business requirements. We have created an enterprise architecture for the R&D department at AUDI AG to improve the alignment of business and IT and to manage the complex dependencies within the heterogeneous IT landscape.

At AST 2008 [CSH08], we presented our intended approach for integrating CAx methodologies and IT systems from a business-oriented view. We aimed to use semantic technologies to represent, formalize, and interconnect the relevant knowledge. Implementing this approach, we focused on supporting the development of the CAx systems with a business focus. This paper shows how we built up an enterprise architecture using semantic technologies. To this end, it is organized as follows. Section 2 introduces the use case. Section 3 describes the ontologies that we developed. In section 4, we present our approach to ontology-aided IT management and the web-based application that is used to access the knowledge and which allows for easily updating the ontology in a collaborative manner. Finally, section 5 summarizes the approach and lessons learned.

2 Use Case

Computer-aided methodologies support the different steps in the product development process. They are especially used for the construction of new parts and prototypes, their validation and their release for the serial production. Using as much as possible virtual validation methodologies and combining them in an appropriate way with physical validation methodologies allows for optimizing the product development process. Computer-aided design (CAD), engineering (CAE) and testing (CAT), referred to as CAx, play thus a major role. They are supported by a number of specialized IT systems ranging from standard products, such as in CAE the tools for pre-processing, solving and post-processing, to individual software that is highly tailored to AUDI's needs. As tools may require the output of other tools as input, there are complex dependencies between the different information systems. The exchange of information between them is not straightforward because product models, data structures and technical platforms differ. Managing the dependencies and bridging the gaps between CAx systems is thus a major task for optimizing their interplay and, in the end, that between the CAx methodologies.

As business architects, we analyze project proposals for new CAx systems as well as for extensions of existing ones. We have to ensure their integration into the IT landscape and their fit with business requirements. Moreover, managing the IT landscape requires to identify issues in the existing infrastructure and to initiate change projects. These may improve the alignment of business and IT or increase the re-usability and flexibility. Methods and frameworks from Enterprise Architecture (EA) provide here a valuable starting point. An EA describes the basic artifacts of an enterprise and their dependencies. These artifacts range from business aspects, such as strategic aspects and organizational structures, to IT-related ones, such as software components and the IT infrastructure. The EA allows for the analysis of the enterprise's current IT infrastructure and the strategic planning of its future state. The Enterprise Architecture Management Method Library defines Enterprise Architecture Management (EAM) as "a continuous and self maintaining management function seeking to improve the alignment of business and IT in an (virtual) enterprise" [EAM].

Different frameworks such as the Zachman Framework [Zac87], TOGAF [The09] and the ARIS framework [Sch99] have been proposed for modeling an EA. Creating an EA for the R&D department at AUDI AG, we followed the methodology suggested by TOGAF. TOGAF, The Open Group Enterprise Architecture Framework, is an industry standard architecture framework that can be freely used to develop an EA. The core of TOGAF is its generic enterprise architecture method, the Architecture Development Method (ADM), which specifies the different steps required for building up an EA. Guidelines and frameworks complete the toolkit. TOGAF comprises the following enterprise architecture domains: The *business architecture* comprises the business strategy, governance, organization, and key business processes. Then, the *application architecture* provides a blueprint for the individual application systems. It contains the interactions between the systems and their relationships to business processes. Then, the *data architecture* structures the organization's logical and physical data assets and the associated data management resources. The *technical architecture* describes the hardware, software and network infrastructure.

3 Architecture Domain Ontologies

Following the TOGAF architecture methodology, we started developing ontologies for the four architecture domains. In a whole, they provide the essential information for the management of the different CAx systems in use. The ontologies for the application and the data architecture domain are presented in the following (see also figure 1). We show the main classes for each domain. In practice, they are refined by subclasses, attributes and relationships.

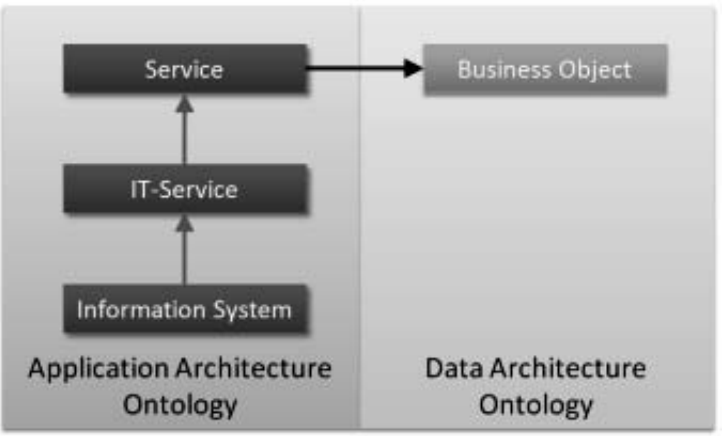


Figure 1: Top Level Classes

3.1 Application Architecture Ontology

The application ontology defines the structure for describing the software systems that are in use in AUDI's R&D department in a business-oriented way. Moreover, it provides a starting point for getting more detailed information on the actual implementations. Following the ideas of a service oriented architecture (SOA), an information system is described in terms of atomic IT-services and business services. Services are independent of other services and are highly reusable. They can be composed to more complex services. A service is self-contained and does not expose any internals to the service consumers.

The Application Architecture Ontology was inspired by the Service-Oriented Architecture Ontology [The08] (SOA Ontology) developed by The Open Group. This ontology aims at describing a service-oriented architecture with the help of an ontology. The SOA Ontology is currently available as a draft technical standard. It is provided as a OWL ontology. It defines classes for describing a SOA and relationships between these classes. The SOA Ontology contains both business and technical aspects. Moreover, it defines classes related to the development and management of a SOA.

Central classes and relationships of our ontology are:

- Business Service: A business service is, following the definition in the SOA ontology "a logical representation of a repeatable business activity that has a specified outcome". It is described from the domain user's point of view. Business services are on the one hand data services for creating and accessing business objects (e.g. create order, modify order) and on the other hand logic services which realize more complex business logics and rules by using one or more distinct business objects (e.g. check availability).
- IT Service: An IT service represents a self-contained functionality that corresponds to a business service. An interface describes its semantics. An IT service is ideally a Web Service. It may also be an EJB. However, in order to be able to represent the functionality by legacy systems, other IT service types may be considered such as DB views or Stored Procedures. We described IT services in a way that potential users of this IT service have enough information to use it. For example, a Web Service has to provide a link to its WSDL. Dependencies between different IT services are captured in the application architecture ontology, too: an IT service uses another IT service.
- Information system: An information system provides IT services. Systems thereby range from individual software to standard software and customized standard software. Standard software is ready-made software and is used as it is by the engineers in AUDI's R&D department. Individual software, in contrast, is specially developed to meet particular requirements. Customized standard software is such standard software that is tailored to fit AUDI's needs.

3.2 Data Architecture Ontology

The central class of our data architecture ontology is the business object. A business object is a description of a business entity that is independent of its implementation. It is important to define a business object precisely in order to ensure a common understanding. A business object can be described in detail by its attributes. For example, we have a business object *order* with attributes like an *order number*. Business objects may be structured hierarchically with general ones on the highest level, and more specific ones on the detailed levels. For instance, a business object *resource* may be further differentiated in resources required for performing a certain validation, such as a *crash test dummy* for performing a crash test. Business objects inherit the attributes of the hierarchically higher ones. For example, a specialized order for building up a prototype, also has an order number, inherited from the more general business object *order*. A further possibility to structure the data architecture are references between business objects. That means that dependencies between business objects are explicitly represented. An example for such a reference is that a client *creates* an order.

4 Ontology-aided IT-Management

In the previous section we presented ontologies which describe the application and the data architecture domain. These descriptions are adopted to the needs of AUDI and are used within different project stages that are relevant for governance in service-oriented architectures (cf. [Bis08]):

- Pre-project stage: During the initial project proposal and funding, the project initiator and an enterprise architect utilize the enterprise architecture to clearly define the scope of the envisioned project in terms of business objects and services that should be implemented as well as IT services to be reused. If the project deals with new business objects or requires more specialized services, the enterprise architect adds these information to the appropriate domain ontology.
- Project stage: The result of the project stage is, among others, a standardized IT specification based on UML 2.0, which covers all relevant aspects for the application architecture domain. Based on the XML metadata interchange (XMI) format, these information are synchronized with the application architecture ontology to reflect the most recent information about a specific IT system.
- Run-time: Enterprise architects are constantly monitoring the different architecture domains in order to identify optimization needs and if necessary trigger change projects. IT architects can analyze dependencies, e. g. which services and systems are affected by a modification of a certain service or have to be informed before the release of a new version.

Users at AUDI, such as enterprise and IT architects, project and requirement managers, can access the information from the application and data architecture ontology via a web-based application that provides means to search and visualize information but also edit and correct them like in a WIKI system. While the core information are represented as ontology facts, users can link documents to any of these ontology individuals to provide additional information about a topic. A semantic search engine enables the search for terms not only in the ontologies but also in all documents that are linked to ontology elements. A ranked result list is presented to the user. Users can generate reports for any element in the ontologies and use a graph visualization to analyze all known dependencies to other elements in the information model.

Most importantly, users can edit any A-Box element in a collaborative way using a wiki-style editor including features for showing the change history and comparing specific versions of an individual as well as simple rating mechanisms. The editor does not require any knowledge on ontological modeling so that business users can easily contribute information and add comments on existing facts. The described web-based tool has been set productive early 2009. Currently, the application has about 80 registered users, of which 25 are actively using and contributing to the knowledge base. The T-Box of all ontologies contains 80 classes, 104 object properties, and 22 data properties while the A-Box has about 25000 individuals and assertions.

5 Conclusion and Outlook

In this paper we have presented an ontology-aided approach to support the IT management in the research and development department at AUDI. The implementation of these ontologies is based on technologies that are currently being developed for realizing the vision of a Semantic Web. We use the Web Ontology Language (OWL) as ontology modeling language and provide a web-based application where users can edit the assertional boxes of the used ontologies without any detailed knowledge about ontology modeling. Most importantly, the use and update of the presented models is embedded into the IT management process, which ensures that the information are up to date and actively used. This is crucial for a high information quality.

While we are convinced that the architecture domain ontologies have to be defined for each enterprise separately, a common application architecture ontology e.g. based on the SOA Ontology by The Open Group could be an ideal starting point for enabling ontology-aided IT-management in other industrial environments. Depending on the available information about information systems, however, the degree of detailing needs to be adaptable.

We are currently working on addressing additional aspects of enterprise architecture management such as IT planning or business process alignment. Also, we are starting to implement IT governance guidelines using logical rules as provided by the Semantic Web technology stack.

References

[Bis08] Todd Biske. *SOA Governance*. Packt Publishing, 2008.

[CSH08] Willy Chen, Thomas Syldatke, and Claudia Hess. Business-oriented CAx-Integration with Semantic Technologies. In *Proceedings of Informatik 2008, Applications of Semantic Technologies 2008 (3rd International AST Workshop)*, 2008.

[EAM] EAMML. Enterprise Architecture Management Method Library (version 0.2). `http://wwwmatthes.in.tum.de/wikis/sebis/eamml`, last accessed January 19, 2010.

[Sch99] August Wilhelm Scheer. *ARIS – Business Process Frameworks*. Springer, 3 edition, 1999.

[The08] The Open Group. SOA Ontology Draft 2.0. `http://www.opengroup.org/projects/soa-ontology/uploads/40/16940/soa-ontology-200-draft.pdf`, last accessed January 19, 2010, 2008.

[The09] The Open Group. *TOGAF Version 9, Enterprise Edition*. 2009.

[Zac87] J.A. Zachman. A framework for information systems architecture. *IBM SYSTEMS JOURNAL*, 26(3), 1987.

Konrad-Zuse-Workshop

Rechnender Raum und Zellulare Automaten

Gemeinsame Fachgruppe

Physik, Informatik und Informationstechnik (PII)

der GI, ITG und DPG

Konrad-Zuse-Workshop

Rechnender Raum und Zellulare Automaten

Gemeinsame Fachgruppe
Physik, Informatik und Informationstechnik (PII)
der GI, ITG und DPG

Vorwort

Prof. Dr. Bernd Klauer
Helmut-Schmidt-Universität
Universität der Bundeswehr Hamburg
bernd.klauer@hsu-hh.de

Die Erfindung der ersten frei programmierbaren Rechenmaschine mit Gleitpunktarithmetik und Anwendung des dualen Zahlensystems ist heute fest mit dem Namen Konrad Zuse verbunden. Neben dieser Erfindung, die unzweifelhaft zu den fundamentalen Erfindungen gehört, die die Menschheit bis zu ihrem Ende begleiten werden, gilt Konrad Zuse aber auch als Urheber wichtiger wissenschaftlicher Erkenntnisse im Umfeld des Computers. Dies sind der Plankalkül, der als Vorstufe der modernen Programmiersprachen gesehen wird und der Rechnende Raum, der den Themeschwerpunkt dieses Workshops darstellt.

Die Fachgruppe PII fühlt sich Konrad Zuse in dieser Thematik besonders verbunden. Am 21. und 22. November 1991 fand in München das *Arbeitsgespräch Physik und Informatik — Informatik und Physik* statt. Auf dieser Veranstaltung, die die Fachgruppe PII als ihre eigentliche Gründungsveranstaltung ansieht, hielt Konrad Zuse den Übersichtsvortrag mit dem Titel *Rechnender Raum*, in dem er selbst sein Modell des Rechnenden Raums mit den Zellularen Automaten vergleicht. Die Fachgruppe nimmt dies zum Anlass den Rechnenden Raum als fundamentalen wissenschaftlichen Beitrag neben der grundlegenden Erfindung des Computers in den Fokus eines Workshops zu Ehren Konrad Zuses im Jahr seines 100. Geburtstages zu stellen.

Bernd Klauer, Hamburg, im Juli 2010

Der Rechnende Raum von Konrad Zuse

Horst Zuse

Es geschah bei dem Gedanken der Kausalität, dass mir plötzlich der Gedanke auftauchte, den Kosmos als eine gigantische Rechenmaschine aufzufassen. Ich dachte dabei an die Relaisrechner: Relaisrechner enthalten Relaisketten. Stößt man ein Relais an, so pflanzt sich dieser Impuls durch die ganze Kette fort. So müßte sich auch ein Lichtquant fortpflanzen, ging es mir durch den Kopf. Der Gedanke setzte sich fest; ich habe ihn im Laufe der Jahre zur Idee des "Rechnenden Raumes" ausgebaut. Es sollte freilich dreißig Jahre dauern, ehe mir eine erste konkrete Formulierung der Idee gelang.

Aus: Konrad Zuse, Der Computer — mein Lebenswerk

Konrad Zuse berichtet in seinem Buch aus einer Zeit um 1945 als er seine ersten Gedanken über den *Rechnenden Raum* skizziert. Physik und Informatik begegenen sich hier in den Bereichen der Quantenphysik und der Automatentheorie in der Frage, ob das Universum schlussendlich ein riesiger Computer ist. Die Zeitschrift Spektrum der Wissenschaft widmete dem Thema im Januar 2005 eine Spezialausgabe mit dem Thema: *Ist das Universum ein Computer ?* Dabei entsteht die Frage, welche Eigenschaften des realen Universums, wie z.B. Quanteneffekte nachweisbar digital sind und welche Eigenschaften zur digitalen Modellbildung gehören.

Der Stau im Computer – Verkehrssimulationen mit Zellularautomaten

Michael Schreckenberg
Fakultät für Physik
Universität Duisburg-Essen

Seit über 50 Jahren beschäftigt sich die Wissenschaft ernsthaft mit dem Thema "Stau". Ob Fahrzeuge, Menschen, Tiere oder Mikroben, überall wo sich viele ähnliche Objekte zur selben Zeit am selben Ort in dieselbe Richtung bewegen wollen, tritt er auf. Oft ohne erkennbaren Grund und plötzlich ("Stau aus dem Nichts"). In der Zwischenzeit sind manche Rätsel geklärt, einiges aber bleibt auch weiterhin rätselhaft und unverständlich.

Ein wesentliches Hilfsmittel dabei waren und sind Zellularautomaten, mit denen man erstmals den Stau aus dem Nichts realistisch nachbilden konnte. Aufgrund der intuitiven Struktur kann man heute sogar psychologische Aspekte mit in den Regelsatz einbeziehen. Dadurch entstehen ganz neue Stauphänomene, die man bisher so nicht erklären konnte. Die Simulationen werden zudem zur Erzeugung von Online-Informationen und Prognosen genutzt.

Trotzdem hat jeder Verkehrsteilnehmer seine eigenen Theorien und Strategien zur Bewältigung von Staus, was am Ende oft zu überraschenden Ergebnissen führt. Mit Informationen, Telematik und Navigation versucht die Technik dem Stau zu entrinnen - oft verbunden mit neuem Ungemach. Der Vortag versucht, den heutigen Stand der Erkenntnisse rund um Verkehr und Stau zu vermitteln.

Marching Pixels — Rechnen mit Hardware-Agenten im 2D Pixelraum

Dietmar Fey
Friedrich-Alexander-Universität Erlangen-Nürnberg
Department Informatik, Lehrstuhl Rechnerarchitektur

Konrad Zuses Konzept des Rechnenden Raums erweist sich nicht nur für die Modellierung physikalischer Vorgänge als besonders tragfähig, sondern es kann auch für den Entwurf paralleler Prozessorarchitekturen in der Hardware eingebetteter Systeme benutzt werden. Letzteres verwundert auch nicht, da Zuse für die Idee des Rechnenden Raumes ohnehin Universalität in Anspruch nimmt. Im Vortrag wird eine auf Zellulären Automaten aufbauende Klasse von Algorithmen vorgestellt, die nach emergenten Prinzipien arbeitet und für den Einsatz in intelligenten Kamerachips geeignet ist. Kern dieser Algorithmen bilden Hardware-Agenten, die als Marching Pixels bezeichnet werden. Diese laufen virtuell auf den Pixeln eines Bildes und suchen dort im Verbund nach Objekten, um deren Position und Drehlage zu bestimmen. Aufgrund des hohen Grades an Parallelität und den rein lokalen Operatoren werden in diesem ”rechnenden 2D-Raum” Objekte sehr schnell und unabhängig von der Größe des Bildes bearbeitet. Die Algorithmen sind für eine Umsetzung in der Hardware einer intelligenten Hochgeschwindigkeitskamera gedacht, die z.B. in der industriellen Automatisierungstechnik eingesetzt werden kann. Es werden unterschiedlich leistungsfähige Marching-Pixels-Algorithmen vorgestellt, die für unterschiedlich komplexe Anordnungen der Objekte im Bild geeignet sind. Ferner wird aufgezeigt, wie man die Idee selbst-organisierender Hardware-Agenten ausweiten kann, um Anforderungen an Robustheit und Zuverlässigkeit zu bewältigen, die sich in künftigen, auf Nanobauelementen beruhenden, Prozessor-Architekturen stellen werden. Auch hierfür bietet der Ansatz des rechnenden Raums und der damit verbundenen Diskretisierung von Raum und Zeit ein hilfreiches Fundament.

Konrad Zuses ”Rechnender Raum” — Kühne Betrachtungen nach 40 Jahren —

Friedel Hoßfeld
Juelich Supercomputing Centre
Forschungszentrum Jülich GmbH

Zurecht sind die Programme der Gedenkveranstaltungen zum 100. Geburtstag Konrad Zuses gefüllt mit Reden und Vorträgen über seine epochalen Leistungen zur Entwicklung des Digitalrechners, die ihm in den vergangenen Jahrzehnten - zwar spät, aber doch noch zu Lebzeiten - die Anerkennung als der große Computer-Pionier gebracht haben. Bei diesen hochverdienten Würdigungen wird zumeist vergessen, dass Konrad Zuse sich über die Computertechnik hinausreichend mit anderen grundsätzlichen Fragen auseinandergesetzt hat, die in zweifacher Hinsicht innovative Konzepte schufen [1, 2].

Die Wirkung dieser Überlegungen wurde jedoch einerseits durch die gleichzeitig und unabhängig - in England durch John H. Conway mit dem Spiel ”Life” und in den USA durch Edward Fredkin - gefundenen Ideen und ihre kräftige wissenschaftliche und technische Verfolgung abgeschattet, die sich seitdem mit dem innovativen Computing-Modell der zellulären Automaten oder Arrays zu einem breiten Forschungsfeld mit weitreichenden theoretischen Ergebnissen und interessanten Anwendungen entwickelt hat /3, 4/.

Es ist überaus erstaunlich, daß und wie weit Konrad Zuse, durch eine der letzten Arbeiten John von Neumanns über ”The Theory of Self-Reproducing Automata” /5/ stimuliert, das Konzept der zellulären Automaten und die Dynamik der involvierten Zustände auf dem Gitter, die er ”Digitalteilchen” nannte, schon 1966 treiben konnte. Die sowohl hinsichtlich der Universalität des Computing-Modells als auch der entwickelten Vielfalt der zellulären Automaten wichtigen Forschungen und Praxisfelder gipfelten in der 2002 erschienenen, mehr als 1200 Seiten umfassenden Monographie von Stephen Wolfram mit dem unbescheidenen Titel ”A New Kind of Science”, deren Duktus und Anspruch vielfach harsche Kritik hervorgerufen hat /6/.

Andererseits wurde Konrad Zuses vom Konzept der zellulären Automaten ausgehende, dann aber auf die Wirkungsweise physikalischer Prozesse und schließlich auf den Entwicklungsprozess des Kosmos übertragene Idee vom ”Rechnenden Raum” in der Welt der Wissenschaft - aus damaliger Sicht - nicht ernst genommen /6/.

So überschrieb sein Schüler Jürgen Alex die Würdigung der Lebensleistungen Konrad Zuses im Spektrum der Wissenschaft 1997 mit dem Titel ”Wege und Irrwege des Konrad Zuse”, und Friedrich L. Bauer schrieb in seinem Nachruf auf Konrad Zuse im Informatik-Spektrum 1996: ” Es sind erstaunlich wenig Dinge, die Zuse verkannte; daß ihm, wie er selbst oft bemerkte, der mathematische Apparat fehlte, hat ihn nicht immer wirklich behindert. Im Falle des Rechnenden Raumes ging er allerdings in die Irre. ” /7, 8/.

Fraglos mußte der theoretische und technische Ansatz Zuses für seine weitreichende Folgerung, daß das Universum ein gigantischer Automat und somit ein massiv paralleler Rechner und daher der Raum "körnig" sei und demzufolge alle Prozesse im Universum nicht kontinuierlich in Raum und Zeit, sondern in diskreten Sprüngen ablaufen, noch als naive Spekulation angesehen werden.

In diesem Vortrag wird der Versuch unternommen, Zuses Idee vom "Rechnenden Raum" nach vierzig Jahren seit ihrer Veröffentlichung im Licht der aktuellen Diskussion über die spekulativen Modelle der theoretischen Physik und der Kosmologie neu zu betrachten.

Das Phänomen des Urknalls mit seinen in der Singularität bedingten Widersprüchen einerseits, die Anstrengungen der theoretischen Physiker und der Kosmologen um die Vereinigung der Quantenmechanik mit der Gravitationstheorie andererseits, wie sie heute vornehmlich mit den Schlagworten "(Super)String Theory" und "Loop Theory" gekennzeichnet sind /9-11/, führen zu der Aussage, daß das Universum - im Grundsatz wie Zuse spekulierte - "körnig" sein muß, das heißt simplifiziert: der Raum baut sich aus Zellen der Dimension der Planck-Länge $(10^{-33}cm)$ auf und die Zeit springt in Quanten der Planck-Zeit $(10^{-43}sec)$. Die Frage, wie sich ein solcher Kosmos prozessmäßig entwickeln kann, führt so gleichfalls zu der Spekulation, daß eine gigantische Zahl kooperierender mikroskopischer Automaten das Universum treibe. Diese Automaten müssen dann aber dem Quanten-Regime folgen, das heißt: Quanten-Computer sein /12, 13/.

Referenzen:

1. Konrad Zuse: Rechnender Raum, Elektronische Datenverarbeitung Bd. 8, S. 336ff., 1967; redigierter Neudruck in: Ist das Universum ein Computer?, Spektrum der Wissenschaft - Spezial 3/07, S. 6-15

2. Konrad Zuse, Rechnender Raum, Verlag Vieweg, 1969

3. Bastien Chopard and Michel Droz, Cellular Automata Modeling of Physical Systems, Cambridge University Press, 1998

4. Martin Gerhardt und Heike Schuster, Das digitale Universum, Zelluläre Automaten als Modelle der Natur, Verlag Vieweg 1995

5. John von Neumann, The Theory of Self-Reproducing Automata, ed. by A. W. Burks, University of Illinois Press, 1966

6. Christoph Pöppe, Ist die Welt ein gigantisches Uhrwerk?, Editorial, Ist das Universum ein Computer?, Spektrum der Wissenschaft - Spezial 3/07, S. 3

7. Jürgen Alex, Wege und Irrwege des Konrad Zuse, Spektrum der Wissenschaft, H. 1 (Januar 1997), S. 78-90

8. Friedrich L. Bauer, Konrad Zuse - Fakten und Legenden, Informatik Spektrum 19 (1996), S. 299-302

9. Lee Smolin, The Trouble with Physics, Houghton Mifflin Company, Boston - New York, 2006

10. Roger Penrose, The Road to Reality - A Complete Guide to the Laws of the Universe, Alfred A. Knopf, New York, 2005

11. Martin Bojowald, Zurück vor den Urknall. Die ganze Geschichte des Universums, S. Fischer Verlag, 2009

12. Seth Lloyd, Programming the Universe, Vintage Books, London, 2007

13. Michael H. Nielsen and Isaac L. Chuang, Quantum Computation and Quantum Information, Cambridge University Press, 2000

Zuses Auffassung vom Rechnen

Bernd Mahr
Technische Universität Berlin

Während der Begriff der *Berechenbarkeit* durch die Church-Turing'sche These fixiert und allgemein akzeptiert ist, ist das, was wir als *Rechnen* verstehen, weitgehend unbestimmt. Das Verständnis von dem, was Rechnen ist, ist stark von den Gebrauchszusammenhängen abhängig, in denen es praktiziert wird. So hat sich mit der Entwicklung der Mathematik auch das Verständnis vom Rechnen sehr stark gewandelt. In Babylonischer Zeit ist Rechnen weitgehend der Umgang mit quantitativen Merkmalen konkreter Objekte in konkreten Problemsituationen, also z.B. die Ermittlung des Anwachsens der Anzahl der Tiere einer Herde oder die Bestimmung der Größe der Fläche eines bestimmten Feldes. Im antiken Griechenland ist Rechnen die Manipulation abstrakter Größen in der Arithmetik und in der Geometrie, also z.B. die Ermittlung einer Steigung oder eines Rauminhalts aus gegebenen Größen und auf der Grundlage konkreter oder abstrakter Maßeinheiten. Durch die Trennung der quantitativen Merkmale von den Objekten, denen sie zukommen, prägte sich in dieser Zeit auch ein abstrakter Zahlbegriff aus. Arithmetik und Geometrie blieben jedoch im Hinblick auf die Art der Größen, auf die sich ihr Rechnen bezog, noch lange getrennte Bereiche.

Erst mit der Definition der reellen Zahlen in der zweiten Hälfte des 19. Jh. verschmolzen die Ideen der Zahl und des Punktes. Das Kontinuum wurde in dieser Zeit in einem vereinheitlichenden mengentheoretischen Modell erfasst, und es kam durch Gottlob Frege zu einer rein logischen Definition der Zahl. Mit der damit beginnenden Mathematisierung der Logik und der mit ihr eng zusammen hängenden Axiomatisierung der Mathematik trennte sich das Rechnen von seiner direkten oder auch nur indirekten Bezugnahme auf Merkmale und Größen der realen Welt. Seit dieser Zeit bezieht es seine Rechtfertigung aus algebraischen Modellen und Theorien, die den Rechenoperationen zugrunde gelegt werden. Mit der Idee der Worttransformation in der abstrakten Algebra und vor allem mit dem Modell der Turingmaschine wird das Rechnen schließlich von seiner Bindung an Sinn und Bedeutungen befreit und zu einem elementaren Prozess des Lesens und Schreibens von Symbolen, die ihre Bezeichnungsfunktion nur noch durch ihre Unterscheidbarkeit und die örtliche Position ihres Vorkommens erfüllen. Dies ist auch die Zeit, in die Zuses Erfindung fällt. Aber er steht mit seinen Gedanken und Arbeiten in der Tradition numerischer Rechenmaschinen, die seit dem siebzehnten Jahrhundert entwickelt wurden.

Im europäischen Mittelalter und in der frühen Neuzeit ist das Rechnen die Anwendung von Rechenverfahren, deren Fortschreiten auf symbolischer Ebene beschrieben werden kann, wie bei der Ausführung einer Addition und einer Multiplikation, oder auch bei der Summenbildung und Integration. Hilfsmittel dieses Rechnens sind Tabellen, mathematische Formeln und Werkzeuge, die die Ausführung grundlegender Rechenoperationen un-

terstützen. Mit der Entwicklung von Computern treten Rechenverfahren, Programme und Softwaresysteme den Vordergrund. Zuses Rechnender Raum, das Rechnen zellularer Automaten und, in jüngster Zeit, die Modelle des Rechnens des natural computing' lösen sich schließlich vom Menschen nicht nur als dem Subjekt des Rechenakts und von der Maschine als dessen Substitut, sondern auch von der Idee eines Rechnens, das einem vorgeschriebenen Verfahren folgt. Sie stellen die Beziehungen von Entitäten mit den sie umgebenden Kontexten ins Zentrum ihrer Betrachtung. Rechnen wird damit zu einem Zustandsverlauf, dessen Definitheit zwar noch ex post, aber nicht mehr notwendig schon ex ante gegeben ist.

Konrad Zuses Auffassung vom Rechnen ist explizit. Sie zeigt sich in seinen Arbeiten, in Interviews und in seiner Biographie, vor allem aber auch in den von ihm entwickelten Rechenmaschinen, dem Plankalkül und in seinen Ausführungen zum Rechnenden Raum. Seine Überlegungen zum Rechnen durchziehen sein Arbeiten und sein Denken wie ein roter Faden. Vor dem Hintergrund der historischen Entwicklung stelle ich mir drei Fragen: 1. Wie weit gehend kann der Gedanke des Rechnens sinnvoll gefasst werden? 2. Gibt es ein Modell des Rechnens, das alle Entwicklungsstadien und bekannten Varianten zu erfassen erlaubt? und 3. Wie ist Zuses Auffassung vom Rechnen in die Entwicklungsgeschichte und in ein allgemeines Modell des Rechnens einzuordnen? Mit der Beantwortung dieser Fragen soll Zuses Leistung als Denker des Rechnens herausgearbeitet werden. Dabei wird sich zeigen, dass sich sein Verständnis sehr deutlich von dem anderer Pioniere des Rechnens unterscheidet.

Workshop

2. Workshop zur IT-Unterstützung von Rettungskräften

Christian Erfurth, Birgitta König-Ries

Vorwort zum 2. Workshop IT-Unterstützung für Rettungskräfte

Christian Erfurth
Fachhochschule Jena
Jena, Germany
christian.erfurth@fh-jena.de

Birgitta König-Ries
Friedrich-Schiller-Universität Jena
Jena, Germany
birgitta.koenig-ries@uni-jena.de

Im täglichen Leben hat sich der Einsatz von IT in den letzten Jahren zu einer Selbstverständlichkeit entwickelt. Neueste Technik ist in nahezu jedem Haushalt vorhanden und verschönert uns den Alltag. Abseits der handelsüblichen Geräte mit Standardsoftware aus industrieller Fertigung gibt es dennoch Anwendungsbereiche, die die Wissenschaft Informatik fordern. So ist die IT-Unterstützung von Einsatzkräften der unterschiedlichen Behörden und Organisationen mit Sicherheitsaufgaben (BOS) bei ihrer täglichen Arbeit dringend notwendig. Der effiziente Einsatz kann entscheidend sein, damit die Rettung von Menschenleben in einer schwierigen und gefährlichen Situation erfolgreich ist.

Der Mensch steht in zweifacher Hinsicht in dieser Anwendungsdomäne im Vordergrund: Als Rettungskraft und damit als Anwender der IT sowie als Betroffener, der geborgen, versorgt oder informiert werden will. IT muss somit einen Dienst dem Menschen erweisen, der in einem chaotisch erscheinenden Einsatz die richtigen Entscheidungen treffen muss: Wie schwer ist die Person verletzt? Wen versorge ich als nächstes? Wo sind Verletzte? Welches Gebiet ist gefährdet? Wie viele Helfer habe ich? Welche Kapazitäten haben die Krankenhaus? Wie soll die Räumung ablaufen? Zu den vielen Fragen kommt das optimale Zusammenspiel der Einsatzkräfte innerhalb der Organisation und insbesondere zwischen Organisationen. Kommunikation ist ein wesentliches Merkmal für gelungene Einsätze.

In unserem ersten Workshop zur INFORMATIK 2009 wurde aufgezeigt, wie umfangreich und vielfältig das Einsatzgebiet von IT im Kontext der Rettungskräfte ist. Hier gibt es für die Informatik noch viel zu leisten. In diesem Sinne ist das Motto der INFORMATIK 2010 „Service Science – Neue Perspektiven für die Informatik“ für unseren Workshop treffend. Den Bedarf an IT signalisieren uns engagierte Anwender, die verschiedenen Forschungsprojekten zur Seite stehen – meist aus starkem Interesse teilweise auch unterstützt durch Fördermittel auf Bundesebene z.B. im Rahmen des Sicherheitsforschungsprogramm vom BMBF oder auch aus Mitteln der EU. Wir hoffen mit dem Workshop einen Beitrag zum Fortschritt der Forschung in diesem Bereich in Deutschland zu leisten.

Der Workshop umfasst zum einen die Präsentation von wissenschaftlichen Beiträgen, zum anderen soll er aber auch Raum für intensiven Erfahrungsaustausch zwischen Wissenschaftlern und Anwendern bieten. Er vermittelt einen Überblick über den Stand der Technik in relevanten Teilbereichen und soll den jeweiligen Handlungsbedarf aufzeigen. Der Workshop kann als Diskussionsplattform dienen, die den Austausch zwischen unterschied-

lichen Projekten fördert. Nicht zuletzt wird die Vernetzung der einschlägig arbeitenden Gruppen im deutschsprachigen Raum angestossen bzw. weiter befördert. Um diesen Zielen gerecht zu werden, gliedert sich das Workshopprogramm in drei Teile:

- In zwei **eingeladenen Vorträgen** wird das Workshopthema aus einem etwas weiteren Blickwinkel beleuchtet. Diese Perspektive erscheint uns bei diesem Thema besonders wichtig, zeigen doch Erfahrungen, dass die beste IT-Lösung im Rettungsumfeld nichts nützt, wenn sie an den Bedürfnissen und Kulturen der Rettungskräfte oder den gesellschaftlichen und rechtlichen Randbedingungnen vorbeientwickelt wurde.
- Die zum Workshop angenommenen **wissenschaftlichen Beiträge** werden beim Workshop in Vorträgen präsentiert. Diese Beiträge geben detaillierten Einblick in unterschiedlichste aktuelle Arbeiten zur IT-Unterstützung von Rettungskräften.
- Um den Austausch zwischen Gruppen, die sich möglicherweise mit ähnlichen Fragestellungen beschäftigen, zu fördern, wird es eine Reihe von **Kurzvorstellungen von aktuell laufenden Projekten** geben. Diese Kurzvorstellungen haben - ebenso wie der Workshop insgesamt - das Ziel, den Erfahrungsaustausch zwischen unterschiedlichen Arbeitsgruppen anzuregen, Anknüpfungspunkte aufzuzeigen und die Vernetzung in der deutschsprachigen Community zu unterstützen.

Dieser Workshop wäre nicht möglich gewesen ohne die bereitwillige Zusage unserer eingeladenen Vortragenden, ohne die Einreichungen aus diversen Projekten, die unterschiedliche Aspekte des Themas beleuchten, und ohne die Unterstützung des Organisationsteams der Informatik 2010. Ihnen allen gilt unser Dank ebenso wie den Mitgliedern des Programmkomitees, die uns tatkräftig unterstützt haben:
Benedikt Birkhäuser (Bergische Universität Wuppertal)
Johannes Hertel (Rittal GmbH & Co. KG)
Carsten Holtmann (FZI Karlsruhe)
Gudrun Klinker (TU München)
Rainer Koch (Uni Paderborn)
Antonio Krüger (DFKI Saarbrücken)
Bernd Lehnert (Rittal GmbH & Co. KG)
Andreas Mitschele-Thiel (TU Ilmenau)
Simon Nestler (TU München)
Jens Pottebaum (Uni Paderborn)
Thomas Rose (Fraunhofer FIT)
Kai-Uwe Sattler (TU Ilmenau)
Thomas Schmidt (TU München Werkfeuerwehr)
Stefan Stein (Rittal GmbH & Co. KG)
Stefan Strohschneider (FSU Jena)
Christian Wietfeld (TU Dortmund)

Wir wünschen allen Teilnehmern einen anregenden Austausch und allen, die den Tagungsband in den Händen halten, eine spannende Lektüre.

Christian Erfurth und Birgitta König-Ries Jena, im Juli 2010

Wissenschaftliche Optimierungsmethoden für die strategische Planung und die operative Einsatzführung in der Gefahrenabwehr - Implementierung und praktische Anwendung

Dr. K. Danowski

Gruppenleiter Disposition und strategische Optimierung
Fraunhofer-Institut für Verkehrs- und Infrastruktursysteme
Zeunerstraße 38
01069 Dresden
kamen.danowski@ivi.fraunhofer.de

Abstract: Für die wirksame Softwareunterstützung bei komplexen strategischen und operativen Entscheidungen ist die Vernetzung von Praxiswissen und modernen wissenschaftlichen Methoden von entscheidender Bedeutung. Die Technologie MobiKat - Mobile Information Kommandoarbeit und Taktik integriert in einem Framework hochleistungsfähige Analyse- und Optimierungsalgorithmen, benutzerfreundliche Bedienungs- und Visualisierungskomponenten sowie Geo- und Sachdatenmodule. Durch die implementierte flexible Architektur werden die Optimierungsmodule je nach Konfiguration von Planern, Stäben und Einsatzleitern in den Bereichen Feuerwehr, Rettungsdienst, Katastrophenschutz und Polizei eingesetzt. Die Technologie befindet sich seit mehreren Jahren erfolgreich im alltäglichen Einsatz und wird mit den Anwendern weiterentwickelt.

1 Motivation

Die Vorsorge und Bewältigung von Großschadenslagen sowie die alltägliche Gefahrenabwehr sind mit komplexen planerischen und operativen Entscheidungen verbunden, welche für die Effektivität der Maßnahmen von entscheidender Bedeutung sind. Wissenschaftliche Berechnungs- und Optimierungsmethoden im Rahmen von Entscheidungsunterstützungssystemen bieten hierbei eine effektive Hilfe. Praxiserfahrungen zeigen, dass eine Reihe von Aufgaben und Problemgruppen sowohl im organisatorisch-administrativen als auch im operativ-taktischen Bereich Gemeinsamkeiten aufweisen und somit eine multivalente Nutzung von Grundalgorithmen und Verfahren ermöglichen.

Aus diesem Grund wurde unmittelbar mit Praxisanwendern aus den Bereichen Feuerwehr, Katastrophenschutz und Rettungsdienst die Technologie MobiKat[1] entwickelt, welche der unmittelbaren Überführung von Forschungsergebnissen in die praktische Anwendung der Gefahrenabwehr dient. Einen Hauptschwerpunkt hierbei stellt neben der Bereitstellung und Aktualisierung von relevanten und aktuellen Informationen im Bereich der Gefahrenabwehr vor allem die wirkungsvolle Entscheidungsunterstützung im strategischen sowie im operativ-taktischen Bereich mittels geeigneter Analyse- und Optimierungsverfahren in einem Freamwork dar. Im Mittelpunkt stehen dabei gleichermaßen effektive Planung, schnelle Entscheidungen in zeitkritischen Situationen bei Krisen- bzw. Großschadenslagen und Entscheidungsunterstützung in alltäglichen Einsätzen.

2 Architekturkonzept

Das gemeinsam mit den Praxisanwendern erarbeitete mehrstufige Konzept zur Entwicklung und Bereitstellung von flexiblen Analyse- und Optimierungswerkzeugen definiert in seiner Gesamtheit kein starres und geschlossenes System, sondern stellt eine Reihe von Funktionen zur flexiblen Integration bereit. Die eigentliche Systemarchitekturausprägung ist vor allem vom konkreten Anwendungsumfeld - Arbeitsprozesse, Rahmenbedingungen und Aufgaben, Strategien und Ziele und Verantwortlichkeiten u. a. - abhängig. Grundlage dafür sind einerseits Muster zur Wiederverwendung von Komponenten und andererseits die konkreten Module und Funktionsgruppen, welche auf unterschiedlichen Architekturebenen agregiert werden. Einige wichtige Funktionsgruppen in MobiKat sind unter anderem Schnittstellen, Datenmanagement, Visualisierung, Kommunikation, Entscheidungsunterstützung, serverorientierte Bereitstellung von Diensten.

Zum Bereich Entscheidungsunterstützung gehören unter anderem Basismodule und -algorithmen, Module zur Analyse, Komponenten für strategische Optimierung, operativ-taktische Einsatzplanung und -führung. Diese kommunizieren je nach Anwendung und Einatzumgebung über Schnittstellen mit Einsatzmittellokalisierungsmodulen, Datenbankabfragekomponenten, kartenbasierten Visualisierungsmodulen, Komponenten zur Bereitstellung von Diensten für Fremdsysteme u. a. Somit können diese flexibel in unterschiedlichsten Architekturen und Anwendungskontexten gemäß folgender Klassifizierung eingesetzt werden:

- Aufgabe - Analyse, Planung und operative Einsatzführung
- Einsatzbereich - Feuerwehr, Rettungsdienst, Katastrophenschutz und Polizei
- Nutzungsplattform - Desktop, browserbasiert im Internet und Intranet, Dienste für Fremdsysteme
- Nutzerschnittstelle - mit oder ohne grafischer Benutzeroberfläche.

[1] MobiKat - Mobile Kommandoarbeit und Taktik

3 Datengrundlagen und Aktualisierung

Eine primäre Voraussetzung für die effektive Entscheidungsfindung in der Praxis stellt die Bereitstellung von aktuellen relevanten Geo-, Infrastruktur- und Sachdaten dar. Maßgeblich für die zu integrierenden Themen und resultierenden Verarbeitungsschritte ist die Spezifik der zu lösenden Aufgaben und der geforderte Detaillierungsgrad der Ergebnisse. Die inhaltliche Verknüpfung und die Sicherung der laufenden Aktualität der komplexen Datenbasis wird in MobiKat aus räumlich verteilten, heterogenen Datenquellen über Schnittstellen realisiert, so dass man sowohl mit amtlichen, mit kommerziellen sowie frei verfügbaren Daten arbeiten kann. Die Informationen werden in MobiKat automatisiert klassifiziert und entsprechend ihrer Relevanz gefiltert und auf Widerspruchsfreiheit, Konsistenz und Vollständigkeit geprüft. Anschließend werden die Abhängigkeiten und Beziehungen zwischen den unterschiedlichen Datenkategorien im System hergestellt. Die hohe Aktualität wird mit minimalem Pflegeaufwand gewährleistet, da die Informationen unmittelbar von den Stellen bezogen werden, welche sie primär erzeugen und pflegen.

Im Rahmen der zu lösenden Aufgaben und der von den Anwendern definierten Implementierungen wurden nach diesem Konzept eine Vielzahl themenbezogener sowie regionbezogener Datenbasen erstellt und ihre laufende Aktualisierung gesichert. Beispiele für integrierte Datenthemen sind: Straßen-, Wege- und Schienennetz, Brücken, Tunnel und aktuelle Baustellen, Gewässerflächen und Linien, Überflutungsflächen für verschiedene Pegelstandorte, Adresskoordinaten, Einwohnerzahlen, Gebäudeumrisse, Flurstücksgrenzen und Flächennutzung, kritische Infrastruktur und Schwerpunktobjekte der Gefahrenabwehr, Objekte mit Brandmeldeanlagen, Feuerwehrstandorte, Rettungs- und Polizeiwachen, Einsatzmittel, Hydranten, digitales Geländemodell (DGM25), Unfallschwerpunkte (wie Kletterfelsen u. a.), hochaufgelöste Orthofotos u. v. m. [SD10]

4 Implementierte Optimierungsalgorithmen und -verfahren

4.1 Strategische Planung

Die Module für strategische Analysen und Planung in MobiKat werden bei der Lösung von Aufgaben im Zusammenhang mit längerfristigen Maßnahmen im administrativ-organisatorischen Bereich eingesetzt. Dazu gehören Module, wie:

- Risiko-, Hilfsfrist- und Schutzzielanalysen
- Optimale Standortplanung für Feuer- und Rettungswachen
- Ermittlung der geeignetsten Wach- und Ausrückebereiche
- Standortkonzepte für Spezialfahrzeuge und -geräte
- Alarm- und Ausrückeordnungen

- Szenarienorientierte Überprüfung von geplanten Verkehrsmaßnahmen
- Analysen für externe Notfallpläne, Gefahrenabwehrpläne u. a.

4.2 Operativ-taktische Einsatzplanung

Im operativ taktischen Bereich unterstützen MobiKat-Komponenten komplexe Entscheidungen in der Einsatzführung. Module für die Ermittlung von optimierten Handlungsstrategien werden in folgenden Bereichen bereitgestellt:

- Lageabhängige Auswahl und Heranführen von Einsatzkräften
- Dislozierung von Kräften und Mitteln
- Berechnung von langen Löschwasserförderstrecken
- Optimierung des Kräfte- und Mitteleinsatzes bei der Vermisstensuche
- Einsatzplanung und -führung bei Großveranstaltungen
- Optimale Verletztenversorgung in Großschadenslagen (MANV)
- Situationsabhängige Marschroutenermittlung für Einsatzkräfte
- Optimierung der Einsatzplanung bei Evakuierungsmaßnahmen.

5 Anwendung in der Praxis

Das vorgestellte Konzept kam erstmals während der Hochwasserkatastrophe an der Elbe im Frühjahr 2006 zur praktischen Anwendung. Die benötigten Optimierungs- und Visualisierungsmodule von MobiKat wurden innerhalb kürzester Zeit im Katastrophenschutzstab eingerichtet und für die Lagevisualisierung, die Analyse der Hochwasserlage, Lageprognose und die Einsatzplanung eingesetzt. Die MobiKat-Komponenten unterstützten rund um die Uhr die Lageführung und die optimale Ressourcenplanung und stellten digitale Karten der überfluteten und gefährdeten Gebiete, optimale Verkehrswege für Einsatzkräfte, Evakuierungslisten für die aktuelle und prognostizierte Hochwasserentwicklung u. v. m. bereit. [Be08] [Ur08]

MobiKat bewährte sich als Planungs- und Führungsinstrument auch bei einer Reihe von Großeinsätzen, wie Stadtfesten und Großveranstaltungen mit mehreren Zehntausend Besuchern, internationalen Sportveranstaltungen und -events, ein Krankenhaus-Umzug u. a., sowie bei der Konzipierung und Durchführung von mehreren Feuerwehr- und Katastrophenschutzübungen.

Mit angepassten Funktionsinhalten wird MobiKat in Fachabteilungen der Verwaltung, als mobiles Entscheidungsunterstützungssystem auf Tablet-PC's für die Einsatzleiter sowie auf Einsatzfahrzeugen (s. Abbildung 1) angewendet. Das System wird auch täglich als Informations- und Entscheidungsunterstützungssystem in zwei Rettungsleitstellen genutzt. [Be08]



Abbildung 1: Anwendung von MobiKat im Führungsfahrzeug des DRK Kreisverbandes Dresden und im Einsatzleitwagen ELW 2+ Sachsen

Die MobiKat-Komponenten zur strategischen Planung unterstützen in der Praxis die Brandschutzbedarfs- und Rettungsdienstbereichsplanung. Mit deren Hilfe werden umfangreiche Analysen zur Abdeckung einer Region mit Einsatzmitteln innerhalb einer definierten Hilfsfrist durchgeführt sowie die optimalen Grenzen der Einsatzbereiche benachbarter Wachen berechnet. Das System ermittelt ebenfalls die Auswirkungen von Maßnahmen wie den Bau neuer, der Schließung oder der Zusammenlegung bestehender Wachen. Es trifft Aussagen über die Zielerfüllung möglicher Alternativen und hilft, Varianten nach verschiedenen Kriterien zu vergleichen (s. Abbildung 2). MobiKat wurde zur Lösung von Analyse- und Planungsaufgaben im Landkreis Sächsische Schweiz-Osterzgebirge, im Landkreis Meißen, in Dresden, Leipzig und Pirna eingesetzt.



Abbildung 2: MobiKat-Analysen zur Abdeckung mit Feuerschutzmittel und Besprechung von Varianten der strategischen Planung

Die MobiKat-Technologie wird als eine wesentliche Grundlage im Rahmen des Vorhabens COSMOD[2] für grenzüberschreitende Gefahrenanalysen und Planungen in der Euroregion Elbe-Labe eingesetzt. Hauptaufgaben sind die optimierte Koordination der Zusammenarbeit von Feuerwehr, Rettungsdienst und Katastrophenschutz zwischen Deutschland und Tschechien. Die entwickelten Komponenten tragen dazu bei, mit den verfügbaren Kräften und Mitteln den optimalen Schutz der Bevölkerung sicherzustellen und bei Großschadenslagen eine schnelle grenzüberschreitende Hilfeleistung zu gewährleisten.

6 Zusammenfassung und Ausblick

Die Erfahrungen aus den vielfältigen erfolgreichen Einsätzen in der Praxis zeigen, dass wissenschaftliche Analyse- und Optimierungsverfahren eine wirksame Unterstützung sowohl bei strategischen als auch bei operativ-taktischen Entscheidungen leisten. Das flexible Architekturkonzept von MobiKat ermöglicht eine multivalente Nutzung von Komponenten und Modulen und je nach Konfiguration die Unterstützung von Planern, Stäben und Einsatzleitern in den Bereichen Feuerwehr, Rettungsdienst, Katastrophenschutz und Polizei.

Da das implementierte Architekturkonzept nicht auf ein geschlossenes System zielt, sondern als Basis die Bereitstellung einer Reihe von Funktionen und Modulen zur Integration hat, ermöglicht mittels flexibler Konfiguration einer schnelle Anpassung an konkrete Aufgaben und Anwendungsumgebungen. Auf dieser Grundlage werden ausgehend vom Praxisbedarf gemeinsam mit den Nutzern neue Komponenten konzipiert und implementiert.

Literaturverzeichnis

[Be08] Beul, H.: Zeit rettet Leben. RettMobil - Fachmagazin für Führungskräfte und Helfer in Feuerwehr, Rettungsdienst und Verwaltungen, Ausgabe Juni-August 2008, Drum Verlag S. 38-41.

[SD10] Scholl, H.; Danowski, K.: Das Dispositions- und Entscheidungsunterstützungssystem MobiKat: Innovation für die Einsatzplanung und Durchführung, Im Einsatz - Zeitschrift für Helfer und Einsatzkräfte, Ausgabe 3/2010, Stumpf+Kossendey Verlag S. 42-45.

[Ur08] Urbanski, B.: MobiKat als digitaler Lebensretter im Dienst, Sächsische Zeitung, 22.02.2008.

[2] COSMOD - Cross-border system for management and optimisation of disaster control and crisis management

Entwicklung eines ereignisgetriebenen Systems für intelligentes Ressourcenmanagement – ‚Work-in-Progress'-Report zum Stand des ersten Demonstrators

Benedikt Birkhäuser

Universität Paderborn / C.I.K.
Warburger Str. 100
D – 33098 Paderborn
b.birkhaeuser@cik.upb.de

Jochen Schwenniger

Fraunhofer IAIS
Schloß Birlinghoven
D – 53754 St. Augustin
jochen.schwenninger@iais.fraunhofer.de

Ben Loke

Noldus Information Technology BV
Nieuwe Kanaal 5
NL – 6709 PA Wageningen
b.loke@noldus.nl

Laura Niittylä

Mattersoft Ltd.
P.O. Box 507
FI – 33101 Tampere
laura.niittyla@mattersoft.fi

Pythagoras Karampiperis

NCSR Demokritos
Napoleon St.
GR – 15310 Aghia Paraskevi, Attiki
pythk@iit.demokritos.gr

Robin Marterer

Universität Paderborn / C.I.K.
Warburger Str. 100
D – 33098 Paderborn
marterer@cik.upb.de

Abstract: Entscheidungen über Ressourcenallokationen stellen eine wesentliche Herausforderung im Alltag der Entscheidungsträger von Zivil- und Katastrophenschutzorganisationen dar. Im vorliegenden Beitrag wird der aktuelle Stand der Entwicklung eines ereignisgetriebenen Systems zur Unterstützung mittels intelligenten Ressourcenmanagements (IRM) vorgestellt. Aufbauend auf der Beschreibung der gewählten System-Architektur werden erste Ergebnisse der Demonstratorentwicklung präsentiert.

1 Projekt PRONTO[1]: Einführung und Projektrahmen

Entscheidungen über Ressourcenallokationen stellen eine wesentliche Herausforderung im Alltag der Entscheidungsträger von Zivil- und Katastrophenschutzorganisationen dar. Zunehmend werden Möglichkeiten der Unterstützung derartiger Entscheidungen durch Informationssysteme erforscht. Im Rahmen des Projektes SHARE[2] wurden dazu frühzeitig erste Ansätze aufgabenangepasst Informationen zur Verfügung zu stellen umgesetzt. Ziel des Projektes PRONTO ist es, diese Ansätze zu vertiefen und im Rahmen eines ereignisbasierten Systems weiterzuentwickeln.[3] Zwei Case-Studies im Bereich des ÖPNV und im Bereich der Feuerwehr bilden die Grundlage zur Analyse und Validierung des Ansatzes. Kernidee im Bereich des ÖPNV ist dabei die Unterstützung von Betriebsverantwortlichen beim Auftreten unerwarteter Störungen bzw. die Optimierung des laufenden Betriebs. Für den Bereich der Feuerwehr stehen ein Hochwasserszenario und die damit verbundenen Nicht-Routine-Entscheidungen im Mittelpunkt. Zielsetzung des PRONTO-Systems ist dabei die Unterstützung der Entscheidungsträger auf strategischer Ebene.
Basis für alle Überlegungen ist die Idee, Daten von verschiedenen Quellen in Form von Ereignissen zusammenzuführen und zu für den Nutzer brauchbaren Informationen zusammenzustellen. Hier wird insbesondere der Ansatz des intelligenten Ressourcenmanagements betrachtet [BPK09]. Während des ersten Projektjahres sind dazu im Rahmen der Anforderungsanalyse verschiedene Einsatzszenarien auf Ereignisse hin untersucht [Fr10] und darauf aufbauend eine Systemarchitektur entworfen worden [MBK10]. Parallel sind Module für den Bereich der Ereigniserkennung in verschiedenen Datenströmen (Audio, Video, GPS), sowie im Bereich des Complex Event Processing [Luc07] vom Konsortium [Ar10] entwickelt worden.

2 System-Architektur

Aus technischer Sicht ist eine übergreifende Anforderung für beide Case-Studies die der Echtzeiteigenschaft des zu entwickelnden Systems.[4] Weiterhin müssen eventuelle Netzausfälle kompensiert werden können und Module sowie Sensoren variabel mit dem System verbunden werden können. Als Designentscheidung bildet der Ansatz der asynchronen Kommunikation den weiteren Gestaltungsrahmen für die System-Architektur. Naturgemäß wird das System in einer verteilten Umgebung Anwendung finden.
Das Paradigma der Event-Driven Architecture [Ta09] stellt eine aktuell in der Forschung beleuchtete Möglichkeit dar, diesen Anforderungen zu entsprechen. Die für das Projekt PRONTO gewählte System-Architektur lehnt sich an die von [Du08] vorgeschlagene

[1] „Event Recognition for Intelligent Resource Management“, gefördert im Bereich „Information and Communication Technology“ des 7. Forschungsrahmenprogramms der Europäischen Kommission 2009 bis 2012. Vgl. http://www.ict-pronto.org

[2] „Mobile Support for Rescue Forces, Integrating Multiple Modes of Interaction”, gefördert im Bereich „Information Society Technology “ des 6. Forschungsrahmenprogramms der Europäischen Kommission, 2005 bis 2007. Vgl. http://www.ist-share.org

[3] Vgl. dazu auch [BPK09].

[4] Zur detaillierteren Beschreibung der Case-Studie ‚Feuerwehr‘ siehe [Gr09].

Referenzarchitektur für ereignisbasierte Systeme an und besteht im Wesentlichen aus folgenden Modulen (vgl. Abbildung 1):

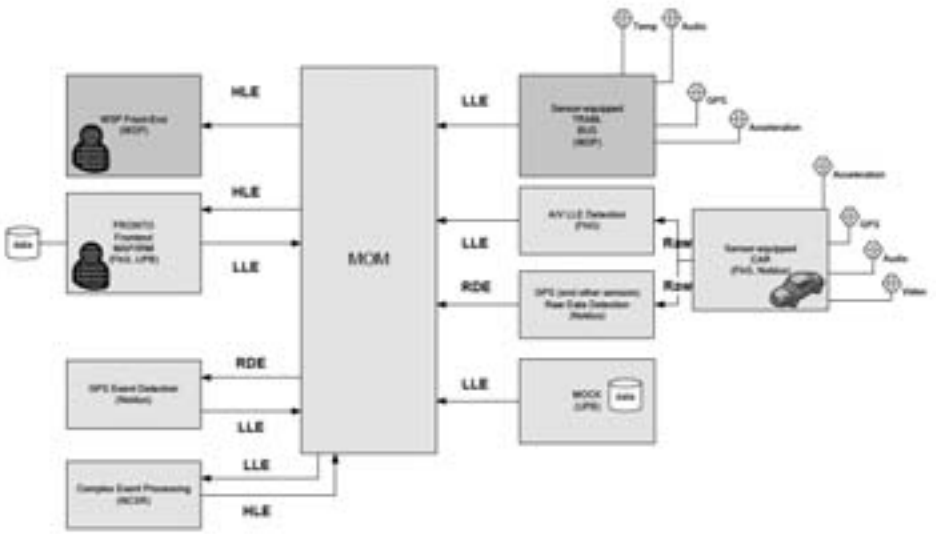


Abbildung 1: System-Architektur des Demonstrators (in Klammern: Verantwortlicher Partner)

- Middleware: Als Middleware wird ein einfaches, nachrichtenbasiertes System auf Grundlage von JMS eingesetzt. HornetQ[5] ist eine frei verfügbare Middleware-Lösung, die den JMS-Standard implementiert. Verglichen mit anderen Produkten in diesem Bereich zeichnet sie sich durch eine hohe Performanz und Möglichkeiten des Journaling aus. HornetQ ist der Nachrichtenserver des JBoss-Projektes und ist derzeit in der Version 2.0.0GA[6] verfügbar. Basis für den Austausch von Informationen bildet eine Syntax zur Abbildung der im System bekannten Ereignisse, die in Form von Java-Klassen für alle Module des Systems zur Verfügung steht. Ereignisse werden als einfache Textnachrichten übertragen, die mittels JSON[7] kodiert werden. In PRONTO werden nach ihrer Position in der Ereignishierarchie Raw Data Events (RDE), Low Level Events (LLE) und High Level Events (HLE) unterschieden.
- Logging / Testing / MockUp Framework: Während der Testphase des PRONTO-Systems ist nicht sichergestellt, das permanent Echt-Daten im System zum Testen zur Verfügung stehen. Aus diesem Grund ist ein MockUp Framework entwickelt worden, das es ermöglicht alle über das System kommunizierten Ereignisse zu Loggen und zu einem späteren Zeitpunkt und unter Beibehaltung der zwischen den Ereignissen auftretenden Zeitabstände wieder in das System einzuspielen.

[5] http://www.jboss.org/hornetq
[6] Stand 16.05.2010.
[7] http://www.json.org

- Sensoren: Für die beiden Case-Studies ÖPNV und Feuerwehr werden derzeit unterschiedliche Sensoren an das System angebunden. Für den Bereich des ÖPNV senden derzeit alle Trams und Busse des Anwendungspartners GPS-Daten. Eine spezielle Tram ist mit weiteren Sensoren wie Beschleunigungssensoren, Temperatursensoren und Mikrophonen ausgestattet. Im Bereich der Feuerwehr sind derzeit im Echtbetrieb nur wenige Einsatzfahrzeuge mit GPS-Sensoren ausgestattet. Für Testzwecke ist ein Fahrzeug mit GPS-, Temperatur- und Audio-/Video-Sensoren ausgestattet worden. Weiterhin ist geplant, das System mit den bestehenden Einsatzleitrechnern zu vernetzen.
- Video-Modul: Das Video-Modul wird eingesetzt um die Auslastung eines Fahrzeuges mit Passagieren (ÖPNV) zu bestimmen. Dazu wird ein Hintergrundmodell des leeren Fahrzeugs angelegt und Keypoints und Deskriptoren extrahiert. Diese Keypoints und Deskriptoren dienen im späteren Einsatz zur Berechnung der aktuellen Fahrzeugauslastung.
- Audio-Modul: Ein aussagefähiges Ereignis für die ÖPNV-Case-Study ist das des Passagierkomforts. Dazu wird der Geräuschpegel innerhalb der Fahrzeuge analysiert und in Form von Ereignissen zur Verfügung gestellt. Für spätere Ausbaustufen des Systems ist eine Analyse von Funksprüchen (Feuerwehr) auf Grundlage der TETRA-Technologie vorgesehen, um auch die Kommunikation zwischen den Einsatzkräften zur Extraktion von Ereignissen zu nutzen.
- GPS LLE Detection: GPS-Sensoren produzieren einen Datenstrom, der insbesondere bei mehreren eingesetzten GPS-Sensoren vom TETRA-Netzwerk aufgrund mangelnder Bandbreite nicht übertragen werden kann. Das ‚GPS LLE Detection'-Modul analysiert jeweils einen GPS-Datenstrom vor der Übertragung über das Netzwerk und beschränkt auf Grundlage von Grenzwerten die Anzahl der über das Netzwerk zu übertragenden GPS-Ereignisse.
- Complex Event Processing: Das ‚Complex Event Processing'-Modul führt verschiedene Events zusammen und bildet neue, höherwertige Ereignisse. Grundlage ist ein PROLOG-Reasoner [ASP10], der auf Basis einer Wissensdatenbank und vordefinierter Ereignishierarchien diese neuen Ereignisse ableitet.

3 Beschreibung des ersten Demonstrators

Der aktuelle Stand des Demonstrators wird in Abbildung 2 dargestellt. Das Nutzerinterface ist als einfache Webanwendung auf der Basis von AJAX-Technologien implementiert worden. Gut zu erkennen ist, dass bereits Echtzeitdaten in Form von Ereignissen in das System eingespielt werden. Die Kartenanwendung bildet über Ereignisse eintreffende Informationen verknüpft mit Ortsangaben in der GUI ab. Die Karte stellt die für den Nutzer gewohnten Funktionen zur Verfügung: Scrollen in die verschiedenen Himmelsrichtungen sowie Hinein- und Herauszoomen. Ebenfalls sind erste Funktionalitäten zur Markierung von bestimmten Regionen auf der Karte bereits implementiert. Diese können in der aktuellen Version bereits als Ereignisse aufgefasst werden und beeinflussen das System dementsprechend.



Abbildung 2: Screenshot des Demonstrators[8]

Abbildung 3 stellt einen Ausblick auf eine mögliche Weiterentwicklung der GUI dar. Zusätzlich zur aktuellen Demonstratorversion werden weitere GUI-Module zur Darstellung von Hilfstexten oder zur genaueren Beschreibung der Ereignisse eingebunden. Weiterhin werden die Funktionen zur Interaktion des Nutzers mit der Anwendung ausgebaut.

[8] Stand 16.05.2010

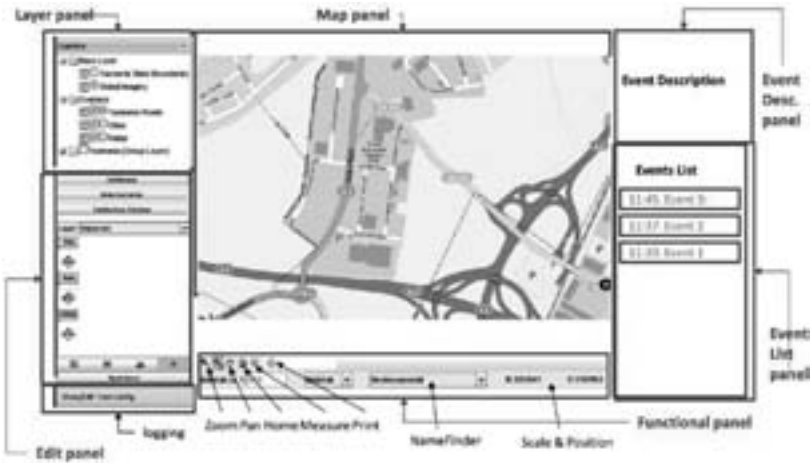


Abbildung 3: GUI-Konzept

4 Fazit und Ausblick

Bereits nach knapp einem Jahr ist im Projekt PRONTO ein erster, lauffähiger Demonstrator entwickelt worden. In ersten Tests im Zuge der Demonstratorentwicklung hat sich die gewählte Systemarchitektur als robust genug für den geforderten Einsatzbereich erwiesen. Auch im Hinblick auf die Flexibilität sind die ersten Erfahrungen positiv. Im weiteren Verlauf des Projektes wird der Demonstrator innerhalb von Evaluationsphasen weiter getestet werden. Ziel ist es, die ersten Eindrücke durch konkrete Messwerte bestätigen zu können.

Der entwickelte Demonstrator bildet die Grundlage für alle weiteren Entwicklungen. Im Zuge der Entwicklung des Demonstrators wurde Wert auf volle Integration aller Module gelegt, d.h. in der Abwägungsentscheidung zwischen vertiefter Modulentwicklung und lauffähigem Gesamtsystem wurde für ein lauffähiges Gesamtsystem optimiert. Entsprechend wird in den nächsten Entwicklungsschritten eine weitere Verbesserung der einzelnen Module im Mittelpunkt stehen.

Literaturverzeichnis

[Ar10] Artikis, Alexander; Paliouras, Georgios; Portet, François; Skarlatidis, Anastasios (2010): Logic-Based Representation, Reasoning and Machine Learning for Event Recognition. In: Proceedings of Conference on Distributed Event-Based Systems (DEBS 2010), ACM Press, 2010.

[ASP10] Behaviour Recognition From Video Content: A Logic Programming Approach (2010). In: International Journal on Artificial Intelligence Tools, Jg. 19, April 2010, S. 192-209.

[BPK09] Birkhäuser, Benedikt; Pottebaum, Jens; Koch, Rainer (2009): Unterstützung von Einsatzentscheidungen der Feuerwehr auf Basis IT-unterstützter Kräftekoordination. In: Fischer, Stefan (Hg.): Informatik 2009. Im Focus das Leben. Beiträge der 39. Jahrestagung der Gesellschaft für Informatik e.V. (GI), 28.9. - 2.10.2009 in Lübeck. Bonn: Gesellschaft für Informatik (GI-EditionProceedings, 154), S. 1393-06.

[Du08] Dunkel, Jürgen; Eberhart, Andreas; Fischer, Stefan; Kleiner, Carsten; Koschel, Arne: Systemarchitekturen für verteilte Anwendungen. Carl Hanser Verlag, München, 2008.

[Fr10] Friberg, Therese; Birkhäuser, Benedikt; Pottebaum, Jens; Koch, Rainer (2010): Using Scenarios for the Identification of Real-World Events in an Event-Based System. In: French, Simon; Tomaszewski, B.; Zobel C. (Hg.): Proceedings of the 6th International Conference on Information Systems for Crisis Response and Management ISCRAM2010.

[Gr09] Grünewald, Uwe; Schümberg, Sabine; Wölleke, Britta; Riesenbeck, Gert Graf-van; Piroth, Klaus (2009) Gutachten zu Entstehung und Verlauf des extremen Niederschlag-Abfluss-Ereignisses am 26.07.2008 im Stadtgebiet von Dortmund. Online verfügbar unter: http://www-docs.tu-cottbus.de/hydrologie/public/files/Hochwassergutachten_Dortmund.pdf. Zuletzt abgerufen am 03.07.2010.

[Luc07] Luckham, David: The Power of Events. Boston : Addison-Wesley, Pearson Education, 2007.

[MBK10] Marterer, Robin; Birkhäuser, Benedikt; Koch, Rainer (2010): Das Paradigma der Event-Driven Architecture als Grundlage für ein entscheidungsunterstützendes IT-System im Zivil- und Katastrophenschutz. In: Workshopband zur Konferenz Software Engineering 2010.

[Ta09] Taylor, Hugh; Yochem, Angela; Phillips, Les; Martinez, Frank: Event-Driven Architecture. Addison-Wesley, Pearson Education, Boston, 2009.

Graph analysis on uncertain measurements for position estimation

Q. Hamp and L. Reindl

University of Freiburg - IMTEK
Department of Microsystems Engineering
Laboratory for Electrical Instrumentation
Georges-Köhler-Allee 103, D-79110 Freiburg
quirin.hamp@imtek.uni-freiburg.de

Abstract: This paper presents a method for multi target detection and localization based on heterogeneous uncertain information produced by multiple agents.

Information association is based on graph analysis which considers georeferences, spatial precision and pre-existing knowledge. The objective of the scoring fusion is to suggest as quickly and as precisely as possible the hypothetic positions of trapped persons by increasing the quality of uncertain information. The overall aim is to ameliorate the search efficiency by increasing the detection capabilities while reducing risks, false alarms and oversight.

1 Introduction

In Urban Search and Rescue (USAR) operations there is a static world of likely multiple and unknown targets. Since rescue is time critical, non-monotonic reasoning should occur after any belief revision. Current state of the art of IT-systems assisting disaster response is the georeferenced presentation of information in order to support the decision-making process [KKHK02, MIH+06], but to our knowledge automatic reasoning for the detection and localization of trapped victims is introduced with this project.

This paper presents a method for information fusion based on graph analysis. Filtering of information can not occur due to ethical considerations. In section 2, we present the crucial process of information association including cut capabilities. The inference machine consisting of classification and localization is explained in section 3.

2 Association

In order to understand association of uncertain information, the source of information i.e. the target has to be distinguished from the uncertain information about the source. Association consists of finding the unknown common information source. Since the link

between pieces of uncertain information and the source is often unknown, the least ambiguous link has to be found. This problem is also referred to multisensor multitarget association. During USAR operations the association framework can not necessarily neither rely on exhaustive search results of every search method for all position hypotheses nor on a known number of targets. Several methods exist (see [HM04],[Das08]), but they are mainly designed for static fusion, i.e. dynamic states.

A measurement might not be reproducible if it is based on physical variables which can vanish such as odor, heart beat or respiration. Another challenge is that hypotheses are supported by only a small number of observations which are collected on the fly. Our geographical data association can be divided in four phases illustrated by the dashed boxes in Fig. 1. Input and output of every process are represented by parallelograms. The primary input are reports or measures whereas the final output are regions of interest (ROIs).

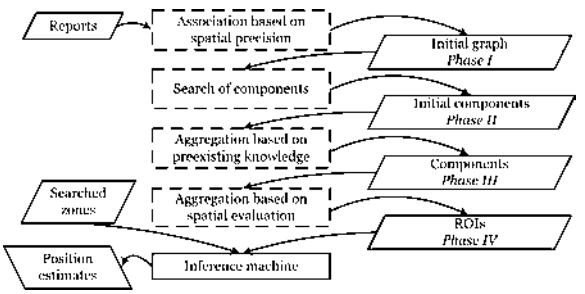


Figure 1: Process of Fusion and Inference

2.1 First phase

An initial graph is created. The spatial precision $\eta(\tilde{\mathbf{r}}_k)$ of a report is used to determine whether there is a link among them or not. An association is given if their surfaces intersect. A report's circular surface is given by its position and its spatial precision which define center and radius, respectively. This condition can be stated as in equation 1. A fulfilled association condition is represented schematically in Figure 2(c).

$$\left\| \tilde{\mathbf{r}}_i - \tilde{\mathbf{r}}_j \right\|_2 \leq \eta(\tilde{\mathbf{r}}_i) + \eta(\tilde{\mathbf{r}}_j) \tag{1}$$

2.2 Second phase

The initial graph represented by an incidence list is searched to identify connected components. A depth first search algorithm is applied to the incidence list to extract N_c components. A report which has no link to another is as well a component. The connected

components can not be considered as ROIs per se because they may grow infinitely in surface.

2.3 Third phase

An analysis of the identified components is undergone with respect to preexisting knowledge such as the expected number of trapped persons. If the expected number is bigger than the number of components ($N_c < \hat{N}$), each component is evaluated with respect to its split ability. The split is performed using the k-Means algorithm of Llyod [Llo82] assuming two clusters. The ranking of components is based on the mean distance to the centroid. As higher the mean distance as more probable is that the association hypothesis is based on more than one information sources. However, this does not take into account systematic errors.

Components with at least two vertices are split unless their mean distance is smaller than a given threshold μ_{min}. Figure 2(a) illustrates this case where the component's mean distance to the centroid is too small, in contrast to Fig. 2(b) where the component is splittable.

If the number of expected number of trapped persons is smaller than the number of components ($N_c > \hat{N}$) merging of components can not occur since the likelihood of an erroneous estimation can not be excluded.

2.4 Fourth phase

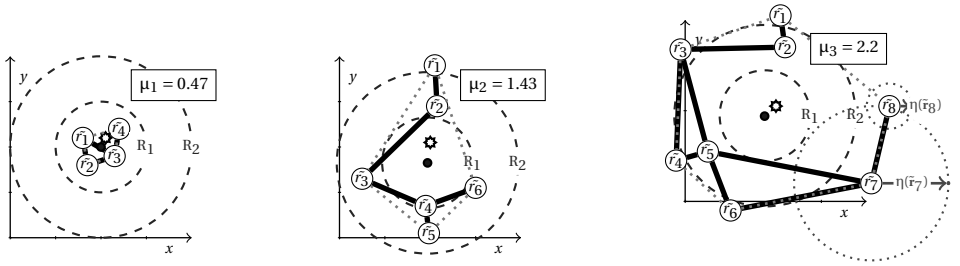
Components are processed with respect to the mean distance to the centroid, which is equivalent to a kind of gating method. If the mean is bigger than a predefined threshold μ_{max} (see Fig. 2(c)), the component is split using the same algorithm as in the third phase no matter how many victims are expected.

The final output (ROIs) correspond to clusters. Spatially a ROI is delimited by the convex hull bounding all vertices of a component.

3 Inference

Performance factors of search methods, ROIs and "*Searched zones*" (i.e. negative search results) are inputs of the inference machine performing the classification and conditionally localization. It is based on a heuristic method for scoring fusion [HM04].

Classification consists of an assessment of all available information within a ROI and after calculation of the *target score* **TS**, outputs whether a position hypothesis is rejected,



(a) $\mu_1 < \mu_{min}$ (b) $\mu_{min} < \mu_2 < \mu_{max}$ (c) $\mu_3 > \mu_{max}$, two dotted circles around $\tilde{r}_7$ and $\tilde{r}_8$ represent the fulfilled association condition

Figure 2: Schematic representation of component splitting cases for $\mu_1 < \mu_2 < \mu_3$ (*reports*: solid circles , *estimated positions* $\hat{\mathbf{x}}$: stars, *edges of component*: solid lines, *convex hulls*: dotted lines, *help circle*s (dashed circles) around *centroids* $\tilde{\mathbf{x}}$: filled dots)

confirmed or still unclear. **TS** expresses how likely the presence of a trapped person is at a given position.

If a search-action resulted in a report within the hull, **TS** increases. The number of reports for each search method $N_{r,i}$ is considered. On the contrary, if a search-action was fruitless within the hull and hence, no correlated report is present, **TS** decreases. The counter of searched zones $N_{s,i}$ is also relative to search methods.

The total number of reports is relevant: as more reports are emitted by search teams as higher is the target score. Eqs. 2 are used to assess the scale factor γ_r and γ_s with respect to the total number of reports N_r and negative search results N_s respectively. The denominator in the magnitude of five is empirically determined by the *Federal Agency for Technical Relief*.

$$\gamma_r = \frac{N_r}{5}, \quad \gamma_s = \frac{N_s}{5} \tag{2}$$

Equation 3 calculates the target score **TS** with a weighted average over the detection probabilities and the scale factors γ (see Eqs. 2). Index i iterates through the search methods and N_i is the total number of methods. This equation is based on an exponential recovery function to either increase or decrease **TS** by summing or subtracting from the starting score value of $1/2$, respectively. For a schematic representation of **TS** refer to Fig. 3.

$$\mathbf{TS} = \frac{1}{2} + \frac{1}{2}\left[1 - \exp\left(-\frac{\gamma_r}{N_r}\sum_{i=1}^{N_i} N_{r,i} P_i(t|\tilde{\mathbf{r}})\right)\right] - \frac{1}{2}\left[1 - \exp\left(-\frac{\gamma_s}{N_s}\sum_{i=1}^{N_i} N_{s,i} P_i(\neg t|\neg\tilde{\mathbf{r}})\right)\right] \tag{3}$$

A hard limiter function $f(\mathbf{TS})$ (see Eq. 4) gives the output u of the inference machine with $C_{min} \in [0, 0.5]$ and $C_{max} \in]0.5, 1]$ to be defined.

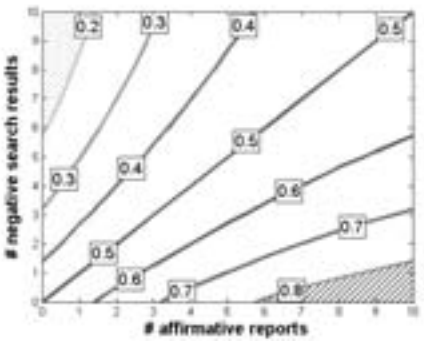


Figure 3: Equipotential plot of **TS** with respect to number of reports and negative search results, for a unique search method with $\mathrm{P}(t|\tilde{\mathbf{r}}) = \mathrm{P}(\neg t|\neg\tilde{\mathbf{r}}) = 0.8, \mathrm{C}_{min} = 0.2, \mathrm{C}_{max} = 0.8$; inference for striped area: *confirmed hypothesis*, for dotted one: *rejected hypothesis*

$$u = f(\mathbf{TS}) = \begin{cases} \text{confirmed } \mathrm{H}^p & \text{if } \mathbf{TS} > \mathrm{C}_{max} \\ \text{unclear } \mathrm{H}^p & \text{if } \mathrm{C}_{min} \leq \mathbf{TS} \leq \mathrm{C}_{max} \\ \text{rejected } \mathrm{H}^p & \text{if } \mathbf{TS} < \mathrm{C}_{min} \end{cases} \tag{4}$$

If the the target score is bigger than a threshold termed maximal certainty $\mathbf{TS} > \mathrm{C}_{max}$, the position hypothesis is confirmed and further actions should be triggered because the existence and the position of a trapped person is credible enough. On the other hand, if the target score is smaller than a threshold called minimal certainty $\mathbf{TS} < \mathrm{C}_{min}$, no trapped person is inferred. Between maximal and minimal certainty threshold not enough certainty can be attributed to the hypothesis.

Localization For ROIs with several reports N_r which are provided by a varying number of search methods N_i, the estimated position $\hat{x}$ based on reports under the condition of a confirmed position hypothesis $\{\tilde{\mathbf{r}}(x, y, z)|\mathrm{H}^p\}$ is computed with Eq. 6. The aim of this weighted average is to enhance the spatial precision of the position estimate of a confirmed position hypothesis. The confidence c_i is a weighing factor for each search method expressing how reliably the position is estimated. It has to be determined by experience η_i and by assessment of the current situation $\eta(\tilde{\mathbf{r}}_k)$. Eq. 5 expresses this confidence with a weighted sum. Both spatial precisions are normalized and κ expresses how much influence experience has over actual assessment. As represented in Figure 2 the estimated position does not necessarily correspond to the centroid.

$$c_i = \kappa \left\| \eta_i \right\| + (1 - \kappa) \left\| \eta(\tilde{\mathbf{r}}_k) \right\|, \; 0 \leq \kappa \leq 1 \tag{5}$$

$$\hat{\mathbf{x}} = \frac{1}{\sum_{i=1}^{\mathrm{N}_i} c_i} \sum_{i=1}^{\mathrm{N}_i} c_i \left[\frac{1}{\mathrm{N}_{r,i}} \sum_{j=1}^{\mathrm{N}_{r,i}} \tilde{\mathbf{r}}_{i,j} \right] \tag{6}$$

4 Conclusions

This paper presented a fusion method which is focusing on the association of measures to a common information source. The classification is ternary and accounts for negative search results. The estimated position is calculated with a weighted average considering the spatial precision of heterogeneous search methods. Hence, this method covers the entire fusion process. It is capable of dealing with heterogeneous search results, with multiple targets and with a dynamic fusion process required for static worlds.

Acknowledgment

We gratefully acknowledge financial support from the *German Federal Ministry of Education and Research* (BMBF) (support code: 13N9759). Furthermore, we are most grateful for the support of the *German Federal Agency for Technical Relief* (THW).

References

[Das08] Subrata Das. *High-Level Data Fusion*. Artech House, Inc., Norwood, MA, USA, 2008.

[HM04] David L. Hall and Sonya A.H. McMullen. *Mathematical Techniques in Multisensor Data Fusion*. Artech House, Inc., 2004.

[KKHK02] S Kakumoto, Y Kosugi, M Hatayama, and H Kameda. Development of spatial temporal geographic information system and risk adaptive regional management information system - Toward development of GIS based on Asian culture for disaster prevention. In *PROCEEDINGS OF THE 41ST SICE ANNUAL CONFERENCE*, volume 1-5, pages 352–357, 2002.

[Llo82] S. Lloyd. Least squares quantization in PCM. *IEEE Transactions on Information Theory*, 28(2):129 – 137, 1982.

[MIH+06] Jun-Ichi Meguro, Kiitirou Ishikawa, Takumi Hasizume, Jun-Ichi Takiguchi, Itsuki Noda, and Mitinori Hatayama. Disaster information collection into geographic information system using rescue robots. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, volume 1-12, pages 3514–3520, 2006.

Indirect Tracking of Patients in Mass Casualty Incidents

Simon Nestler, Tayfur Coskun, Eva Artinger, Peter Pichlmaier, Gudrun Klinker

Fachgebiet Augmented Reality
Technische Universität München
Fakultät für Informatik
Boltzmannstraße 3, 85748 Garching bei München
nestler|klinker@in.tum.de

Abstract:
In MCIs (mass casualty incidents) patient tracking is cumbersome when using paper based approaches. Due to the fact that the spatial positions of the patients are essential in MCIs, we equipped the relief workers with electronic devices. We analyzed two concepts for patient tracking: direct and indirect tracking. The indirect tracking turned out to be superior with regard to the requirements from Feuerwehr TUM (Technische Universität München). An implementation and evaluation of this concept was performed within the scope of the *SpeedUp* project[1]. This project focuses on the development of strategies for reacting in mass casualty incidents (MCIs) in an appropriate way. In this paper we describe the background, the requirements, the concepts, the implementation and the evaluation. Finally, the results are presented and discussed.

1 Background

In MCIs (mass casualty incidents) tracking of patients and relief workers is essential for the success of the crisis response. Tracking is the basis for presenting spatial information on mobile user-interfaces [NK07a]. Concepts for supporting the collaboration between relief workers and incident commanders have been discussed in [NEDK08] and [ENDK09]. General concepts for facilitating an overview in time-critical, life-threatening and unstable situations were presented in [NDP+09].

The first relief workers at the scene initially perform the MCI triage. Recording the patients' positions during triage is cumbersome when using paper based approaches. Nevertheless the spatial positions of the patients are essential for the organization of the crisis response [GHZ+06]. As soon as electronic devices are used by the relief workers in MCIs, the tracking of patients can be simplified a lot. In the past we performed different evaluations on the benefits of assisting relief workers with mobile devices, which have been

[1]The project SpeedUp is funded by the German Federal Ministry of Education and Research (BMBF) within the programme "Research for Civil Security" (May 1st, 2009 - April 30th, 2012, FKZ: 13N10175). Website: http://www.speedup-projekt.de

published previously: [NK07b], [End09] and [Pic09]. These publications focus on usability aspects of computer assisted triage. In this paper we show a complementary feature – the benefits of patient tracking for the incidents commanders.

Different publications deal with the usage of GPS tracking for rescue operations. The *Emergency Management System* from [RNSH09] uses direct tracking by equipping (1) incident commanders, (2) patients, (3) doctors and (4) relief workers with mobile devices. The research project *GÜTER* uses a GPS system for supervising transport routes. In the case of an MCI the spatial information is communicated to the relief units as described by [KHLP07]. In [SDL+07] an approach to simplify the dispatch of ambulances by using GPS tracking is presented. Furthermore, the tracking of all ambulances simplifies the optimization of the overall spatial distribution as discussed by [Hil06]. By tracking the position of emergency calls both, misuses and mistakes, can be significantly reduced according to [GBD05]. Furthermore, [SM06] is confident that GPS and video data can be transmitted from the emergency caller to the control center in the near future.

2 Requirements

In order to facilitate a better assignment of relief workers, the tracking of the patients has to fulfill certain requirements. These requirements have been identified in collaboration with Thomas Schmidt, deputy fire chief of Feuerwehr TUM (Technische Universität München)[2]. We identified five basic functional requirements: **(F1)** The position of the relief workers has to be gathered automatically in order not to distract them, **(F2)** it has to be possible to explicitly initiate the positioning of a certain patient, **(F3)** the completion of the patient's positioning has to be done automatically, **(F4)** the patients' positions have to be transmitted to the incident commanders over the wireless network and **(F5)** grouping on the basis of the tracking information has to be possible. Additionally we identified three non-functional requirements: **(N1)** the solution has to be scalable, **(N2)** the system has to be cost-effective and **(N3)** the maintenance efforts have to be low.

Furthermore, the Feuerwehr TUM had two remarks: **(R1)** The possibility to extend the patient's position with additional information (e.g. triage category) would be useful and **(R2)** the relief workers would benefit from a visual feedback whether the acquisition of the patient's position was successful.

3 Concepts and Implementation

This section analyses two concepts for patient tracking: direct and indirect tracking. Direct tracking determines the patients' positions by attaching devices to the patients. Indirect tracking deduces the patients' positions from the positions of the relief workers. The indirect tracking bases on the assumption that patient and relief worker are very close to

[2]The 60 employees of Feuerwehr TUM are responsible for the research center in Garching, working in three shifts with twenty-four-seven service

each other when the relief worker performs the MCI triage.

When using direct tracking the accuracy **(F5)** is ideal, the explicit initiation **(F2)** and the transmission of the positions is possible **(F4)**. In order to track the relief workers **(F1)**, they can also be equipped with GPS receivers. However, the scalability **(N1)**, cost-effectiveness **(N2)** and maintenance requirements **(N3)** are poor for direct tracking. Extending the paper based patient tag with an GPS receiver and a wireless communication unit raises the price per tag tenfold to hundredfold. These increased costs as well as the necessity to use power supplies reduces the scalability of the solution. When using the indirect tracking, the scalability **(N1)**, cost-effectiveness **(N2)** and the maintenance requirements **(N3)** are ideal. Due to the fact that relief workers are equipped with GPS **(F1)** and wireless communication **(F4)**, these two requirements can be fulfilled as well. In order to provide an explicit initiation **(F2)** and an automatic completion **(F3)** a basic user-interface is needed for the relief workers. The most critical requirement for indirect tracking is **(F5)**. The fulfillment of this requirement can only be completely proven by an evaluation within an MCI exercise. By extending the basic user-interface the input of additional information **(R1)** and the visual feedback **(R2)** can be supported as well.

This short examination shows, that the indirect tracking is more likely to fulfill the requirements from Feuerwehr TUM. Therefore, we realized a prototypic implementation of the indirect tracking and evaluated it in an MCI exercise. The implementation was done in C#[3] on handheld devices with *Windows Mobile*[4]. The result of the implementation is shown in Figure 1 and described in detail in [Pic09].



Figure 1: Indirect tracking of patients with *Windows Mobile* handheld devices

4 Evaluation

The evaluation is of central importance to prove that the requirements are fulfilled, especially the requirement on tracking accuracy (F5). The evaluation of the tracking accuracy was performed within the scope of an MCI triage training at Feuerwehr TUM. In summary 36 people were involved in the evaluation: 10 mimes, 16 relief workers, 4 supervisors, 2

[3] http://msdn.microsoft.com/de-de/vcsharp/default.aspx
[4] http://www.microsoft.com/windowsmobile/de-de/default.mspx

technical coordinators, 2 training coordinators and 2 photographers. The scenario of the evaluation was the following: On a highway an accident with four involved cars has happened. Three severely injured patients are in each of these cars. The cars (A, B, C and D) were spread over an area of 10 meters x 50 meters. Eight teams performed the MCI triage with three different types of mobile user-interfaces and the paper-based approach. In summary 32 MCIs were trained by the relief workers, in 24 of these MCIs tracking data could be collected for the 12 patients. Theoretically 24 x 12 = 288 patient positions could be generated during the whole evaluation.

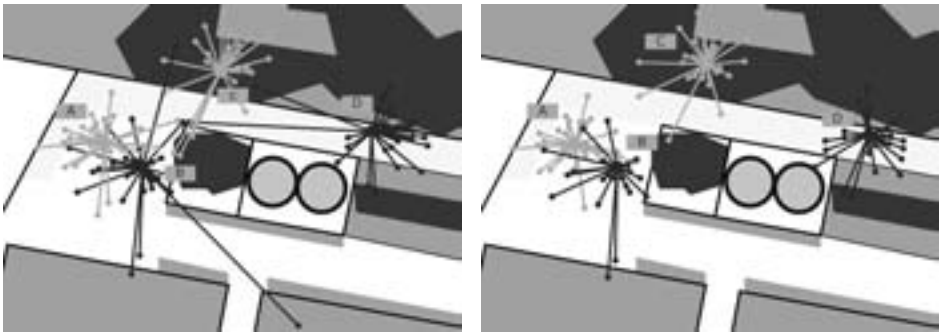
Organizational challenges as well as some technical challenges occurred during the evaluation. The technical coordinators were able to solve most of these technical issues in order to guarantee a proper course of the training. Due to the fact that low-budget handheld devices were used in this evaluation the batteries had to be changed several times during the training. The coupling of mobile handheld devices and GPS receivers via Blutooth turned out to be not that robust. In future evaluations ruggedized handheld devices with integrated GPS receivers will be used to avoid these hardware problems.

5 Results

In the MCI scenario the relative accuracy is more important than the absolute accuracy. Assumed that all GPS receivers have a divergence of 100 meters in the same direction, the relocation of patients is still possible. In the case, however, that one device has no divergence and all other devices have a divergence of 100 meters in the same direction, the relocation of patients is no longer possible. In the evaluation the positioning of the patients was only possible when the GPS receiver had a signal during the initiation of the positioning process. Instead of the mentioned 288 patient positions only 252 patient data sets were created and only 193 patient positions were relayed. There are three reasons for the reduced number of positions: (1) technical problems with the device, (2) technical problems with GPS receiver and (3) problems with the triage workflow. Some relief workers just forgot to initiate the positioning process and no patient position could be stored in the system.

The accuracy of the indirect tracking was influenced by three factors: (1) Influence of the indirect approach, (2) influence of the specific GPS receiver and (3) influence of GPS itself. The video recording of the evaluation showed that the relief worker is often not that close to the patient when initiating the positioning process. Regarding the influence of the specific GPS receivers significant differences between the three GPS receivers could be identified. Due to the fact that only the relative position is relevant, the median was calculated from all positions for each vehicle and all divergences were calculated relatively to the median. The mean divergence is 6,71 ($\pm$ 7,31) meters, the highest divergence is 43,5 meters as shown in Figure 2(a).

When analyzing the different devices isolated, the most inaccurate device has a mean divergence of 11,05 ($\pm$ 7,31) meters. The other two GPS receivers have an isolated error of 4,86 ($\pm$ 3,79) meters or rather 4,35 ($\pm$ 3,81) meters. The maximum divergence of



(a) GPS positions (all) (b) GPS position (corrected)

Figure 2: GPS positions of the patients and allocation to the vehicles

43,5 meters was also caused by the most inaccurate receiver, the other two receivers had a maximum divergence of 14,6 meters or rather 20,1 meters. An ANOVA[5] showed, that positioning is significantly ($> 95\%$) worse with the most inaccurate GPS receiver. The spatial illustration of the errors in Figure 2 demonstrate that no allocation to the vehicles is possible when using the positions from all receivers (Figure 2(a)). When excluding the inaccurate GPS receiver, however, as shown in Figure 2(b), an allocation to the vehicles is possible in cases when the distance between two vehicles is at least 25 meters. The patients of vehicle A and B are still grouped together.

6 Discussion

In summary three groups can be arranged on the basis of indirect tracking information: (1) Vehicle A & B, (2) vehicle C and (3) vehicle D. We discussed the results with Feuerwehr TUM and the grouping was considered to be ideal for three reasons: (1) regions are more important than vehicles, (2) A and B are within sight and (3) C and D are not within sight. In the future we have to analyze in what way the incident commanders can benefit from this indirect tracking approach.

The requirements could be fulfilled by the evaluated implementation in the following way: **(F1)** The positioning of the relief workers could be performed automatically, **(F2)** the positioning of the patients was initiated explicitly, **(F3)** the completion of the positioning was done automatically, **(F4)** the positions could be transmitted to the incident commanders – with the described limitations and **(F5)** grouping of patients is possible – assumed that only the most accurate GPS-devices are used. Furthermore, the scalability and cost-effectiveness is guaranteed by using low-cost RFID chips **(N1, N2)** and the maintenance efforts are low due to the fact that passive tags are used.

[5]The ANalysis Of VAriance (`http://www.statsoft.com/textbook/anova-manova`) tests whether the means of several groups are all equal

References

[End09] T. Endres. Unterstützung der Einsatzkräfte eines Rettungseinsatzes durch softwarebasiertes Informationsmanagement auf mobilen Geräten. Diplomarbeit, Technische Universität München, Department of Computer Science, Dezember 2009.

[ENDK09] F. Echtler, S. Nestler, A. Dippon und G. Klinker. Supporting Casual Interactions between Board Games on Public Tabletop Displays and Mobile Devices. *Personal and Ubiquitous Computing*, Oktober 2009.

[GBD05] A. Gries, M. Bernhard und B. Dirks. Leitbild für die Leitstelle – Gefahren im Einsatz – Notfallmedizin im DRG-Zeitalter. *Notfall & Rettungsmedizin*, 8(5):358–363, 2005.

[GHZ+06] W. Gutsch, T. Huppertz, C. Zollner, P. Hornburger, M. Kay, U. Kreimeier, W. Schäuble und K. Kanz. Initiale Sichtung durch Rettungsassistenten. Ergebnisse bei Übungen zum Massenanfall von Verletzten. *Notfall & Rettungsmedizin*, 9(4):384–388, Juni 2006.

[Hil06] E. Hildebrand. Effizientere Notrufbearbeitung durch integrierte Regionalleitstelle. *Notfall & Rettungsmedizin*, 9(5):473–477, 2006.

[KHLP07] R. Koch, R. Harnasch, B. Lee und J. Pottebaum. Rapid and Precise Mobile Data Processing for Fire Brigades and Rescue Services (SAFeR / GUTER / SHARE). *Lecture Notes in Computer Science*, 4558:1050, 2007.

[NDP+09] S. Nestler, M. Duschl, D. Popiv, M. Rakic und G. Klinker. Concept for visualizing concealed objects to improve the driver's anticipation. In *17th World Congress on Ergonomics*, Beijing, China, August 2009.

[NEDK08] S. Nestler, F. Echtler, A. Dollinger und G. Klinker. Collaborative problem solving on mobile hand-held devices and stationary multi-touch interfaces. In *PPD'08. Workshop on designing multi-touch interaction techniques for coupled public and private displays*, May 2008.

[NK07a] S. Nestler und G. Klinker. Towards adaptive user-interfaces: Developing mobile user-interfaces for the health care domain. In *Mobiles Computing in der Medizin (MoCoMed)*, Augsburg, Germany, September 2007.

[NK07b] S. Nestler und G. Klinker. Using Mobile Hand-Held Computers in Disasters. In *UbiComp Workshop on Interaction with Ubiquitous Wellness and Healthcare Applications (UbiWell)*, Innsbruck, Austria, September 2007.

[Pic09] P. Pichlmaier. Software-basierte Planung und Informationsmanagement auf mobilen Geräten zur Unterstützung der Leitung von Rettungseinsätzen. Diplomarbeit, Technische Universität München, Department of Computer Science, Dezember 2009.

[RNSH09] A. Rausch, D. Niebuhr, M. Schindler und D. Herrling. Emergency Management System. In *Proceedings of the International Conference on Pervasive Services 2009 (ICSP 2009)*, 2009.

[SDL+07] T. Schlechtriemen, B. Dirks, C.K. Lackner, H. Moecke, D. Stratmann, H. Krieter und K.H. Altemeyer. Leitstelle – Perspektiven für die zentrale Schaltstelle des Rettungsdienstes. *Notfall & Rettungsmedizin*, 10(1):47–57, 2007.

[SM06] R. Stastny und M. Merka. Next Generation Emergency Services: Die Zukunft der Notrufe. *e & i Elektrotechnik und Informationstechnik*, 123(7):323–332, 2006.

Hermes – Erforschung eines Evakuierungsassistenten für den Krisenfall bei Großveranstaltungen

Stefan Holl, Armin Seyfried, Maik Boltes

Jülich Supercomputing Centre
Forschungszentrum Jülich GmbH
52425 Jülich, Germany
st.holl@fz-juelich.de
a.seyfried@fz-juelich.de
m.boltes@fz-juelich.de

Abstract: Das vom Bundesministerium für Forschung und Bildung geförderte Forschungsprojekt Hermes soll die Sicherheit der Personen in großen, multifunktionalen Bauwerken wie auch bei Großveranstaltungen im Freien durch die Erforschung eines Evakuierungsassistenten verbessern. Ziel des Projektes ist es, Echtzeit-Simulationen für bis zu 60.000 Personen durchzuführen und den Einsatzkräften von Polizei, Feuerwehr und Sicherheitsdienst eine Prognose zur Verfügung zu stellen, an welchen Stellen es in den nächsten Minuten zu gefährlichen Stauungen kommen wird. An dem interdisziplinären Forschungsvorhaben mit einer dreijährigen Laufzeit sind Partner aus Forschung und Wirtschaft sowie die vor Ort verantwortlichen Einsatzkräfte beteiligt. Im Mai 2011 wird die Testphase in der ESPRIT Arena in Düsseldorf beginnen.

1 Motivation

Der Trend zu großen, multifunktionalen Bauwerken wie auch die zunehmende Anzahl von Großveranstaltungen im Freien stellt neue Anforderungen an die Qualität der Sicherheitskonzepte. Im Notfall ist es erforderlich, dass alle anwesenden Personen den Gefahrenbereich schnell verlassen können. Zwar wird dies im Regelfall durch Anwendung der rechtlichen Vorgaben gewährleistet, bei Überbelegungen oder Ausfall einzelner Rettungswege kann es jedoch zu gefährlich hohen Personendichten und lang anhaltenden Stauungen kommen.

Die Erfahrung der Einsatzkräfte vor Ort zeigt, dass es bei Großveranstaltungen kaum möglich ist, die Anzahl der Personen in einem Bereich zuverlässig festzustellen. Für die Sicherheit der Menschen ist es daher erforderlich, neue Strategien und Hilfsmittel zu entwickeln, welche sowohl die Anzahl der anwesenden Personen erfassen als auch die optimale Räumung eines Gefahrenbereichs unterstützen können.

2 Konzept des Evakuierungsassistenten

Der Evakuierungsassistent basiert auf einer Echtzeit-Simulation der Personenströme, bei welcher neben den geometrischen Randbedingungen auch die aktuellen Daten der automatischen Personenzählanlage sowie bei Veranstaltungen in Gebäuden die Informationen des Gefahrenmanagementsystems berücksichtigt werden. Die Ergebnisse der Evakuierungssimulation werden durch ein Kommunikationsmodul visualisiert. Den Einsatzkräften von Polizei, Feuerwehr und Sicherheitsdienst steht damit eine aktuelle Prognose zur Verfügung, an welchen Stellen es innerhalb der nächsten Minuten zu gefährlichen Stauungen kommen kann.

Für das Forschungsvorhaben konnte der Betreiber der ESPRIT Arena in Düsseldorf als Kooperationspartner gewonnen werden. Am Beispiel dieser Multifunktionsarena mit einer Kapazität für bis zu 66.000 Zuschauer soll gezeigt werden, wie bei Großveranstaltungen die Menschenmassen unter Berücksichtigung der konkreten Gefahrenlage so geführt werden können, dass eine bestmögliche Ausnutzung der Rettungswege erreicht wird. Ab November 2010 soll der Evakuierungsassistent in die ESPRIT Arena eingebaut und ab Mai 2011 in einer halbjährigen Testphase hinsichtlich seiner Funktion und seines Nutzens überprüft werden.

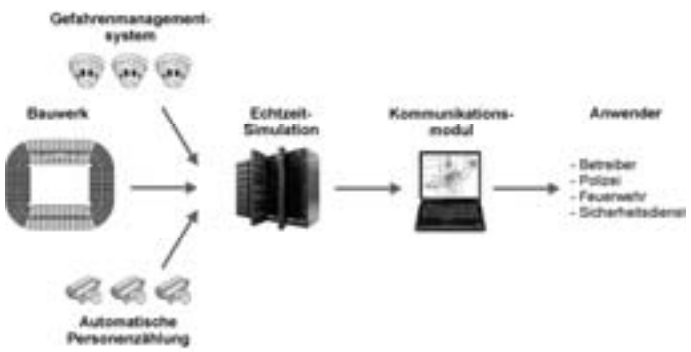


Abbildung 1: Prinzip des Evakuierungsassistenten

3 Berücksichtigung aktueller Daten

Für die Echtzeit-Simulation der Personenströme stehen dem Evakuierungsassistenten aktuelle Daten aus der automatischen, videobasierten Personenzählanlage zur Verfügung. Das Bauwerk wird hierfür in Zählbereiche unterteilt, an deren Übergängen Videokameras mit einer integrierten Rechner-Einheit (embedded system) installiert sind. Die Rechner-Einheit wird zur Identifikation von Personen innerhalb der Videobilder genutzt. Damit brauchen von den Kameras statt der datenintensiven Videosignale ausschließlich die relevanten über ein definiertes Zeitintervall integrierten Zählwerte für die jeweilige Laufrichtung übertragen werden. Die Information über die aktuellen Belegungszahlen

steht den Anwendern unmittelbar zur Verfügung und fließt gleichzeitig in die Evakuierungssimulation ein.
Der Zustand des Bauwerks wird durch das vorhandene Gefahrenmanagementsystem erfasst. Über die Brand- und Gefahrenmeldeanlage sowie ggf. durch manuelle Eingabe seitens der Anwender steht dem Evakuierungsassistenten die Information über die Nutzbarkeit von Bereichen, Türen etc. zur Verfügung.

4 Modellierung, Simulation und Parallelisierung

Die der Räumungsprognose zu Grunde liegenden Simulationen basieren auf einer Kombination aus „makroskopischem“ Netzwerkmodell [SNS10] und „mikroskopischen“ Modellen der Fußgängerdynamik. Bei den mikroskopischen Modellen kommen sowohl Zellularautomaten wie auch raumkontinuierliche Modelle zum Einsatz.

Das Jülich Supercomputing Centre am Forschungszentrum Jülich hat die Aufgabe übernommen, die raumkontinuierlichen Modelle der Fußgängerdynamik, in welchen die individuelle Bewegung der Personen anhand eines Systems gekoppelter Differentialgleichungen berechnet wird, weiter zu entwickeln [C09, PS10]. Im Ansatz ähneln diese Modelle denen der Molekulardynamik. Fußgänger werden allerdings als selbstgetriebene Teilchen in einem kontinuierlichen Raum dargestellt, in welchem sie untereinander und mit der Raumgeometrie (Wände, Türen, Hindernisse etc.) wechselwirken.

Bisherige Forschungen in diesem Gebiet konzentrierten sich auf die qualitative Reproduktion von Selbstorganisationsphänomenen, wie die Linienbildung in Engstellen oder bidirektionalen Personenströmen und sind bis heute nur ansatzweise validiert. Teilziel des Forschungsprojektes Hermes ist es, die bisherigen Ansätze zur Modellierung der Fußgängerdynamik soweit zu entwickeln, dass Personenströme in beliebigen Anlagentypen (Flure, Treppen, Engstellen etc.) nicht nur qualitativ, sondern auch quantitativ richtig beschrieben werden. Bislang liegt jedoch noch keine ausreichende und widerspruchsfreie experimentelle Datenbasis vor [S09].

Um zuverlässige Daten für die Kalibrierung und Validierung der Modelle zur Verfügung zu haben, führen die Universität Wuppertal, die Universität zu Köln und das Forschungszentrum Jülich bereits seit 2005 Experimente unter Laborbedingungen durch [Se05, Se08, Se09]. Im Projekt Hermes wurden diese Arbeiten durch Experimente in der ESPRIT Arena sowie der Messe Düsseldorf, an denen bis zu 350 Probanden mitgewirkt haben, fortgeführt. Durch den Einsatz eines am Forschungszentrum Jülich entwickelten Verfahrens zur automatischen Extraktion von Trajektorien (Raum-Zeit-Kurven) aus Videodaten [B10] ist es möglich, die individuellen Laufwege jeder einzelnen Person zu beschreiben. Die Trajektorien liefern uns die Grundlage für die Optimierung und Verifizierung der verwendeten Modelle.



Abbildung 2: Automatische Extraktion der Trajektorien

Für die Analyse der experimentellen Daten konnte zusätzlich ein neues Messverfahren entwickelt werden, welches auf einer Voronoi-Zerlegung des Raumes beruht [SS10, LSS10]. Dieses Verfahren ermöglicht es, Daten aus Experiment und mikroskopischer Simulation zu vergleichen.

Wesentlicher Baustein für die Realisierung des Evakuierungsassistenten ist die effiziente Implementierung der Simulationsmodelle [Se10]. Die Laufzeitanforderungen an eine Simulation schneller als Echtzeit für mehr als 60.000 Personen in einer komplexen Geometrie machen den Einsatz modernster Prozessortechnologie wie System-on-Chip (SoC) und Fließkommabeschleuniger unabdingbar. Die Nutzung von Nehalem-Prozessoren verspricht, eine entsprechende Beschleunigung der Simulationen zu ermöglichen. Allerdings müssen die üblichen Parallelisierungskonzepte aus der Molekulardynamik hierfür erweitert und an die verwendete Systemarchitektur angepasst werden. Der parallele Einsatz mehrerer solcher Systeme ermöglicht schließlich eine automatisierte Abschätzung der Zuverlässigkeit der Vorhersagen. Darüber hinaus soll eine automatische Anpassung der Simulation anhand der realen Personenzählung erfolgen.

5 Input/Output über das Kommunikationsmodul

Entscheidend für den Nutzen und die Akzeptanz des Evakuierungsassistenten ist die Aufbereitung der berechneten Informationen für die verantwortlichen Einsatzkräfte. Im Projekt Hermes ist geplant, einen Front-End-Rechner im Bereich der „Skybox“[1] der ESPRIT Arena zu installieren. In einem speziell eingerichteten Lagebesprechungsraum werden die Informationen aus dem Evakuierungsassistenten auf einem großformatigen Bildschirm visualisiert. Ein speziell ausgebildeter Operator hat zusätzlich die Möglichkeit, über das Kommunikationsmodul manuelle Eingaben in den Evakuierungsassistenten vorzunehmen oder – bspw. bei einer spontanen Lageänderung – den aktuellen Simulationslauf abzubrechen und eine neue Simulation mit aktualisierten Randbedingungen zu starten.

6 Kooperationspartner im Projekt Hermes

- Forschungszentrum Jülich GmbH, Jülich Supercomputing Centre (JSC), Verbundkoordination
- Bergische Universität Wuppertal, Abteilung Bauingenieurwesen, Lehrstuhl für Baustofftechnologie und Brandschutz
- Imtech Deutschland GmbH & Co. KG
- Multifunktionsarena Immobiliengesellschaft mbH & Co. KG (ESPRIT Arena)
- PTV Planung Transport Verkehr AG
- TraffGO HT GmbH
- Rheinische Friedrich-Wilhelms-Universität Bonn, Geographisches Institut
- Universität zu Köln, Institut für Theoretische Physik
- Vitracom AG
- Feuerwehr Düsseldorf
- Landesamt für Zentrale Polizeiliche Dienste NRW und Polizeipräsidium Düsseldorf
- Special Security Service SSSD GmbH

[1] In der Skybox befinden sich die Räume für die Entscheidungsträger der Einsatzkräfte, den Stadionsprecher etc. Von der Skybox aus ist der gesamte Innenbereich der Arena überschaubar.

Literaturverzeichnis

[B10] Boltes, M. et al.: Automatic Extraction of Pedestrian Trajectories from Video Recordings. In (Schreckenberg, M. et al. Hrsg.): Pedestrian and Evacuation Dynamics 2008, Springer, Berlin, 2010, 43-54

[C09] Chraibi, M. et al.: Quantitative Description of Pedestrian Dynamics with a Force-based Model. 2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology, 2009, 3, 583-586

[LSS10] Liddle, J.; Seyfried, A; Steffen, B.: Analysis of bottleneck motion using Voronoi diagrams. Contribution to the PED 2010, Washington, March 2010

[PS10] Portz, A.; Seyfried, A.: Analyzing Stop-and-Go Waves by Experiment and Modeling. Pedestrian and Evacuation Dynamics 2010 (to appear)

[S09] Schadschneider, A. et al.: Evacuation Dynamics. Empirical Results: Modeling and Applications. In (Meyers, R. A. Hrsg.): Encyclopedia of Complexity and System Science, Springer, Berlin, 2009, 5, 3142-3176

[Se05] Seyfried, A. et al.: The fundamental diagram of pedestrian movement revisited. J. Stat. Mech., 2005, P10002

[Se08] Seyfried, A. et al.: Enhanced empirical data for the fundamental diagram and the flow through bottlenecks. In (Schreckenberg, M. et al. Hrsg.): Pedestrian and Evacuation Dynamics 2008, Springer, Berlin, 2010, 145-156

[Se09] Seyfried, A. et al.: New insights into pedestrian flow through bottlenecks. Transportation Science, 2009, 43, 395-406

[Se10] Seyfried, A. et al.: Runtime Optimization of Force Based Models within the Hermes Project. Pedestrian and Evacuation Dynamics 2010 (to appear)

[SNS10] Schomborg, A; Nökel, K.; Seyfried, A.: Evacuation Assistance for a Sports Arena Using a Macroscopic Network Model. Pedestrian and Evacuation Dynamics 2010 (to appear)

[SS10] Steffen, B.; Seyfried, A.: Methods for measuring pedestrian density, flow, speed and direction with minimal scatter. Physica A, accepted for publication, http://dx.doi.org/10.1016/j.physa.2009.12.015 (2010).

Räumungsszenarien bei Großveranstaltungen: Modellierung einer Datenbasis für Planung und Forschung*

Steffen Schneider, Jens Pottebaum, Tobias Becker, Rainer Koch

Computeranwendung und Integration in Konstruktion und Planung (C.I.K.)
Universität Paderborn, Fakultät Maschinenbau
Warburger Str. 100
33098 Paderborn
{st.schneider, pottebaum, t.becker, r.koch}@cik.uni-paderborn.de

Abstract: Die Planung einer Großveranstaltung beruht häufig stark auf Erfahrungswerten und weniger auf wissenschaftlichen Erkenntnissen. Aus diesem Grund widmet sich diese Arbeit der Fragestellung, wie die Planung von Großveranstaltungen durch einen geeigneten Forschungsansatz ergänzt werden kann, um die Planungsprozesse effizienter zu gestalten. Durch eine Use Case-Analyse werden im Wesentlichen zwei Gruppen von Bedarfsträgern identifiziert: Der Planer, der für die Vorbereitung der Großveranstaltung verantwortlich ist, und der Analyst bzw. Forscher, der die Veranstaltung selbst und das dafür verwendete Planungskonzept untersucht. Aufbauend auf den Anwendungsfällen wird ein konzeptuelles Schema der Datenbasis für ein IT-System vorgestellt, welches sowohl den direkten Anwendungsfall (Planung) als auch die darauf aufbauende Prozessanalyse (Forschung) unterstützt.

1 Einleitung

Schutz und Rettung von Menschen stellen die oberste Priorität der planerischen Aufgaben der Behörden und Organisationen mit Sicherheitsaufgaben (BOS) dar. Dies erfordert bei den immer komplexer werdenden Gegebenheiten ein optimiertes Informationsmanagement, das insbesondere einsatztaktische Aspekte berücksichtigt und Erfahrungsdaten einbezieht. Oftmals bleibt aber bei dessen Entwicklung die analytische Perspektive zur Prozessoptimierung unberücksichtigt. In dieser Arbeit wird ein Konzept vorgestellt, um sowohl für die Aufgabe der Sicherheitsplanung als auch für die vor- und nachbereitende Aufgabe der einsatzbezogenen und -übergreifenden Analyse eine IT-Unterstützung bereitzustellen. Explizit soll dabei die Forschungsgemeinschaft als Teil des Stakeholderkreises Berücksichtigung finden.

*Dieser Beitrag präsentiert Ergebnisse des Projektes "EVA: Risiko Großveranstaltungen - Planung, Bewertung, EVAkuierung und Rettungskonzepte" (Förderkennzeichen 13N10303), welches im Bereich "Schutz und Rettung von Menschen" des Programms "Forschung für die zivile Sicherheit" im Rahmen der High-Tech-Strategie der Bundesregierung durch das Bundesministerium für Bildung und Forschung (BMBF) gefördert wird. Die Autoren danken dem BMBF für die Unterstützung des Projektes sowie dem EVA-Projektverbund.

Bedarfsgesteuert stellt dabei die Vorbereitung und Unterstützung von Großveranstaltungen durch BOS den konkreten Anwendungsfall dar: „Die Entwicklung der gesellschaftlichen Bedürfnisse und Wünsche hat in den letzten Jahren zu einem merklichen Anstieg an Versammlungsstätten für unterschiedliche Zielgruppen geführt, wo auf immer 'optimaler' genutzten Grundrissen erhebliche Menschenmassen zusammenkommen“ [Sch10]. Dies impliziert jedoch, dass die Planung zur Gewährleistung der Sicherheit aller beteiligten Personen immer komplexer wird. Insbesondere der Aspekt der Räumung von Objekten und Flächen wird in der Forschung intensiv diskutiert und in der empirischen Brandschutzforschung seit einigen Jahren adressiert (vgl. z.B. [SSL04, SKK+09, SK10]). Die bisherigen Planungskonzepte beruhen zum Großteil auf persönlichen Erfahrungen, werden nur im geringen Maße durch Informationstechnik unterstützt und sind im Nachhinein nur schwerlich unter der Zielsetzung der Optimierung analysierbar.[1]
Die hier beschriebene Arbeit basiert auf der Erkenntnis, dass verfügbare Informationstechnologien hohes Potential zu einer Optimierung in diesem Bereich beinhalten. Dies soll mittels eines Informationssystems genutzt werden, dessen Kern eine Datenbasis zur übergreifenden Nutzung durch alle Stakeholder ist. Im Folgenden wird ein entsprechendes konzeptuelles Schema der Datenbasis vorgestellt, welches die Grundlage zur Eingabe von Daten und zur Bereitstellung von Informationen und Handlungsempfehlungen in diesem Bereich bildet. Da die vollständige Darstellung des Datenmodells den Rahmen des Beitrages deutlich überschreiten würde, wird hier lediglich das Konzept vorgestellt[2].

2 Vorgehen

Zwei wichtige Gruppen von Bedarfsträgern wurden einleitend bereits hervorgehoben: Anwender und Forscher müssen bei der Analyse von Anforderungen an ein Datenmodell berücksichtigt werden. Beide Gruppen führen dabei planerische und analytische Aufgaben durch, die jedoch auf Grund ihrer eigenen Zielabsichten unterschieden werden müssen. Daraus ist zu folgern, dass die Akteure *Sicherheitsplaner* und *Forscher* sowie die Aufgaben *Planung* und *Analyse* unterschieden werden müssen.
In Anlehnung an Kemper und Eickler ([KE09], S. 34) ist die Anforderungsanalyse für Datenbanksysteme in die Betrachtung von Informations- und Datenverarbeitungsanforderungen zu gliedern. Die Ergebnisse der Anforderungsanalyse sowie die Informationsanforderungen im Speziellen werden in einen konzeptuellen Entwurf überführt. Der Begriff der Information und somit der Informationsanforderung impliziert dabei, dass die Handlung des Bedarfsträgers in die Analyse einbezogen werden muss. Es kann festgestellt werden, dass die beiden vorgestellten Akteure in der Ausführung der beiden zentralen Aufgaben generell komplementäre Ziele verfolgen. Diese können grob in die optimale Planung von Großveranstaltungen und den Erkenntnisgewinn bezüglich verschiedener Forschungsfragen unterschieden werden; ihre Erreichung hängt jedoch von einer gemeinsamen Datenbasis ab. Für die Anforderungsanalyse wurde daher das Vorgehen um Aspek-

[1] Schlussfolgerung aus der Analyse von Interviews mit Verantwortlichen der Feuerwehr und Auswertungen von Workshopbeiträgen im Projekt EVA.

[2] Die Dokumentation des vollständigen Datenbankmodells kann bei den Autoren nachgefragt werden.

te des ziel- und szenariobasierten Requirements Engineering nach Pohl (vgl. [Poh08]) erweitert. Im Rahmen von Interviews mit Stakeholdern wurden Zielsetzungen und dazu aktuelle Vorgehensweisen erhoben. In den Ergebnissen dieser Analyse können Use Cases für ein Datenbanksystem identifiziert werden: Beide Akteure müssen Daten abgelaufener Großveranstaltungen pflegen, abrufen und mit den zugehörigen Planungsdaten und -konfigurationen vergleichen. Such- und Auswertungskriterien sind dabei teilweise gleich (z.B. Planungsdaten und allgemeine Statistik), teilweise unterschiedlich (z.B. spezifischer Abruf der Trajektorien einzelner Gruppen mit gemeinsamen Merkmalen) und betreffen insbesondere diejenigen Parameter, die in der Planung bzw. in der Analyse auf Annahmen beruhen. Der Sicherheitsplaner muss Berechnungen zur Gewährleistung der Sicherheit bei einer Großveranstaltung durchführen und Parameter für entsprechende Algorithmen bewerten. Er muss darauf aufbauend Veranstaltungsdaten mittels seiner Vorschriften und Arbeitsanweisungen eingeben. Der Abruf von Maßnahmeempfehlungen unterstützt bei der räumlichen, funktionalen und kräftebezogenen Disposition. Bei Integration eines Simulationswerkzeugs in die Arbeitsabläufe müssen beide Akteure dieses Werkzeug mit allen notwendigen Parametern initialisieren. Der Sicherheitsplaner nutzt dies insbesondere dazu, die Simulationsergebnisse hinsichtlich erforderlicher Maßnahmen auszuwerten.

3 Ergebnisse

Abbildung 1 zeigt das konzeptuelle Schema der umzusetzenden Datenbank. Dieses beruht auf einer Zerlegung in sieben Kategorien, deren untergliederte Entitäten wiederum in Abhängigkeiten untereinander stehen (Pfeile).

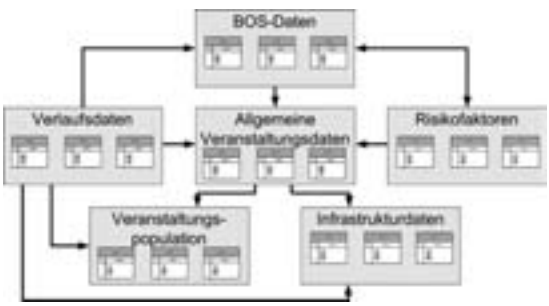


Abbildung 1: Konzeptuelles Schema der Datenbasis

Im Zentrum der Datenbank stehen die **allgemeinen Veranstaltungsdaten**, zu denen insbesondere die spezifizierenden Daten (z.B. Veranstaltungsdauer und -ort) und Annahmen wie das erwartete Besucheraufkommen zählen. Die Daten werden von einem Anwender während des Prozesses der Veranstaltungsplanung direkt eingegeben. Diese Daten werden zum einen für die Identifikation von **Risikofaktoren**, zum anderen zur Abbildung von **Verlaufsdaten** benötigt. Während erstgenannte die Veranstaltungsdaten auf Basis definierter Regeln bzw. Algorithmen aus wissenschaftlichen Erkenntnissen und Erfahrungswissen auf mögliche Risiken abbilden, ergänzen letztere die statischen Daten um die zeit-

lich veränderliche Datenlage im Veranstaltungsverlauf. Hier ist insbesondere die Veranstaltungspopulation und der Verlauf von BOS-Einsätzen zu nennen. Die **BOS-Daten** charakterisieren die Akteure, welche für Schutz und Rettung von Menschen wirken. Dabei müssen nur die veranstaltungsspezifischen Daten durch den Sicherheitsplaner eingegeben werden. Für eine vollständige Datenbasis muss ein Informationssystem hier Daten- und Systemschnittstellen zu existierenden IT-Systemen vorsehen. Die BOS-Daten bilden die Grundlage für Einsatzdaten, die innerhalb der Kategorie Verlauf abgelegt werden. Zudem wirken die abgebildeten Elemente der BOS auf verschiedene Risikofaktoren, die im Umkehrschluss Auswirkungen auf Maßnahmen der Aufstellung von Sicherheitskräften haben. Risikofaktoren werden im Rahmen dieser Arbeit anhand des Berechnungsverfahrens „Kölner Algorithmus“ veranschlagt, der aus diesen Empfehlungen kalkuliert[3]. Die **Infrastrukturdaten** sind für die Sicherheitsplanung einer Großveranstaltung unverzichtbar.[4] Objektbezogene und räumliche Daten müssen einbezogen und Akteuren in der Planung und der Analyse zur Verfügung gestellt werden. Die Informationen können sowohl Verlaufs- als auch Veranstaltungsdaten näher spezifizieren. Zudem bilden sie die Basis, um eine Kubatur für eine Personenstromanalyse zu generieren. **Populationsdaten** beschreiben die zu erwartenden Besuchergruppen und deren Anteil an der Gesamtpopulation. Die Informationen diesbezüglich werden im Vorfeld als Standardwerte generiert und müssen nicht vom Sicherheitsplaner eingegeben werden. In der Analyse haben Sicherheitsplaner und Forscher die Möglichkeit, auf eine konkrete Veranstaltung oder eine Veranstaltungsart bezogene Daten auszuwerten und daraus Erkenntnisse zu gewinnen bzw. Standardwerte zu optimieren. Die dadurch zur Verfügung stehenden Informationen bieten sowohl einen Überblick über die mit dem Personenstrom verbundenen Gefahrenpotentiale als auch Eingabematerial für eine Personenstromanalyse.
Die Verlaufsdaten sind an dieser Stelle herauszuheben, da sie in der Nachbereitung von Einsätzen, Wetter- und Verkehrsveränderungen ausgewertet werden können und somit beiden Akteuren erst in der Analyse Nutzen bringen.

4 Diskussion

Die Bereitstellung eines fundierten Datenmodells führt zu einer Standardisierung, die eine verlässliche Grundlage für die Gegenüberstellung von verschiedenen Veranstaltungsplanungen bildet. Ergebnisse einer wissenschaftlichen Untersuchung können dazu beitragen, das System und die Planungsmethoden zu optimieren (verbesserte Algorithmen, exaktere Parameter, neue Standardwerte etc.). Eine übergreifende Nutzung von Datenbasen ist notwendig, da zur optimalen Gewährleistung der Sicherheit ein hohes Maß an Interdisziplinarität notwendig ist. Langfristiges Ziel ist die Validierung (vgl. [Kro00] und DIN EN ISO 17025) der eingesetzten Planungsmethoden, um dem Sicherheitsplaner eine verlässliche

[3] Dieser Algorithmus wird derzeit von vielen Sicherheitsplanern von Feuerwehren in Deutschland angewendet, so dass sich das System in den Planungsprozess verschiedener Feuerwehren einfach integrieren lässt. Vgl. [Tec10]

[4] Aussage aus Interviews mit Verantwortlichen der Sicherheitsplanung. Die Eingabe und Pflege der Daten fällt in den Bereich des vorbeugenden Brandschutzes. Eine Referenzierung von Datensätzen aus diesem Bereich wird in diesem System gewährleistet.

Grundlage bereitzustellen. Dabei muss die Verlässlichkeit als Qualitätskriterium begründet vermittelt werden. Als Planungsmethoden können neben dem Kölner Algorithmus weitere Berechnungs- oder Simulationsverfahren eingebunden werden.

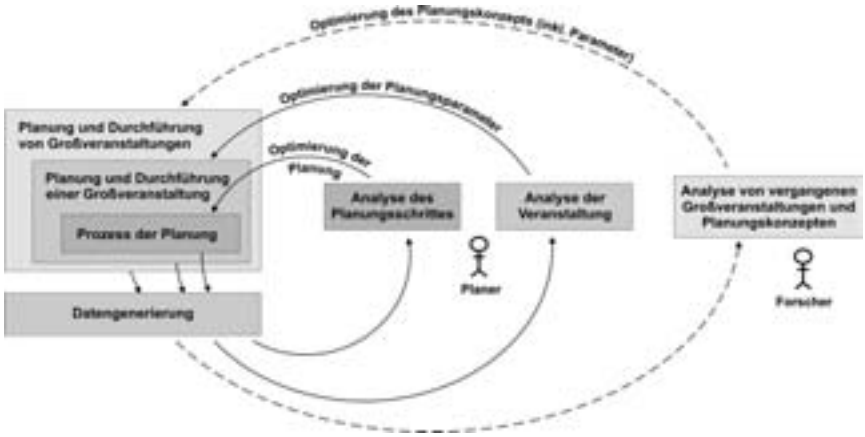


Abbildung 2: Planung und Analyse - eine ”Win-Win-Beziehung”

Im Sinne eines kontinuierlichen Verbesserungsprozesses (s. [Pfe01]) soll hier ein geschlossener Regelkreis entstehen. Die Abbildung 2 zeigt drei Kreisläufe, die durch den Prozess der ”Datengenerierung” verbunden sind. Die standardisierte Dateneingabe vereinfacht den Prozess der ”Planung einer Großveranstaltung”. Informationen zum Gefahrenpotential und zum Evakuierungsstrom sowie automatisch generierte Handlungsempfehlungen unterstützen den Planer bei der Planung und Analyse. Dieser innere Kreislauf wird solange iterativ durchlaufen, bis die Veranstaltung hinreichend genau geplant ist. Im Nachgang einer Großveranstaltung sollten alle relevanten Planungsdaten mit dem tatsächlichen Ablauf verglichen werden, um die Planung und Durchführung einer vergleichbaren zukünftigen Großveranstaltung zu verbessern. In erster Linie bietet sich hier die Chance, Planungsparameter zu optimieren (mittlerer Kreislauf).
Alle gesammelten Planungs- und Verlaufsdaten können vom Forscher dazu genutzt werden, vergangene Großveranstaltungen und die dafür verwendeten Planungskonzepte zu analysieren. Dadurch können Planungskonzepte und -parameter verbessert werden (äußerer Kreislauf). Die beiden Akteure Sicherheitsplaner und Forscher gehen damit eine ”Win-Win-Beziehung” ein. Der Forscher erhält Zugriff auf eine große Menge von Planungs- und Verlaufsdaten, der Sicherheitsplaner profitiert von einem verbesserten Planungssystem (neue Algorithmen, neue Vorgabedaten, etc.) mit validiertem Planungskonzept.
Bei der kontinuierlichen Verbesserung des Planungssystems ist jedoch zu hinterfragen, ob die in dieser Arbeit akzentuierte Vergleichbarkeit noch gegeben ist. In der Analyse von Großveranstaltungen, die mit aktuellen oder aber älteren Konzepten geplant wurden, könnten sich bei ansonsten gleichen Rahmenbedingungen Abweichungen ergeben. Dies ist ein Punkt, der bei der Modellierung derart berücksichtigt wurde, dass Konfigurationen der Algorithmen nachgehalten werden.

5 Ausblick

Das in diesem Beitrag vorgestellte Konzept soll zu einer Einbeziehung der verschiedenen Akteure motivieren. Die Hinzunahme der Perspektive des Forschers auf Meta-Ebene bei Betrachtung von grundlegend analogen Aufgabenstellungen bewirkt im dargelegten Anwendungsfall eine direkte Verbesserung des Gesamtsystems. Dies legt nahe, Systementwicklungen in dieser Anwenderdomäne auf diese Fragestellung hin zu prüfen. Zudem könnte die Integration weiterer Akteure eine weitere Optimierung im Hinblick auf die Anwendung der Software bringen. Dies sollte eine Grundlage für weitere Forschungsarbeiten liefern. Außerdem ist die Integration des Datenmodells in die Entwicklung einer Domänenontologie für den Brandschutz (vgl. [PJPK10]) geplant. Für die vorgestellte Arbeit stellen die Implementierung des Datenmodells in ein Informationssystem und seine Evaluierung die folgenden Forschungs- und Entwicklungsaktivitäten dar.

Literatur

[KE09] Alfons Kemper und André Eickler. *Datenbanksysteme - Eine Einführung*. Oldenbourg, 7. Auflage, 2009.

[Kro00] Stavros Kromidas. *Handbuch Validierung in der Analytik*. Wiley-VCH, 2000.

[Pfe01] Tilo Pfeifer. *Qualitätsmanagement - Strategien, Methoden,Techniken*. Hanser Verlag, 2001.

[PJPK10] Jens Pottebaum, Anna Japs, Stephan Prödel und Rainer Koch. Design and modelling of a domain ontology for fire protection. In Simon French, Brian Tomaszewski und C. Zobel, Hrsg., *Proceedings of the 7th International Conference on Information Systems for Crisis Response and Management ISCRAM2010*, 2010.

[Poh08] Klaus Pohl. *Requirements Engineering: Grundlagen, Prinzipien, Techniken*. dpunkt.Verlag GmbH, 2. Auflage, 2008.

[Sch10] Michael Schreckenberg. Grundsätzliche Aspekte zu Evakuierung und Räumung. In *VdS-Fachtagung Evakuierung und Räumung*. VdS-Verlag, 2010.

[SK10] V. Schneider und R. Könnecke. Microscopic modelling of crowd movement at major events. In *Proceedings Interflam2010, Fire Science & Engineering Conference*, Nottingham, UK, July 2010.

[SKK+09] Andreas Schadschneider, Wolfram Klingsch, Hubert Klüpfel, Tobias Kretz, Christian Rogsch und Armin Seyfried. Evacuation Dynamics: Empirical Results, Modeling and Applications. In *Encyclopedia of Complexity and Systems Science*, Seiten 3142–3176. 2009.

[SSL04] Armin Seyfried, Marcus Strupp und Thomas Lippert. Verfeinerte Auswertungsmethoden für Evakuierungssimulationen. Bericht, Nov 2004. vfdb-Zeitschrift für Forschung, Technik und Management im Brandschutz, Seite 214, Heft 4 (2004).

[Tec10] Technisch-Wissenschaftlicher Beirat (TWB) der Vereinigung zur Förderung des Deutschen Brandschutzes e.V. Richtlinie Einsatzplanung Großveranstaltungen. vfdb-Richtlinie 03/03 (Entwurf), 2010.

Risiko und Rettung

- Die Rolle der Informatik in den Phasen des Desaster Managements -

Horst Kremers

CODATA-Germany
PO Box 20 05 48
13515 Berlin
office@horst-kremers.de

Abstract: Der Begriff Risiko ist zentrales Element in allen Phasen des Desaster Managements bei natürlichen und technischen Katastrophen. In diesem Beitrag werden auf der Basis des interdisziplinären Erfahrungsaustausches Methoden der Informatik für Strukturierung, Implementierung und Betrieb entsprechender Informationssysteme aufgezeigt. Die Erarbeitung und Bereitstellung von entscheidungsunterstützenden Informationen ist bei der im Krisenfall typischerweise äußerst hohen Komplexität von Akteuren in den jeweiligen Situationen mit unterschiedlichsten Handlungsanforderungen eine besondere Herausforderung. Die Aspekte der Informationsgesellschaft im Sinne der Bedürfnisse an Information, Kommunikation und Kooperation bedingen nicht nur technische sondern auch innovative rechtliche Entwicklungen auf nationaler und internationaler Ebene.

Decentralized, Resource-Aware Information Management and Delay Tolerant Networks in Command-and-Control

Markus Brückner, Liz Ribe-Baumann
{markus.brueckner, liz.ribe-baumann}@tu-ilmenau.de

Abstract: A robust, decentralized information management system and reliable transport of information between disconnected agents is essential to the success of command and control activities in disaster relief scenarios. We present work-in-progress aimed at developing a distributed hash table (DHT) protocol and a delay-tolerant network that meet the specific requirements of disaster relief scenarios through, among other things, the integration of location and resource awareness.

1 Introduction

During the response and recovery phases after a disaster, large amounts of data are produced and must be stored and transfered to the relief workforces in order to ensure timely, informed, and well coordinated relief efforts. The success of relief efforts depends not only on the right people being sent to the right place, at the right time, with the right equipment, but also their access to up-to-date information pertinent to their tasks and their ability to save and share important information that they are gathering. Such information may include data about environmental conditions, locations of resources, progress of relief efforts, positioning and expertise of relief workforces, or lists of tasks. Unfortunately, the storage and transfer of data in a widespread disaster scenario is complicated by several factors: The amount of information and number of users in the system increases incredibly fast (even disaster in a confined space like the WTC attacks spur a high amount of communication [K$^+$04]); much of the system operates on limited battery power; landlines and base stations may be damaged; and relief workforces must be sent to areas without means for communicating directly with the command center.

The work-in-progress presented in this paper considers two approaches aimed at ensuring reliable storage and transfer of data for the control and command of disaster relief efforts where each node knows its geographic position. As we can not anticipate all possible uses of this work, we focus on one specific type of data generated in a disaster: operation and technical logs which are maintained at every field unit (see [fKuK99] for details). These logs contain information about nearly every event a unit encounters as well as reactions to it. As it is typically used for documentation purposes only, we do not consider it to be time critical. On the other hand, we argue that providing this data to non-data-generating units in a timely manner might prove beneficial both as a backup, improving data availability, and for data mining, improving situational awareness. Transporting the information drawn from this heightened situational awareness back into the field in an automatic fashion would

lessen the burden on both the communication backbone as well as the operators in the control centers.

To facilitate such applications we present partial solution ideas for the development of a highly scalable, location and resource aware distributed data management system (with resources such as power, bandwidth, etc.) as well as a data transport system which enables the forwarding of the data through poorly connected areas. The rest of the paper is organized as follows: We present the data management system in Section 2, which is based on a distributed hash table (DHT) and maps data to network nodes and provides routing algorithms for lookups (see for example [SMK+01]). In Section 3, we present the data transport system which uses a delay tolerant network (DTN) [Fal03] that, in order to minimize delay and maximize delivery probability, exploits prior knowledge about node movements (in our use case, couriers). Section 4 provides a brief conclusion and an outlook to future work.

2 Data Management - Resource Aware Overlay

We consider a DHT to be resource aware when it spreads load among nodes such that node failures or timeouts due to the (over)use of specific resources are minimized, and assume that each query routed over a node uses a portion of that node's available resources. Our overlay construction is motivated by the small-world graphs explored by Watts and Stogatz [WS98], the construction of the virtual network coordinates Vivaldi [DCKM04], and the small-world DHT Symphony [MBR03]. We integrate resource levels into nodes' coordinates and then choose links based on nodes' coordinates' distances in the hopes of reaching a "small-world-like", resource aware DHT similar to Chord [SMK+01].

We assume that each node x possesses two dimensional geographic coordinates $x^g = (x_1, x_2)$ and a resource level $x_r \in \{1, \ldots, r_{max}\}$ for some network-wide maximal resource level r_{max}. Each node also possesses an additional "height" dimension, similar to Vivaldi network coordinates [DCKM04]. Vivaldi coordinates emulate a mass-spring system to estimate nodes' locations in a self-organized manner and use an additional height dimension to distance single nodes from the *entire* network. For the special case of 3-dimensional Vivaldi coordinates, two nodes x and y with coordinates $x^c = (x_1, x_2, x_h)$ and $y^c = (y_1, y_2, y_h)$, have Vivaldi-distance:

$$d_V(x^c, y^c) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2} + x_h + y_h.$$

Ledlie et al. [LGS07] showed that this additional height dimension, which was designed to accommodate the expensive last hop in routing that separates a node from the highly connected internet core, is integral to accurate latency estimates. We use a height dimension to express each node's resource level and distance nodes with low resource availability from the entire network. Thus, a node x with resource level x_r receives resource height $x_h = (r_{max} - x_r)/r_{max}$. Essentially, the shorter the distance between two nodes, the more likely they will be linked, and thus, nodes with lower resource levels have fewer incoming neighbors to incur routing load.

HandleSearchMessage(x^c, d, $NodeList$):
1. if ($|d_V(x^c, thisNode^c) - d| \leq \varepsilon$)
   then sendSearchMessageReply()
// if thisNode at correct dist. to x, reply to x
2. else
   addToNodeList($thisNode$, $NodeList$)
   y = findClosestNeighbor(x^c, d, $NodeList$)
   sendSearchMessage(y, x^c, d, $NodeList$)
// find $y \notin NodeList$ in routing tables with
// smallest $|d_V(x^c, y^c) - d|$, forward to y

Figure 1: Pseudo-code for $thisNode$'s handling of a coordinate link search message.

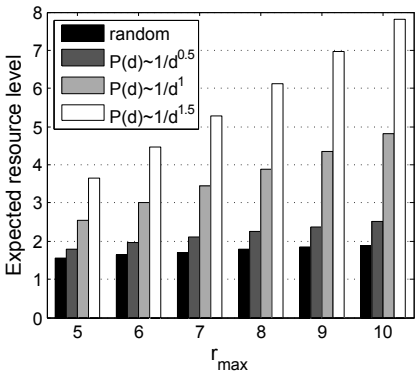


Figure 2: Expected resource level of random coordinate link for various r_{max}'s and probability distributions over distances.

Overlay Initialization. Our DHT functions similar to other ring shaped DHTs (e.g. Chord [SMK+01]), with keys in the interval $[0, 1)$ which wraps around, typical join and failure protocols, and unidirectional routing in the key space. To join the DHT, a node x chooses an identifier $NodeID_x$ uniformly at random from the keyspace and contacts some participating node y with $NodeID_x$. Then y performs a key-lookup to find the node responsible for $NodeID_x$ and x positions itself on the ring of keys, establishes links to its k nearest neighbors on the ring, and assumes responsibility for the range of keys preceding $NodeID_x$ as defined in Chord. Next, x establishes links for two kinds of routing tables which are used jointly for routing: location aware *finger links* based on nodes' NodeIDs as in DHash++ [DCKM04] and *coordinate links* chosen based solely on their distance to x in the network coordinate space.

Coordinate Links. Coordinate links are established based on a discrete probability distribution over the distances between nodes in the network coordinate space: the smaller the distance between two nodes' network coordinates, the higher the probability that they are joined by a coordinate link. To find a new coordinate link, a node x draws a *distance* d from its probability distribution and then initiates a search for a node y at approximately distance d from x in the network coordinate space:

$$d_V(x^c, y^c) \in (d - \varepsilon, d + \varepsilon),$$

for some tolerance level $\varepsilon > 0$. Thus, x sends a search message to a random neighbor with its coordinates x^c and distance d, which is forwarded at most some fixed number of iterations (see Figure 1 for more detail). If a suitable node is found, it is added to x's coordinate-link routing table, otherwise a new distance is drawn and the process restarts. Since nodes have dynamic coordinates, links' coordinates must be monitored and a link must be renewed if its distance has changed significantly.

A node x's probability distribution over the distances depends on x_{rad}, the radius of the coordinate space as measured from x's position (details are omitted here), and the maximal

resource level r_{max}. Distances are chosen as increments of $1/r_{max}$ for factors from 1 to $x_{rad} \cdot r_{max}$. For example, the probability distribution for which the probability for choosing d is inversely proportional to d is:

$$P_x(d = k/r_{max}) = \frac{1/k}{\sum_{i=1}^{x_{rad} r_{max}} 1/i}, \quad k \in \{1, 2, \ldots x_{rad} r_{max}\}. \tag{1}$$

In order to examine the effectiveness of the coordinate-links or the entire system, many assumptions must be made about the distribution of resources and nodes. As an example, Figure 2 shows the expected resource level of a single coordinate link for Zipf-distributed resources - where high resource availability is rare - and uniformly distributed nodes within a cone-shaped area with radius 100. In this case we see that for the distribution in (1) the resource level for a random link is substantially higher than for $P_x(d) \sim d$ (labeled 'random'). However, such observations depend strongly on the assumptions of node and resource distributions.

3 Data Transfer

We propose that the minimization of transport overhead and delay could be reached by optimal adaption of the routing system to nodes' current movement patterns (i.e. passing messages to nodes that will move towards the target as soon as possible). As our solution is intended for relief workers where we expect a certain amount of pre-known missions (e.g. couriers approaching field units at known locations). We intend to use the mission data as a base for routing optimization. This work draws ideas from the work of Lindgren et.al. [LDS03] who proposed using repetitive contact patterns of nodes to optimize message forwarding. We were also inspired by geographic forwarding algorithms as in [Fin87] which rely on the geographic position of nodes to transport data.

System structure. In order to keep routing flexible and extendable, we minimize the interfaces between the routing algorithm and the rest of the system. As the routing optimization is based on the knowledge of node trajectories, we expect a per-node information store to provide access to this information. The necessary data should be exchanged whenever two nodes meet in order to provide system-wide dissemination. Each node provides its own trajectory data based on information from the mission control or – in the case of human-controlled vehicles – route guidance. Our proposal would be to derive the necessary information from mission descriptions given electronically. This is based on the assumption that at least the operation controllers should know where field units are located and are therefore able to provide rather accurate descriptions of the target of a mission. By keeping the information about node movement abstract and separate from the actual source of this data we facilitate later system extensions with other sources of trajectory data like movement prediction algorithms which provide limited data even on systems without pre-known mission goals (e.g. units in search-and-rescue missions).

Movement-based forwarding. Routing – especially geographic routing – is essentially a well-understood problem for which a set of solutions exists. As with the data storage, we use these solutions and adapt them to the problem at hand. Looking at nodes' trajectories in a common space immediately displays a similarity to a two-dimensional map. While this similarity is clear, it can be misleading: trajectories may intersect but routing of messages along them is only possible if the corresponding nodes meet. In order to capture this we have to extend the model by one dimension: the time.

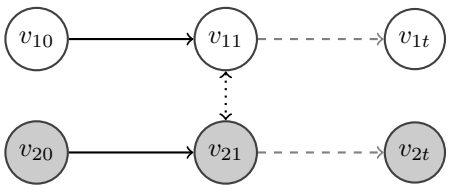


Figure 3: A connectivity graph of two nodes meeting once

By introducing the time into the picture we define the concept of a *temporal map*. The movement of the node is now a function of time: $T : \mathbb{R} \rightarrow \mathbb{R}^2$ where T is the trajectory of a node. By calculating the distances of all trajectories within the prediction window, we are able to derive a connectivity graph that models the possible paths of messages through the system by adding virtual intermediate nodes to the graph for every communication opportunity. An example for such a graph is given in Figure 3. The white and gray sets of vertices each represent one node at different points in time. The black solid edges represent messages traveling inside a node's buffer for some time, while the dotted edge represents a communication opportunity. The gray, dashed edges are mere helpers to simplify future route calculation. Their edge weights are set to a neutral element with regard to the cost function for the path calculation.

With the right weights assigned to the edges, optimizing the forwarding of a packet from node n_1 to n_2 is modeled by finding the minimum-cost path through the connectivity graph from v_{10} to v_{2t}, a problem that is well understood and solved e.g. by Dijkstra's algorithm. Whenever a node has to take a forwarding decision (i.e. whenever the opportunity to communicate arises) it can calculate the shortest path to the target and check whether the crossover edge representing the current communication attempt is contained in that path. If it is, the message is forwarded to the communication partner, otherwise it is kept in the internal buffer. Depending on the selection of the cost function, the path could be optimized for minimum delay or maximum delivery probability.

4 Conclusion - Future Work

We presented solution ideas to the problem of storing and transporting non-time-critical data in a disaster. Our solution is divided into a DHT-based storage subsystem providing reliability despite widespread failures and limited resource availability and a DTN-based

data transport system adapting to existing movement patterns. While our use cases are currently aimed at an electronic version of operation and technical logs, we can not fully anticipate future uses of the system. We expect that the widespread availability of the data will increase situational awareness and prevent mistakes due to incomplete or outdated information, but this must still be investigated. Future work should focus on improving our understanding of the scalability and robustness of the system. On the one hand, replication schemes for the data management system need to be developed in order to maintain a desired level of availability given nodes' locations and resources. On the other hand, data prioritization and reliability levels should be integrated into the transport system to provide further optimizations of message transport. Furthermore, the potential for the optimization of message transport using the remaining infrastructure after widespread failures should be assessed. However, since this system is dissimilar to existing communication infrastructures, we are currently not pursuing the direct integration with existing networks.

References

[DCKM04] Frank Dabek, Russ Cox, Frans Kaashoek, and Robert Morris. Vivaldi: a decentralized network coordinate system. *SIGCOMM '04*, pages 15–26, 2004.

[Fal03] Kevin Fall. A delay-tolerant network architecture for challenged internets. In *SIGCOMM '03*, pages 27–34. ACM Press, 2003.

[Fin87] G.G. Finn. Routing and Addressing Problems in Large Metropolitan-Scale Internetworks. Technical report, University of Southern California, Marina del Rey, Information Sciences Institute, 1987.

[fKuK99] Ständige Konferenz für Katastrophenfürsorge und Katastrophenschutz. Führung und Leitung im Einsatz. Download: http://www.katastrophenvorsorge.de/pub/publications/DV100-SKK.pdf, 12 1999.

[K+04] Thomas H. Kean et al. The 9/11 Commision Report. Technical report, National Commission on Terrorist Attacks upon the United States, 6 2004.

[LDS03] Anders Lindgren, Avri Doria, and Olov Schelén. Probabilistic Routing in Intermittently Connected Networks. In *SIGMOBILE Mobile Computing and Communication Review*, volume 7, 7 2003.

[LGS07] Jonathan Ledlie, Paul Gardner, and Margo Seltzer. Network coordinates in the wild. In *In Proceedings of USENIX NSDI'07*, 2007.

[MBR03] Gurmeet S. Manku, Mayank Bawa, and Prabhakar Raghavan. Symphony: Distributed Hashing in a Small World. In *In Proceedings of the 4th USENIX Symposium on Internet Technologies and Systems*, pages 127–140, 2003.

[SMK+01] I. Stoica, R. Morris, D. Karger, M.F. Kaashoek, and H. Balakrishnan. Chord: A scalable peer-to-peer lookup service for internet applications. In *SIGCOMM'01*, pages 149–160, 2001.

[WS98] D. J. Watts and S. H. Strogatz. Collective dynamics of'small-world'networks. *Nature*, 393(6684):409–10, 1998.

Is it possible to interact with a handheld device while holding it in both hands?

Tayfur Coskun, Simon Nestler, Eva Artinger, Amal Benzina, Gudrun Klinker

Fachgebiet Augmented Reality
Technische Universität München
Fakultät für Informatik
Boltzmannstraße 3, 85748 Garching bei München
coskun|klinker@in.tum.de

Abstract: Intuitiveness, attractiveness and efficiency are in general important characteristics of high quality **U**ser **I**nterfaces (UI). In the case of a **M**ass **C**asualty **I**ncident (MCI) and other life threatening situations good user interfaces are essential as every second counts in these time-critical situations. Finding the best way to interact with the system is a challenge with many open issues. Therefore, it is essential to test different UI alternatives to determine important usability issues as those mentioned before. This paper focuses on the research question how patients can be selected on a map application which runs on a rugged tablet PC. A rugged tablet PC is in general heavy because of its robustness, which introduces a new special requirement in addition to the existing UI requirements. The implementation and the evaluation of the UI alternatives presented in this paper were performed within the scope of the SpeedUp project[1].

1 Motivation - Usability brings order in the chaos?

A **M**ass **C**asualty **I**ncident (MCI) is an incident which generates more patients than the locally available rescue workers can manage simultaneously. Since there are too many injured people to handle at once, they need to be categorized. There is already an ongoing research which digitizes this triage approach using PDAs [NK07]. Using PDAs equipped with GPS sensors has the advantage to track the position of the patients. The PDAs are supposed to be used by the relief units in the field. The tracking positions are transferred from the distributed PDAs in the field to the tablet PC used for this work. The positions of the patients are visualized in an overview map on the tablet PC in the scope of this work. Currently, the **E**mergency **M**edical **C**hief (EMC) is getting the information about the situation from the rescue units verbally. The number of injured people and their classification are written down by the relief units on paper. So this information can be incomplete, redundant and/or wrong. Obviously, using an IT solution can improve and introduce order

[1]The project SpeedUp is funded by the German Federal Ministry of Education and Research (BMBF) within the program "Research for Civil Security" (May 1st, 2009 - April 30th, 2012, FKZ: 13N10175). Website: http://www.speedup-projekt.de

in the chaotic situation in such incidents. This is why we propose that the EMC should use the tablet PC to acquire the required information about the patients. However, because of the users' tendency to resist or even reject new technologies with inadequate user interfaces or bad usability, all conceivable improvements can be useless. It is therefore essential to develop the user interface close with the user to find out which UI alternative is the most intuitive, efficient and attractive. Special requirements on the user interface suitable for an MCI situation have to be considered and explored. This paper focuses on designing a usable UI to select items (patients or relief units for instance) fulfilling these special UI requirements mentioned in 2.

2 Special Requirements

The requirements in general can be split into hardware and software requirements. The hardware requirements and common software requirements are described in the long version of this paper. Therefore, only the special requirements are described here. To date, the information in an MCI situation is mainly captured on paper. Since this method provides an easy, familiar, intuitive and flexible way to realize the information capture task, the first requirement is to achieve a comparable performance with the tablet PC UI as compared to paper. Another special requirement from the fire department of the Technische Universitaet Muenchen (Feuerwehr TUM) is to use just the thumbs for all interaction tasks while holding the tablet PC with both hands. The fire department staff is using a heavy ruggedized tablet PC during their daily work (xplore iX104). The staff complained about the fact that holding it with one arm during interacting with the free hand causes their arm fatigue. Therefore, to solve this problem, the users should be able to interact with the application while holding the tablet PC with two hands. This leads to the fact that only the user's thumbs are able to reach the screen. As a consequence the application must provide all interaction elements on the left and/or the right edge of the tablet PC screen as shown in figure 1.



Figure 1: Tablet PC xplore iX104 holding in both hands

3 Concepts to select patients

The UI of the map application will be developed gradually to increase the chance of its acceptance by the target group (EMC). The goal of this paper is to find an answer to the

question: **How can patients be selected?** There will be different items like vehicles, special places, relief units or patients in the final application. Since the UI has to be completely controllable from the edge of the screen, items cannot be selected by just touching them. Three different UI concepts which fulfill this main requirement are introduced in this section. All of them will create special UI elements (buttons) at the edge of the tablet PC screen which can be reached with the thumb while holding the tablet PC in both hands. And since the space on the edge is limited, only a subset of all visible patients can be selectable at once. Therefore, how to determine this subset is the main topic of this section. To be able to compare the three selections methods which each other, all three should use the same buttons to finally select one patient among the determined subset. The three investigated selection methods are: Horizontal Line (SE1), Selection Quad (SE2) and Automatic Mapping (SE3). **Horizontal Line:** One approach to select patients which are not reachable with the thumb is to draw a horizontal line from the user's right thumb position to the left hand-side of the screen (see figure 2 (a)). This line disappears when the display is not touched anymore. If the user slides his right thumb up or down the horizontal line will follow it. The speed of the line's movement is one-to-one mapped to the thumb's speed in order to make the movement more intuitive. Each patient intersected by the line will be a member of the subset described before, and is therefore selectable through pressing the corresponding button on the left-hand side edge of the screen. This way every patient which is visible on the current segment of the map can become selectable by moving the horizontal line. **Selection Quad:** The approach here is to render a square on the tablet PC screen as shown in figure 2 (b). Each patient who is inside this square is mapped to a member on the left side of the tablet PC and thus selectable by the left thumb. The user is able to move this square by a graphical joystick. In the contrary to the selection with the horizontal line there is a risk that there are too many patients inside the square, so there has to be a possibility to resize it. For this issue the joystick is extended with a plus and a minus button to control the size of the square. Even though this will increase the complexity of this selection method compared to the horizontal line, it also provides more flexibility to the user. **Automatic Mapping:** Our last approach is to map all of the visible items to both edges of the screen. On the one hand an obvious advantage is that the user can select an item without the step of determining a subset (see figure 2 (c)). This alternative provides a button to toggle the selection on and off which makes it easier and faster compared to the horizontal line (SE1) and the selection quad (SE2). On the other hand this method introduces some undesirable effects. Since each patient can be selectable, every patient needs a correspondent button on the edge which requires more space from the limited interaction space.



(a) Horizontal Line (SE1)



(b) Selection Quad (SE2)



(c) Automatic Mapping (SE3)

Figure 2: Concepts

4 Evaluation

The first part of this section explains the participants used for this evaluation, followed by a subsection describing the procedure to be performed by the users. After that the results are presented and discussed.

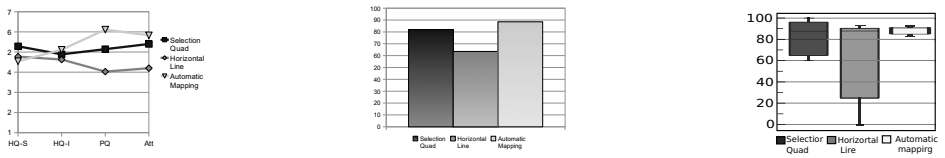
Participants According to literature we will learn the most from our first few test users. It is recommended to have between three to five test users, since more users will not necessarily improve the results of the evaluation [Nie94]. Therefore, the system was evaluated by five test users. In this first approach those test users were not from the rescue service. This decision was made because the **initial** usability issues are the same for most of the people during the **first** evaluation of an UI, whether they are from the rescue service or not. Actually, it is more important to choose people with low experience in the field of human computer interaction in this first step. People from different fields were randomly chosen for that reason: Two biologists, a political scientist, a sociologist and a jurist. The results of this evaluation will highly improve the presented UI elements and may lead to filtering out of some elements. The remaining enhanced and optimized UI elements will **then** be evaluated with the rescue service team (EMC) in future work. The remaining enhanced UI elements will then be evaluated with the rescue service team (EMC) in future work. All of the five test users (two females and three males) have low to average experience with computer interaction in general and with touchscreens in particular according to their own estimation. All of them were between 25 and 28 years old. To find out usability issues and to be able to rate the different UI alternatives, the test users have to use and experience the UI elements. Therefore, it makes sense to define the same task in all three alternatives to be performed by the users. The time duration it takes to finish a task, is measured for all users for each alternative. For each of the interface elements the users fill out two questionnaires: SUS (System Usability Scale) [Bro96] and AttrakDiff [HBK03]. After that a short interview is held with each of the test users. In this interview the test users are asked how they feel and think about the different UI techniques. This is important to find out the reasons for the scores of SUS and AttrakDiff and to be able to determine specific difficulties using these UI techniques.

Procedure During the execution of each task the test user is observed to detect some usability issues and clarify potential ambiguities faced by the test users. To avoid the learning effects, the sequence of the different alternatives are randomized. During the interview phase, the test users give their feedback and experience multiple UI alternatives. Finally, the test users are asked to choose and argue about the preferred UI alternative for selecting patients. Concerning the task description, there are 26 patients at the beginning displayed at the center of the map, and the ability to move or zoom the map is disabled. The patients‘ locations and triage states will be the same for each alternative. The test users are first asked to select every red categorized patient, then every yellow one with each alternative. Every selected patient will disappear immediately in this test scenario from the map. The task is completed when no patient to select is left on the map.

5 Results

In this section the results are presented for each UI selection alternative and discussed. In the first paragraph the results of AttrakDiff are recapitulated, while the next paragraph summarizes the results from SUS. Finally the outcome of the test users' feedback during the interview will be summed up and discussed.

AttrakDiff The results of AttrakDiff are divided into four dimensions (see figure 3 (a)). The first two dimensions describe the hedonic quality of the UI and enclose stimulation and identity. For the stimulation part the selection quad (SE2) shows the best results while the automatic mapping has the highest score in the remaining three dimensions. The differences between UI alternatives are higher for the pragmatic quality (PQ) and the attractiveness (Att). Moreover, the order of the scores for the identity quality, pragmatic quality and the attractiveness respectively are the same. That means, the automatic mapping (SE3) has scored the best results in all three qualities followed by the selection quad and finally the horizontal line. Nevertheless, the results are good for all alternatives since the score is never below four, but there are evident tendencies for the automatic mapping alternative.



(a) AttrakDiff (b) SUS - Means (c) SUS - Boxplot

Figure 3: Results for AttrakDiff and SUS

SUS The results for the system usability scale support the results from AttrakDiff (see figure 3 (b)).Actually, the usability of the automatic mapping alternative obtained as well the highest score of 88.5 among the three alternatives. Figure 5 (c) shows clearly that the results of the automatic mapping alternative are very decisive since all test users rated this alternative in an interval between [82.5 ; 92.5]. Even though, the selection quad alternative obtained the second best score with a mean of 82.0, the test users rating was spread over a bit larger interval of [60.0 ; 100.0] compared to the automatic mapping alternative. The reasons for those scores become clearer after the interview and the feedback of the test users.

Interview The interview shows that the reason for the low score of the horizontal Line (SE1) is that the line was too thin to intersect the patients easily. That is even more disadvantageous for the specific task performed by the test users during the evaluation where they had to select all patients. However, if the specific task is to select one patient among a huge number of patients geographically, the horizontal Line (SE1) might score the best among the three UI alternatives. This has been also confirmed by one of the test users. The selection quad is not immediately understood from all the test users which make this alternative not as intuitive and practical as the automatic mapping. A better visualization of the functionality and the usage of the selection quad could improve its usability and

practicality. Even though the complexity of the selection quad is higher than the automatic mapping, it provides also more flexibility to the user since scrolling and zooming can be integrated in a very compact way with no extra UI element. This benefit could not be seen by the test users during this evaluation because the scrolling and zooming feature of the selection quad was disabled for this evaluation.

6 Conclusion and Future Work

During this evaluation, the automatic mapping was clearly rated as the best selection technique. But this result has to be regarded carefully, since this was an initial evaluation with a predefined task. This task is not representative for all possible real tasks in a real MCI, in our case the result of the evaluation is strongly bounded to this task. Nevertheless the evaluation points out important issues for all three techniques. Investigating those issues and learning from the test users' feedback, improve greatly the proposed UI selection alternatives. Within the SpeedUp project there is ongoing work researching the best UI scrolling and zooming alternatives based on the same requirements presented in section 2. So the next step is to combine the alternatives for these three basic map features (selecting, scrolling and zooming) which was rated with the highest scores and integrate them to a compact system. Additionally, a new UI alternative with all basic map features has to be introduced for the purpose of evaluation. Finally, defining new tasks is also needed to heavily evaluate the UI alternatives; those tasks should make the user employ all provided UI alternatives features.

References

[Bro96] John Brooke. SUS: a 'quick and dirty' usability scale. *Usability evaluation in industry*, 1996.

[HBK03] Marc Hassenzahl, Michael Burmester, and Franz Koller. AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität. In G. Szwillus and J. Ziegler, editors, *Mensch & Computer 2003: Interaktion in Bewegung*, pages 187–196, Stuttgart, 2003. B. G. Teubner.

[Nie94] Jakob Nielsen. Using Discount USability Engineering to Penetrate the Intimidation Barrier. *Guerrilla HCI*, 1994.

[NK07] S. Nestler, , and G. Klinker. Using Mobile Hand-Held Computers in Disasters. In *UbiComp Workshop on Interaction with Ubiquitous Wellness and Healthcare Applications (UbiWell)*, Innsbruck, Austria, September 2007.

Kultur und Technik - Ein kulturanalytischer Ansatz als Grundlage der Entwicklung eines mobilen Endgerätes für Einsatzkräfte in Großschadenslagen

Schumann, S., Yildirim-Krannig, Y., Mähler, M., Wucholt, F., Strohschneider, S.

Fachgebiet für Interkulturelle Wirtschaftskommunikation
Friedrich-Schiller-Universität Jena
Ernst-Abbe Platz 8
07743 Jena
sandy.schumann@uni-jena.de
yeliz.yildirim@uni-jena.de

Abstract: Es wird ein kulturanalytischer Ansatz vorgestellt, der die Einbindung zukünftiger Nutzer in den Entwicklungsprozess eines mobilen Endgerätes zur Kommunikation in Großschadenslagen garantiert. Am Beispiel der Polizei in Deutschland werden Ergebnisse der Kulturanalyse abgeleitet, die in direkten Implikationen für die Gestaltung einer technischen Lösung resultieren. Vorteile der Methode, wie eine stärkere Akzeptanz durch die Endanwender, werden dokumentiert.

1 Zielstellung

Bei der Entwicklung technologischer Lösungen ist die Schnittstelle zu zukünftigen Anwendern bedeutsam, um die Akzeptanz und tatsächliche Nutzung der Produkte sicher zu stellen. Ansätze, die *den Faktor Mensch* in die Technologieforschung einbinden, reichen von klassischen Usability Studien mit quantitativen Methoden (Vergleich beispielsweise [BKM08]) bis zu Kulturstudien auf Basis qualitativer Daten. Während Usability Studien bereits weithin Anwendung im Entwicklungsprozess von technologischen Produkten finden, sind Kulturstudien als Grundlage noch Einzelfälle. Ziel dieser Ausführungen ist es zu verdeutlichen, warum qualitative Kulturanalysen bei der Gestaltung technischer Lösungen eine wichtige Rolle spielen und Beachtung finden sollten. Der Anwendungsfall, auf den sich die Darstellungen beziehen, umfasst die Entwicklung eines mobilen Endgerätes (eines PDA, BCMU oder Tablet PC), mit dem die Kommunikation und Informationssammlung sowie -weitergabe innerhalb und insbesondere zwischen verschiedenen Behörden und Organisationen mit Sicherheitsaufgaben (BOS) in Großschadenslagen unterstützt werden soll. Denn aktuelle Systeme betrachten primär die intraorganisationale Kommunikation und bieten kaum Schnittstellen zu anderen BOS für einen, deren Kommunikations- und Entscheidungsstrukturen angemessenen Informationsaustausch. Als Fallstudie werden Daten, die aus Interviews mit der Polizei gewonnen wurden, vorgestellt. Der Einzelfall gilt als Allgemeines insofern, als er sich im Kontext allgemeiner Regelhaftigkeit gebildet hat. Der Fall gilt als Einzelnes, insofern er sich im Kontext besonderer Bedingungen

entwickelt hat. Hier ist entscheidend, die Dialektik von Allgemeinem und Besonderem herauszuarbeiten. Das grundlegende methodische Vorgehen orientierte sich an den typischen Ansätzen der freien Beobachtung, dem Führen von Leitfaden- oder narrativen Interviews und der Dokumentenanalyse.

1 Polizei als Kultur

Polizeikultur, verstanden als Organisationskultur, wird in der neueren Polizeiwissenschaft überwiegend hinsichtlich ihrer identitätsstiftenden Merkmale analysiert[1]. Dazu zählen u.a. „(…) typische polizeiliche Handlungen, Symbole, Zeremonien, Rituale, Stile usw., die natürlich von der jeweiligen Organisationsstruktur (den harten Fakten), aber auch von den praktizierten Werthaltungen, Normen, Orientierungsmustern, Leitbildern usw. (den sogenannten weichen Faktoren) abhängig sind“ [Ah00]. Grundlage des Verständnisse von Organisationskultur hier ist Hansens Kulturmodell[2], welches als solche nicht Nationalstaaten (Homogenisierung von Kultur), sondern vielmehr Kollektive im lebensweltlichen Sinne versteht, in denen sich Mitglieder einer Gesellschaft mehrfach verorten lassen [Ha03]. Diese Kollektive[3] unterscheiden sich durch die von ihren Mitgliedern entwickelten Gewohnheiten (sogenannte Standardisierungen) in folgenden Kategorien: „der Kommunikation, de(m) Denken( ), de(m) Empfinden( ), de(m) Verhalten( ) und Handeln( )“ [Ha03]. Durch diese Standardisierungen wird eine Basis dafür geschaffen, dass Mitglieder der Kollektive etwas oder jemanden als „normal“ empfinden. Kultur ist somit sowohl Resultat als auch Mittel sozialer Interaktion, die sich in Organisationsformen (Organisationen, Unternehmen, Institutionen) und Strategien (Entscheidungsprozesse und Verhaltensformen) niederschlägt [St10]. Eine differenzierte Betrachtung der Organisationskultur erlaubt das Schichtenmodell von Dülfer[4] [Dü01]. Die Organisation selbst wird dabei als ein eigenständiges Subjekt (Kollektiv zweiten Grades bei Hansen) aufgefasst, innerhalb dessen unterschiedliche Interessengruppen (Kollektive ersten Grades), wie z. B. Verkehrspolizei, Landeskriminalamt oder Wasserschutzpolizei agieren. Nach Dülfer wird die Kultur einer Organisation durch Interaktionen und Einflüsse geprägt, die sich sowohl in horizontaler als auch in vertikaler Ebene vollziehen. Auf horizontaler Ebene wird die Kultur durch die Interaktionen zwischen der Organisation und seiner Aufgaben-Umwelt geprägt, z. B. Staatsanwaltschaft, Ordnungsamt bzw. Bürger, mit denen es zusammenarbeitet. Auf vertikaler Ebene wird die Kultur durch die globale Umwelt beeinflusst, z. B. durch rechtlich-politische Normen. Die Vorschriften und Gesetze (StGB, StPO etc.) schaffen eine gewisse Homogenität der Kultur. Die Multikausalität von Organisationskulturen zeigt, dass die

[1] Organisationskulturanalysen beziehen sich traditionell auf Fragen der Humanisierung der Arbeitswelt (Themen wie Persönlichkeitsentfaltung, Selbstverwirklichung, Zufriedenheit) und auf die Erhöhung der Leistungsbereitschaft und Innovationsfähigkeit [Be06]

[2] Kultur wird bei Hansen auf allen Ebenen menschlicher Gruppierungen angesiedelt und kultureller Zusammenhalt nicht auf der Voraussetzung von Homogenität aufgebaut. Kultur ist demnach Produkt kommunikativer Prozesse [Ha00].

[3] Hansen postuliert zwei Typen von Kollektiven: Kollektiv ersten Grades (besteht aus Individuen) und Kollektiv zweiten Grades (besteht aus ähnlichen Kollektiven, sogenannten Zwillingskollektiven) [Ha09].

[4] Als wesentliches Element des Schichtenmodels gilt, dass Kultur ein multikausales Konstrukt ist, welches auch für einzelne Subkulturen innerhalb einer Organisation gelten muss [Dü01].

Entstehung und Entwicklung von Kulturen dynamischen Prozessen unterliegen und Veränderungen fortwährend stattfinden. Technologische Lösungen können dort ansetzen, aber nur dann Akzeptanz finden, wenn sie ein Teil der Kultur werden und somit normales, routinehaftes und plausibles Handeln (i.S.v. Kommunikation) ermöglichen. Damit wird auch deutlich, dass für die Entwicklung von Technik *für* Kulturen eine tiefgründige Analyse dieser unabdingbar ist. Genauer soll dies an dem bereits angekündigten Beispiel gezeigt werden. Im Folgenden werden Analysen der Organisationskultur der Polizei vorgestellt und Hinweise für die Entwicklung technischer Lösungen abgeleitet.

2 Polizeikultur und Führungsstil

Ein Ergebnis der Kulturanalyse der Polizei in Deutschland ist ihr praktizierter Führungsstil. Neben den Ausführungen der Polizeidienstvorschrift 100 (PDV 100) [nn99], die eine Art Soll-Zustand darstellen, weisen auch inhaltsanalytische Untersuchungen [SMYWS10] zum tatsächlichen Ist-Zustand darauf hin, dass die Polizei den Ansatz des partizipativen Führungsstils verfolgt, um ihre Aufgaben der Gefahrenabwehr und Strafverfolgung zu gewährleisten (PDV 100). Dieser Führungsstil postuliert, dass Menschen grundsätzlich motiviert sind, verantwortungsvoll und kooperativ zu arbeiten. Es wird auch als Auftragstaktik in Abgrenzung zu einer Befehlstaktik beschrieben (PDV 100). Entscheidungen in Einsätzen werden in selbstständigen Einheiten getroffen. In Lagen, in denen Organisationen an ihre Grenzen der Leistungsfähigkeit gebracht werden und Informations- und Kommunikationsmanagement schwierig ist, können so Führungsmodelle flexibel gestaltet und Mitarbeitern Autonomie und Vertrauen zugesprochen werden. Wie können diese Erkenntnisse der Kulturanalyse nun Einfluss auf die Entwicklung und Gestaltung technologischer Lösungen nehmen? Die Auftragstaktik, die einen kreativen Spielraum für die Ausführung der Aufträge offen lässt und der partizipative Führungsstil sind nur im Zusammenspiel mit bestimmten Kommunikationswegen innerhalb der Polizei funktional und plausibel. Dabei sind die Kommunikationswege und Umfang der Kommunikation aufgabenabhängig und lageangepasst auszurichten. Informationswege müssen hinsichtlich ihrer Belastbarkeit und der Einrichtung zusätzlicher Sender oder Empfänger erweiterbar sein. Damit wird für die Entwicklung mobiler Endgeräte ein Fokus auf dynamische Systeme deutlich, die es auch erlauben, ad hoc gewisse Filter für den Informationsfluss einzurichten oder aufzuheben. Da existierende Organisationsstrukturen lagebedingt erweitert oder auch alternativ gestaltet werden können - Kommunikation aller Einsatzabschnittsleiter beispielsweise mit einem bestimmten Unterabschnitt prinzipiell nötig oder eben nicht gewünscht sein kann - können technische Lösungen, die die Arbeit der Polizei unterstützen sollen, nur dann Akzeptanz finden, wenn sie dieses Spezifikum der Organisationskultur beachten.

3 Sprache als Kulturträger

Einen Zugang zur Polizeikultur als Organisationskultur kann außerdem durch die Betrachtung dessen, was in der Wahrnehmung eines lebensweltlichen zusammenhangsprägend ist, erlangt werden. Sogenannte „alltagskulturelle“ Artefakte (verstanden als Indikatoren für ein zugrundeliegendes Konzept) spielen für das

spezifische Funktionieren einer Organisationskultur eine wesentliche Rolle. Eine bestimmte Terminologie und polizeiliche Fachbegriffe sind Ausdruck dieser Artefakte. Sie werden im Laufe der Ausbildung und der täglichen Arbeit erlernt und verfestigt. Durch die gemeinsame Sozialisation von Mitgliedern einer Organisationskultur kann sichergestellt werden, dass die Konnotation der Begriffe sowie der Zeichenvorrat gleich sind und somit Missverständnisse vorgebeugt werden. Zeichen haben inhaltsvermittelnde und symbolische Funktion und können auch Gesten, grafische Abbildungen und Laute umfassen [Bo07]. Dabei muss allerdings beachtet werden, dass Zeichen von dem Kontext des Beschreibenden abhängig sind und sich hinter identischen Begriffen durchaus sehr unterschiedliche Konzepte verbergen. Erst wenn Zeichen, konventionalisiert als Code beschrieben, zumindest teilweise zwischen Kommunikationspartnern geteilt werden, erlauben sie eine Verständigung [Bo07]. Sie sind die Grundlage für gemeinsame mentale Modelle einer Situation, Zielvorstellung und einzuleitende Maßnahmen. Für die Polizei ließen sich hier als Beispiel für Codes eine Reihe von Begriffen und Abkürzungen aufführen, die von Vertretern anderer Behörden und Organisationen mit Sicherheitsaufgaben, aber auch Bürgern und Opfern nicht geteilt werden. Gleich erscheinende Begriffe sind mit unterschiedlichen Bedeutungen verbunden, da sie auf unterschiedlichen Wissensvorräten basieren und das vermeintlich Gleiche wird mit verschiedenen Konzepten verbunden. Die Allgemeine Aufbauorganisation der Polizei beispielsweise, die als AAO bezeichnet wird, ist innerhalb der Feuerwehr mit der Alarm und Ausrückordnung (ebenfalls als AAO bezeichnet) verbunden. Zusätzlich zu Begriffen werden innerhalb der Organisationen für gleiche Konzepte unterschiedliche taktische Zeichen zur Dokumentation und Kommunikation im Einsatz genutzt. Für die Entwicklung eines mobilen Endgerätes, das von verschiedenen BOS in einer Großschadenslage zum Informationsaustausch genutzt werden soll und Schnittstellen zwischen den BOS bieten soll, ist deshalb die Beschreibung der Codes einer Organisationskultur unabdingbar. Hierbei kann man die sich gegenseitig bedingenden verbale, nonverbale, paraverbale und extraverbale Ebenen unterscheiden [Bo07]. Besonders bei der Gestaltung von Benutzeroberflächen müssen diese Ebenen und ihre jeweiligen Spezifika mit einfließen, da sie den kommunikativen Stil der Kultur charakterisieren und Grundlage der Normalitätserwartung der zukünftigen Anwender sind. Die standardisierten Begriffe der Polizei können durch eine Kulturanalyse herausgearbeitet werden, ebenso wie taktische Zeichen und paraverbale Konventionen. Mobile Endgeräte können außerdem die Kommunikation zwischen verschiedenen BOS unterstützen, in dem sie die Kommunikationsspezifika aller Organisationskulturen aufgreifen und Begriffe entsprechend ihrer Konnotation übersetzt werden.

4 Zusammenfassung und Ausblick

Zusammenfassend geben die vorgestellten Beispiele Einblicke, wie Kulturanalysen als Grundlage für die Entwicklung von technologischen Lösungen genutzt werden können. Der Fokus lag dabei auf der Polizei als einer Organisationskultur, die bei der Bewältigung verschiedener Einsatzlagen beteiligt ist. Dazu wurden zwei Ergebnisse der Kulturanalyse beschrieben und deren Relevanz für die Gestaltung eines mobilen Endgerätes, das in Großschadenslagen die Kommunikation von BOS optimieren soll, betont. Zusammenfassend wird deutlich, dass der methodische Ansatz für

technologische Entwicklungen eine interessante Option darstellt, da sie im Gegensatz zu Usability-Studien den Nutzer bereits vor der eigentlichen Konzeption eines Gerätes einbindet und somit ein tiefgründiges Einbeziehen erlaubt. Damit sollte die Wahrscheinlichkeit der Akzeptanz der Technologie und tatsächliche Nutzung in dem angestrebten Kontext steigen sowie Iterationszyklen zur Anpassung der Produkte nach der ersten Entwicklung reduziert werden. Neben der Akzeptanz einer spezifischen Lösung können durch Kulturanalysen auch Aussagen zur allgemeinen Akzeptanz technischer Innovationen in Organisationskulturen gemacht werden. Denn auch wenn bei der Bewältigung von Einsätzen die Polizei mit Funkgeräten, Mobiltelefonen und Laptops arbeitet, werden auch immer noch 4-fach Vordrucke genutzt. Letztere scheinen mit technisierten Lösungen teils identisch zu sein. Ein Blick auf Strukturen, die diesen Artefakten zu Grunde liegen, lässt erkennen, dass diese Verwendung von Papier über die Jahre zur Normalität geworden ist. Polizisten fühlen sich nicht in ihrem Routinehandeln der Dokumentation gestört.

Literaturverzeichnis

[Ah00] Ahlf, E. H.: Ethik im Polizeimanagement, BKA-Forschungsreihe, Wiesbaden, 2000.

[Be06] Behr, R.: Polizeikultur. Routinen – Rituale – Reflexionen Bausteine zu einer Theorie der Praxis der Polizei, VS Verlag für Sozialwissenschaften, Wiesbaden, 2006.

[BKM08] Bangor, A., Kortum, P.T., & Miller, J.A.: An empirical evaluation of the System Usability Scale (SUS). International Journal of Human-Computer Interaction 24(6), 2008.

[Bo07] Bolten, J.: Einführung in die Interkulturelle Wirtschaftskommunikation, UTB Verlag, Berlin, 2007.

[Dü01] Dülfer, E.: Internationales Management in unterschiedlichen Kulturbereichen, 6. Aufl., Oldenbourg Wissenschafts Verlag, München, 2001.

[Ha02] Hansen, K. P.: Kultur und Kulturwissenschaft. Eine Einführung. 2. Auflage. UTB Verlag, Tübingen, Basel, 2002.

[Ha03] Hansen, K. P.: Kultur und Kulturwissenschaft. 3. Auflage. UTB Verlag, Berlin, 2003.

[Ha09] Hansen, K.P.: Zulässige und unzulässige Komplexitätsreduktion beim Kulturträger Nation. Interculture Journal 8, 2010.

[nn99] n.n.: PDV 100 - Führung und Einsatz der Polizei (Ausgabe 1999), n.n., 1999.

[SMYWS10] Schumann, S., Mähler, M., Yildirim-Krannig, Y., Wucholt, F., & Strohschneider, S.: Eine Momentaufnahme der Polizei im Wandel, Unveröffentlichtes Manuskript, 2010.

[St20] Strähle, J.: Was ist Unternehmenskultur. In (Barmeyer, C.; Bolten, J., Hrsg.): Interkulturelle Personal- und Organisationsentwicklung. Verlag Wissenschaft und Praxis, Sternenfels, 2010; S. 57-77

Data Management in Mass Casualty Incidents: The e-Triage Project

Thomas Greiner-Mai

Euro-DMS Ltd.
Anzengruberstraße 10 A
82140 Olching
greiner-mai@euro-dms.de

Anton Donner

DLR
Institute of Communications and Navigation
PO Box 11 16
82230 Weßling-Oberpfaffenhofen
anton.donner@dlr.de

Abstract: The main drawback of paper-based registration systems for organizing mass casualty incidents (MCIs) is that information about affected persons remains among the persons themselves. Data can be duplicated/aggregated by laborious manual copying triage tags only, and the normal medium for exchanging information are voice-based radio systems. Since MCIs normally overwhelm the regularly available rescue resources a particularly effective crisis management has to be applied. Operational command centers and rescue forces need information as fast as possible about type and number of injuries, so that each affected or injured person gets optimal care. This paper gives a general survey of the e-Triage project, in which a coherent overall concept for a MCI data management system is studied.

1 Introduction

Within the e-Triage project [etr10], which is sponsored by the German Federal Ministry of Education and Research, an integrated concept for electronic registration of affected persons is under development. Apart from the technical challenges the degree to which emergency forces accept the e-Triage system will depend primarily on psychological factors. A pre-emptive design of the technology, which accommodates the reduced cognitive abilities of emergency personnel operating under extreme stress, is crucial for a successful deployment. At time of writing, the e-Triage demonstrator system is under implementation; a first trial is planned for January 2011.

The design of the technical concept for the system described within this paper is based on use cases developed from rescue professionals and the derived system requirements,

regarding the user interface, the communication infrastructure, and the database system. Among these requirements it must be assured that the system accomplishes the following general necessities:

(1) Scalability: The system must be able to assist rescue forces in a wide range of mass casualty incidents (MCIs), considering on the one hand their size, and on the other hand, the geographical impact. Moreover, different types of organizations may be active on the disaster area. The system must assure inter-operability among the different forces.
(2) Dynamicity: Rescue teams may join/leave the system at any point in time and space during the rescue operation. The system must support dynamics in order to manage late joins and users leaving the system.
(3) Ease of use: In a general case, the users of the system will be members of rescue forces with (normally) limited technical background. The system must be easy to deploy in the field and must provide a user-friendly interface in order to be operated under heavy stress.
(4) Security: Communication among members of the rescue teams and transmission of data about the victims must be done in a secure way in order to assure integrity and confidentiality of the data being transmitted. Moreover, use of the system (i.e. especially communication bandwidth) by non-authorized persons must be prevented.

The e-Triage approach consists of four main elements: autonomous communication infrastructure, electronic data recording, a distributed database system, and psychological acceptance research. In more details, the technical concept comprises a satellite-based communication system with terrestrial radio cells, matching end devices with dedicated application software for the registration of victims, and a distributed, self-synchronizing database system guaranteeing maximal availability without a single point of failure.

2 Global Architecture

2.1 Communication

During a rescue operation two different areas can be defined: the on-site segment (OSS) with rescue forces in the field and the disaster-safe segment (DSS) with remote coordination facilities. Both segments can be connected using a backhaul communication solution, but an autonomous local coordination of the operation in the field has to be possible, too. It was decided on this basic approach to be able to represent two disjunct areas of security level inside the system (see necessity (4)). At the same time a high level of scalability for the system is secured (see necessity (1)).

Normally, the different elements found in the OSS are mobile actors (see section 2.2) and one or more on-site emergency communications equipment (OSECE). An overview of the general communication architecture is presented in Figure 1 and a detailed description of the satellite-based OSECEs can be found in [VWD10]. Mobile actors' user terminals must have at least intermittent connectivity with one of the OSECEs in order to transmit their data regarding the victims' registration. These terminals are typically laptops or PDAs equipped with WLAN and GSM/GPRS interfaces, GSM/GPRS smart phones,

or TETRA handhelds, supporting voice and data. The authentication and authorization

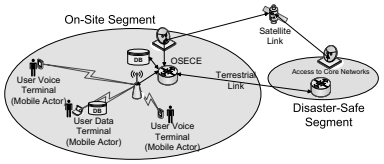


Figure 1: General e-Triage system architecture overview.

on-site is different depending on the access technology: WLAN, GSM/GPRS, TETRA (optional), and DECT (optional). WLAN will use WPA(2)-PSK for access control, and to simplify WLAN access certificates will not be used.

2.2 On-Site Segment (OSS)

In the disaster area there are scores of mobile actors working independently in different roles. A mobile actor can be a physician, doing the classification at the scene, or can be the emergency squad leader who is using the reported data to get an overview of the situation. Earlier works (e.g., [NK09]) have already shown that using electronic devices for MCI data gathering is a worthwhile approach.

The architecture of the end device consists of three main parts: the *user interface*, the *local database instance*, and the *communication unit*. By grouping the functional entities application/database and database/network the user will have a maximum of application performance by writing his data directly into the local database. In the background the database/network entity redistributes the data across the whole system. The flow of events when a mobile actor enters data using an application is used in this section to describe the functional architecture of the mobile actor's end device (Figure 2). In the application/database entity, the steps followed by the mobile actor are:

1. The user starts the e-Triage user interface (e.g., classification, reporting).
2. The user enters data. This can be in the context of the first or advanced classification.
2a. The user enters open data (e.g., classification color, sex), which is written directly into the local database instance.
2b. The user enters secure/advanced data (e.g., name, address). Secure data is locally encoded by an advanced encryption standard (AES) [Ert07] (e.g., Twofish, Serpent).

Necessity (3) "ease of use" has to be the main requirement for the user interface because the acceptance of the overall system will be a matter of the GUI design.

The services related to the database/network entity are:

1. The local database instance makes a local redundant copy of itself on a secondary storage medium (e.g., flash-drive). This is necessary if the primary medium fails or the user terminal cannot be used anymore for some reason, so the mobile actor can use the data with another terminal.

2. The communication unit decides which medium can be used to send the local data to the nearest OSECE (e.g., WLAN, GSM/GPRS) and notifies the local database system.
3. The local database service sends the local data to the nearest OSECE by using the selected network adapter.

In reverse direction, when the application requests data to display (e.g., a report for the squad leader), the local database instance synchronizes itself with the nearest OSECE. The problem of synchronizing distributed databases residing in user terminals and OSECEs via heterogenous networks is another central objective of the e-Triage project and is briefly described in section 2.4. e-Triage local applications implement programmatic authentication,

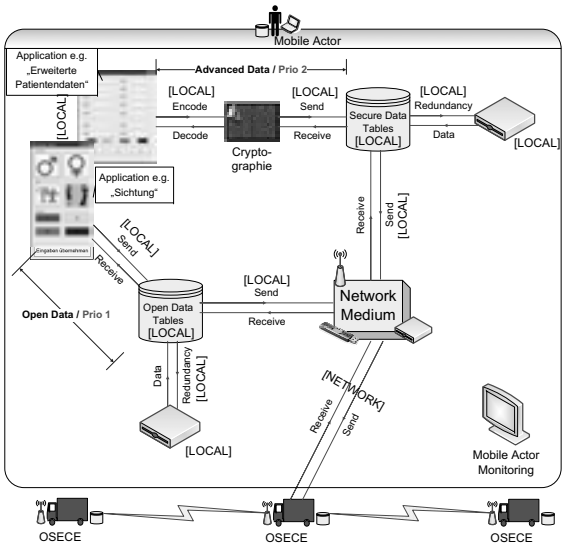


Figure 2: e-Triage mobile actor.

requiring the user to explicitly supply credentials, which are then validated using the local operating system security mechanisms with its underlying database. It is ensured that the secure messages stay private, because they are encrypted/decrypted inside the mobile actor's device. e-Triage uses the approach of authorization through roles. Various roles are created for the envisaged system and respective access rights are assigned to these roles. When each party is authenticated, the credentials applied to the party are mapped to the party's role (through groups in the directory service), which then determines whether or not the party is authorized to access a particular function. In the DSS everyone has to be certified by a security token service, whereas on-site the application signs the messages.

2.3 Disaster-Safe Segment (DSS)

An OSECE establishes the secure interface between OSS and DSS. In the DSS, control centers or hospitals have connection to the e-Triage gateway. To enable this connection,

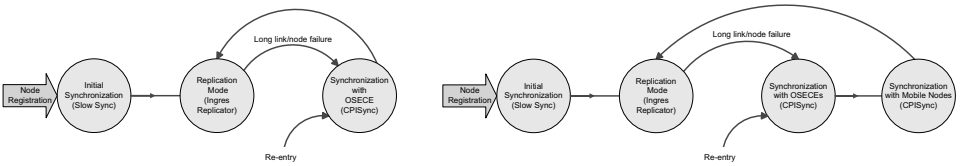
secure tunnels will be built statically because all nodes are known beforehand, including the OSECEs. Each request has to pass through the e-Triage gateway to get authorization for writing or reading from the database server inside a DSS.

e-Triage is based upon a service oriented architecture (SOA) which was chosen to give the system a maximum of adaptiveness to other systems. To make SOA secure, external security is used against outside attacks together with internal security, so that e-Triage supports data confidentiality according to legal requirements. The users in the DSS, namely hospital personnel or control center operators, are authenticated through web-based mechanisms when they connect to the web server. These mechanisms follow a sophisticated mutual certificates authentication schema, based on our specialized applications databases. On software level they communicate via web services [wsa04], where the use of the emergency data exchange language (EDXL) [RWA06] may be considered. There are several broad architectural options for implementing security in the context of web services. For e-Triage it was decided to use an XML gateway. The XML gateway enforces access control rules by processing security tokens contained within incoming messages, and by ensuring that the XML format and content are appropriate for the target. It may use the security assertion markup language (SAML) [CKPM05] to establish the authentication status of an end user or to request attribute information, which is used to make an access control decision. e-Triage strives to re-use the existing security infrastructure, including the pre-configured users, groups, and roles. To do this the XML gateway contains security adapters to existing security technologies such as lightweight directory access protocol directories, traditional firewalls, and public-key-infrastructure. If the incoming traffic bypasses the gateway and reaches the service endpoint, it would also bypass the security provisions implemented by the gateway. Consequently, using this architecture requires additional security provisions. In case of e-Triage the physical nodes, from which the service endpoint can receive service traffic, are restricted. Auditing (internal and external monitoring) is used by each application on the mobile actor's device, the OSECEs and the entities of the DSS.

2.4 Database Management

The current available replication solutions for distributed database systems (DDBSs) do not match the requirements of the e-Triage project, in which there are two main challenges [TDCM10]: (i) Replicated data should be transmitted directly or indirectly in a quasi real-time way towards any nodes existing in the entire system, which have a relatively stable network connection available at the moment. The replication system needs to be robust enough to face temporary (short-term) network interruptions. (ii) In case of intermittent long-term network outages or dynamic changes of the network architecture, the replication system should have an efficient comparison and synchronization solution by using a minimum of communication to retrieve the consistency of data sets when the network connection becomes available again. These requirements define two different modes of operation, which are *quasi real-time replication* and *comparison and re-synchronization*. Both modes must be seamlessly integrated without any user interaction. Nevertheless, it is necessary to make a distinction between the processes performed by an OSECE, and the

operation of the mobile nodes on the field (Figures 3(a) and 3(b)).



(a) Mobile node operation.

(b) OSECE operation.

Figure 3: State diagrams of the mobile node and OSECE operation.

3 Conclusions

The paper has briefly described the global architecture of the e-Triage system. On the one hand data management in MCIs is challenging, and on the other hand a key issue is the development of graphical user interfaces which are intuitive and self-explaining without causing additional stress. The consortium is currently implementing the system, and for 2011 a series of lab tests and field trials are planned. Although earlier and other current works (e.g., [sog09]) have studied single components of electronic MCI management, a unique feature of e-Triage is that for the first time a coherent overall concept consisting of mobile devices, location-independent communication infrastructure, database management, and psychological acceptance research is studied.

References

[CKPM05] Scott Cantor, John Kemp, Rob Philpott, and Eve Maler. Assertions and Protocols for the OASIS Security Assertion Markup Language (SAML). Technical Report v. 2.0, OASIS Security Services TC, March 2005.

[Ert07] Wolfgang Ertel. *Angewandte Kryptographie*. Hanser Verlag, 3rd edition, October 2007.

[etr10] Elektronische Betroffenenerfassung in Katastrophenfällen. [Online]. Available: `http://www.e-triage.de`, 2010. In German.

[NK09] S. Nestler and G. Klinker. Mobile computing in mass casualty incidents (MCIs). In *Mobiles Computing in der Medizin (MoCoMed)*, Lübeck, Germany, September 2009.

[RWA06] Michelle Raymond, Sylvia Webb, and Patti Iles Aymond. Emergency Data Exchange Language (EDXL) Distribution Element. Technical Report v. 1.0, OASIS Emergency Management TC, 2006.

[sog09] SOGRO - Sofortrettung bei Großunfall mit Massenanfall von Verletzten. [Online]. Available: `http://www.c-lab.de/de/forschungsprojekte/sogro/index.html`, 2009. In German.

[TDCM10] Chen Tang, Anton Donner, Javier Mulero Chaves, and Muhammad Muhammad. Performance of Database Synchronization via Satellite. In *Advanced Satellite Multimedia Systems (ASMS) Conference*, Cagliari, Italy, September 2010. Accepted for presentation.

[VWD10] Angels Via, Markus Werner, and Anton Donner. Satellite Communications for Management of Mass Casualty Incidents: The e-Triage Project. In *International Conference on Satellite and Space Communications (ICSSC)*, Anaheim, California, USA, August/September 2010. AIAA. Accepted for presentation.

[wsa04] Web Services Architecture. [Online]. Available: `http://www.w3.org/TR/ws-arch/`, February 2004.

Workshop

5. Workshop: Vorgehensmodelle in der Praxis – Lösung oder Ursache aller Probleme?

Marco Kuhrmann, Patrick Keil, André Schnackenburg und Oliver Linssen

Vorgehensmodelle in der Praxis – Lösung oder Ursache aller Probleme?

Vorwort

Obwohl von vielen als „alter Hut“ abgetan, werden Vorgehensmodelle in den letzten Jahren verstärkt diskutiert. Die Diskussion findet auf verschiedenen Ebenen statt, die von der strikten Ablehnung bis hin zur zertifizierten Umsetzung eines umfassenden Standards reichen. In diesem Spannungsfeld sind insbesondere die Facetten der Agilität, der Anpassbarkeit sowie der Werkzeugunterstützung und Weiterentwicklung einzuordnen. Für viele dieser Bereiche sind in den vergangenen Jahren unterschiedliche Lösungen entstanden, die durch ihre Vielfalt die Auswahl und Abstimmung einzelner Komponenten erschweren. Trotzdem setzen viele Unternehmen verstärkt auf die Etablierung des *strategischen* Elements Vorgehensmodell. Erleben wir einen zyklisch wiederkehrenden Hype oder ist ein gesteigertes Bewusstsein für die Notwendigkeit strukturierter Vorgehensweisen festzustellen?

Dieser zum fünften Mal stattfindende Workshop „Vorgehensmodelle in der Praxis“ widmet sich, gewissermaßen als Zwischenfazit, dem Thema: Wie geht es mit Vorgehensmodellen weiter? Besonderes Interesse liegt dabei auf den Fragestellungen, ob die heute verfügbaren Vorgehensmodelle bereits eine Reife besitzen, die für die Anforderungen aktueller (und zukünftiger) Projekte ausreichend ist. Oder falls nicht, welchen Anforderungen Vorgehensmodelle stattdessen gerecht werden müssten. Damit einher gehend stellen sich dann die Fragen: Beinhalten heutige Vorgehensmodelle zu viel, zu wenig und das richtige? Adressieren bzw. Lösen heutige Vorgehensmodelle die Probleme der Praxis? Wie viel oder wie wenig davon sollen und können überhaupt erfasst werden? Weiterhin ist es ein Bestreben des Workshops festzustellen, ob die Menge der Vorgehensmodelle durch Standards konsolidiert werden kann oder ob mit einer weiteren Diversifikation auf Basis einiger stabiler Grundprinzipien zu rechnen ist.

Von besonderem Interesse sind neben Erfahrungen aus dem agilen Umfeld auch Lösungsansätze aus dem Bereich der Standardvorgehensmodelle wie z.B. RUP, Prince2 und V-Modell XT (das zufälligerweise ebenfalls seinen fünften Geburtstag feiert).

Inhalte des Workshops

In den vier ausgewählten Beiträgen werden unterschiedliche Facetten dazu aufgegriffen: Im Bereich der Erfahrungen und Erfolgsfaktoren der Einführung von Vorgehensmodellen stellen Wind und Schrödl eine Fallstudie zur Einführung agiler Projektmanagementmethoden vor. Auch Bombosch et. al widmen sich der Einführung agiler Verfahren, legen den Fokus jedoch nicht nur auf Entwicklungsprojekte sondern auch auf die ausführenden Organisationen. Einen stärkeren Bezug zur Werkzeugen such der Beitrag von

Romahn, in dem die Frage der Einführung eines Vorgehensmodells im Kontext der Werkzeugunterstützung als Akzeptanzkriterium diskutiert wird. Lewitz stellt ergänzend die grundsätzliche und Frage, warum auch Projekte einen Business Case benötigen und diskutiert diese Frage im Rahmen einer Gegenüberstellung verschiedener Vorgehensmodelle.

Neben den eingereichten Beiträgen aus Forschung und Praxis beleuchten eingeladene Beiträge aus Wirtschaft, Verwaltung und Forschung verschiedene Aspekte von Vorgehensmodellen. Insbesondere wird hier das Spannungsfeld zwischen dem Bedürfnis nach Verlässlichkeit der Auftraggeberseite und der Anforderung der Flexibilität durch Software herstellende Unternehmen aufgezeigt. Auch dem V-Modell XT wird an dieser Stelle zum 5. Geburtstag gratuliert.

Dem Programmkomitee gehörten folgende Personen an:

Prof. Dr. Dr. h.c. Manfred Broy	TU München
Dr. Jens Calamé	Hamburg
Jens Coldewey	it-agile GmbH, München
Dr. Christof Ebert	Vector, Stuttgart
Jan Friedrich	4Soft GmbH, München
Dr. Thomas Greb	freiberuflicher Berater, Bremen
Patrick Keil	TU München
Dr. Marco Kuhrmann	TU München
Dr. Ralf Kneuper	freiberuflicher Berater, Darmstadt
Dr. Christian Lange	BVA/BIT, Köln
Dr. Oliver Linssen	Liantis GmbH & Co. KG
Dr. Jürgen Münch	Fraunhofer IESE, Kaiserslautern
Prof. Dr. Roland Petrasch	TFH Berlin
Prof. Dr. Andreas Rausch	TU Clausthal
Dr. André Schnackenburg	BIT, Köln
Thomas Ternité	TU Clausthal

Wir bedanken uns bei allen Beteiligten des Workshops und der Tagungsleitung vor Ort in Lübeck.

Marco Kuhrmann, Patrick Keil, André Schnackenburg und Oliver Linssen

Einführung agiler Projektmanagementmethoden für Softwareentwicklung in der IuK-Industrie – Fallstudie und Handlungsempfehlungen

Stefan Wind, Holger Schrödl

Lehrstuhl für Wirtschaftsinformatik und Systems Engineering
Universität Augsburg
Universitätsstr. 16
86159 Augsburg
stefan.wind@wiwi.uni-augsburg.de
holger.schroedl@wiwi.uni-augsburg.de

Abstract: Agile Vorgehensmodelle haben zum Ziel, die Erfolgsbilanz von Projekten – insbesondere im Bereich der Softwareentwicklung – zu erhöhen. Trotz erster positiver Ergebnisse in der Praxis werden die Rahmenbedingungen, Möglichkeiten und Effekte agiler Vorgehensmodelle in der Fachliteratur und im Projektalltag kontrovers diskutiert. In diesem Beitrag wurden im Rahmen einer Fallstudie für ein repräsentatives Projekt in der IuK-Industrie die Anwendungsmöglichkeiten verschiedener agiler Methoden untersucht und bewertet. Es werden die Ergebnisse sowie die gesammelten Erfahrungen aus der Anwendung der ausgewählten agilen Methode dargelegt. Auf diese Weise werden allgemeingültige sowie methodenspezifische Einführungsvorschläge für die praktische Anwendung abgeleitet. Bei der Einführung agiler Vorgehensmodelle in vergleichbaren Projekten kann somit auf diese Handlungsempfehlungen zurückgegriffen werden.

1 Motivation

Unternehmen allgemein und im speziellen im Bereich der Informations- und Kommunikationsindustrie (IuK) stehen vor der Herausforderung, im Projektmanagement adäquat mit einem dynamischen Umfeld, sich ändernden Zielen, Märkten und Anforderungen umgehen zu können [Ho08], [Gl10]. Klassische Vorgehensmodelle stoßen im Hinblick auf die veränderten Rahmenbedingungen häufig an ihre Grenzen. Agile Vorgehensmodelle bieten hier eine Verbesserung, sind jedoch nur selten „out of the box“ anwendbar [St10], [Sc10]. Die Frage bleibt, welche Faktoren die Einführung einer agilen Methode positiv beeinflussen können. Das Ziel dieses Beitrags ist es daher, die Erfahrungen in der Anwendung einer agilen Methode auf Grundlage einer Fallstudie darzustellen und praktische Handlungsempfehlungen für die Einführung zu geben.

2 Agile Vorgehensweisen

Im Jahr 2001 wurde das agile Manifest [Ag01] veröffentlicht, in dem siebzehn namhafte

Vordenker (u.a. Kent Beck, Martin Fowler, Jim Highsmith und Bob Martin) eine Grundsatzerklärung über die Werte eines agilen Vorgehens abgeben. Um die an einem agilen Projekt beteiligten Personen in ihrem Handeln zu unterstützen, wurden aus dem agilen Manifest Prinzipien für die Softwareentwicklung abgeleitet [DK05], [Ba08], [Ob08]. Diese wurden zum Teil von bestehenden Vorgehensmodellen aufgegriffen (z.B. die Entwicklung vom V-Modell hin zum V-Modell XT) oder ganz neue agile Projektmanagementmodelle (z.B. SCRUM oder Extreme Programming) gingen daraus hervor. Agile Vorgehensmodelle wie SCRUM finden in der Praxis vermehrt Anwendung [Ge09] und in der Literatur wird schon von einem Paradigmenwechsel gesprochen [Gl10].

2.1 Agiles Projektmanagement

In der Literatur existieren für den Begriff des agilen Projektmanagements unterschiedliche Beschreibungen und Definitionen, die alle auf den agilen Werten und Prinzipien beruhen. Schröder sieht diese Form des Projektmanagements bezogen auf den Begriff „agil“ als „leichtes“ oder „leichtgewichtiges“ Projektmanagement an und betont die Vorteile einer geringen Führungsintensität in solchen Projekten [Sc04]. Highsmith sieht in dieser Form das aktive Steuern eines Projekts; nicht mit dem Ziel einer Projektverwaltungstätigkeit, sondern eines klaren Mehrwerts für den Kunden [Hi04]. Gernert geht hingegen in ihrer Definition speziell auf die praktische Eignung ein und definiert es als eine Sammlung von Vorgehensweisen, die sich in der Praxis schon bewährt haben [Ge03]. Das Projekt Magazin sieht es als einen Oberbegriff für verschiedene agile Vorgehensmodelle in der Softwareentwicklung (z. B. SCRUM) [An09]. Diese Erweiterung der Betrachtungsweise ist nicht unwesentlich, denn ohne passende Vorgehensweisen und Methoden kann auch ein agiles Projektmanagement nicht zum Erfolg führen.

2.2 Ziele für agile Vorgehensweisen

Mit der Einführung agiler Vorgehensweisen werden in jedem Unternehmen unterschiedliche Zielsetzungen verfolgt. Die wesentlichen Teilziele sind dabei eine Verkürzung der Einführungszeit, eine Qualitätsverbesserung, die Reduktion von Ausschuss, eine verbesserte Voraussagbarkeit und eine Verbesserung der Moral im Team [Ec04], [Scc07], [Ob08], [Gl10].

2.3 Forschungsmethode

Als zentrale Forschungsmethode wird in diesem Beitrag die Fallstudie verwendet. Diese wurde auf Basis einer umfassenden Literaturanalyse konzipiert und während der Durchführung konsequent wissenschaftlich begleitet. Erkenntnisse aus den Ergebnissen der Fallstudie wurden unter Annahme einer Ähnlichkeitsbewertung induktiv in Handlungsempfehlungen für zukünftige Szenarien überführt. Es wurde dabei ein konstruktionsorientierter Zugang verwendet [WH07].

3 Beschreibung der Fallstudie

3.1 Ausgangssituation

Das im Rahmen der Fallstudie betrachtete Unternehmen ist im Bereich der Informations- und Kommunikationsindustrie (IuK) angesiedelt und bietet weltweit Produkte und Servicedienstleistungen an. Die Anwendung des agilen Projektmanagements erfolgte erstmalig im Rahmen des Projekts „Design und Implementierung eines Management Cockpits für den Bereich Finance und Controlling mit SAP Strategic Enterprise Management“. Das Projekt verteilte sich in seinem Umfang von ca. 300 Personentagen auf acht Mitarbeiter, welche am selben Standort tätig waren. Die beteiligten Mitarbeiter verfügten noch über keinerlei praktische Erfahrungen in der Anwendung agiler Vorgehensmodelle. In den Bereichen, die an dem Projekt beteiligt waren, bestand seit Jahren ein organisationsspezifisches Projekthandbuch, das das Vorgehen im Projektmanagement regelte. Dieses Vorgehensmodell war wasserfallartig aufgebaut und es waren weder weitergehende Rückkopplungen noch Iterationen zwischen den einzelnen Phasen vorgesehen. Das bisherige Vorgehen lieferte insbesondere im Hinblick auf das Projektergebnis unbefriedigende Ergebnisse. Aus diesem Grund fiel die Entscheidung, agile Vorgehensmodelle im Unternehmen zu evaluieren und in der praktischen Anwendung zu erproben.

3.2 Durchführung

3.2.1 Auswahl der agilen Methode

Im Rahmen des Auswahlprozesses wurden drei agile Methoden (SCRUM, XP, FDD) anhand einer ausführlichen Analyse der Projektziele, -eigenschaften und -risiken untersucht und in Rahmen einer Bewertungsmatrix gegenübergestellt. Als Ergebnis wurde SCRUM als für das Projekt geeignete Methode ausgewählt, da es gegenüber anderen Methoden zwei entscheidende Vorteile bietet. Zum einen eignet sich SCRUM gut im Falle vager Anforderungen und zu antizipierenden chaotischen Bedingungen [Scc07]. Zum anderen lässt sich das Verfahren knapp beschreiben und leicht erlernen, was bei einer agilen Erstdurchführung von Vorteil ist [BW08]. Dagegen wiesen sowohl FDD als auch XP entscheidende Nachteile auf. So wäre es schwer gewesen, mit FDD ein stabiles Gesamtmodell zu erstellen. XP fordert zum einen radikal neue Vorgehensweisen, die oft Unverständnis bzw. Abwehr bei den Beteiligten hervorrufen. Zum anderen liegt der Fokus von XP stark auf programmiertechnischen Ansätzen, wodurch die Konzeption – im vorliegenden Fall entscheidend wichtig – zu kurz gekommen wäre [Co02].

3.2.2 Durchführung des agilen Projektmanagements

Um SCRUM erstmalig im Projekt einsetzen zu können, war eine Vorbereitung der Organisation und der betroffenen Mitarbeiter notwendig. Im Projektrahmen waren die verfügbaren Möglichkeiten jedoch begrenzt. So konnte weder eine Organisationsänderung noch eine groß angelegte Einführung als eigenständiges SCRUM-Projekt durchgeführt werden. Im Rahmen der Möglichkeiten wurde somit nachfolgende Vorbereitung durchgeführt:

- Änderungsbedarf anhand einer Fallstudie aufzeigen
- Agiles Projektmanagement anhand eines praktischen Beispiels vorstellen
- Sicherung der Management-Unterstützung durch Aufzeigen von Potenzialen
- Ausgewähltes Pilotprojekt anhand des Projektauftrags vorstellen
- Beteiligte Mitarbeiter intensiv schulen

Daran anschließend wurden die Rollen „Product Owner“, „SCRUM Master“ und „Team“ besetzt. Für die Rolle des SCRUM-Masters wurde ein zertifizierter und erfahrener, externer Mitarbeiter eingesetzt. Dieser verfügte über umfassende Erfahrung in der Ein- und Durchführung agiler Vorgehensmodelle, die im Unternehmen aufgrund der erstmaligen Anwendung noch nicht vorhanden war. Nach Durchführung der vorbereitenden Schritte wurde das Projekt mit dem gewählten Modell realisiert. Zusätzlich wurde zu Beginn neben dem Produktbacklog ein kontinuierlich angepasster Releaseplan erstellt sowie eine für alle zugängliche Sprint-Infoseite angelegt.

3.3 Bewertung und Erfahrung für das konkrete Projekt

Der unklaren Ausgangslage des Projekts konnte mit SCRUM gut begegnet werden. So war es zu Beginn nicht möglich, alle Anforderungen vollständig und verlässlich zu erfassen. Hinzu kam, dass großteils nur vage oder sogar konträre Vorstellungen über das zu erstellende Cockpit bestanden. Das bisherige wasserfallmodellbasierte Projektvorgehen hätte hier aus den Erfahrungen der letzten Projekte mit ähnlichen Rahmenbedingungen zu Problemen geführt. Die Aufteilung des Projektablaufs in mehrere Sprints erlaubte es hingegen, Anforderungen schrittweise umzusetzen und frühzeitig Feedback von den zukünftigen Anwendern zu erhalten. Im Gegenzug konnten diese die korrekte Umsetzung ihrer Anforderungen zeitnah überprüfen und bei Bedarf bis dato nicht berücksichtigte, zusätzliche Anforderungen in das Projekt einzufließen lassen. Somit profitierten beide Seiten durch die frühzeitige und iterative Vorstellung des entstehenden Cockpits und die Möglichkeit der produktiven Nutzung bereits fertiggestellter Teile. Neben den Vorteilen brachte das agile Vorgehen aber auch Herausforderungen und Probleme mit sich. So galt es zunächst, in der Organisation falsche Vorstellungen von einem agilen Projektvorgehen und Widerstände dagegen zu überwinden.

Dies beinhaltete Argumente wie „Ein wichtiges Projekt muss mit den klassischen und gewohnten Methoden bearbeitet werden“ zu entkräften und falschen Hoffnungen wie „Toll, endlich keine Regeln und Statusberichte mehr!“ in bilateralen Gesprächen entgegenzuwirken. Eine weitere Herausforderung lag darüber hinaus in der Vertragsgestaltung für den Einkauf externer Ressourcen. Obwohl es in der Literatur bereits einige Empfehlungen für die Gestaltung von Festpreisverträgen in agilen Projekten gibt (siehe [CP03]), gelang der Abschluss eines bis dahin üblichen Festpreisvertrages nicht. Die Lösung lag schließlich in einem Vertrag nach Aufwand mit definierter Obergrenze.

Zusammengefasst hat sich SCRUM als agile Managementmethode für das durchgeführte Pilotprojekt bewährt. Das Projekt wurde erfolgreich abgeschlossen und hat die Erwartungen der Beteiligten erfüllt. Insbesondere konnten die Ziele, die in Kapitel 2.2 aufgeführt sind, erreicht werden.

3.4 Handlungsempfehlungen

Aufbauend auf den gesammelten Erfahrungen und bestehenden theoretischen Empfehlungen (siehe [Ec04], [Hi04], [Co05], [Sc07], [Pi08]) sollen nachfolgend Hinweise gegeben werden, die bei der Anwendung agiler Vorgehensweisen in anderen Bereichen und Unternehmen helfen können (Tabelle 1).

Grundlegende Empfehlungen	• Beteiligte Personen und Unternehmensbereiche sind auf die zu erwartenden Neuerungen vorzubereiten. • Bei der Auswahl des Pilotprojekts ist darauf zu achten, dass es geschäftsrelevant und typisch für das Umfeld, nicht aber überlebenswichtig ist. • Im Vorfeld sind die Projekteigenschaften und Risiken mit einer Bewertungsmatrix zu ermitteln und kontinuierlich zu überprüfen. • Es sind je nach Projekt angemessene Methoden auszuwählen und jene kontinuierlich auf ihre Praktikabilität hin zu überprüfen.
SCRUM-spezifische Empfehlungen	• Die wenigen klar definierten Regeln in SCRUM sind absolut konsequent einzuhalten, um den Erfolg der Methode sicherzustellen. • Der Product Owner soll neben seiner fachlichen Kompetenz auch über entsprechende Entscheidungsbefugnis verfügen. • Alle benötigen Rollen, die zur Umsetzung der Projekt- und Sprintziele benötigt werden, müssen auch im Team vertreten sind. • Für die erstmalige Einführung empfiehlt es sich, einen zertifizierten und erfahrenen externen SCRUM-Master einzusetzen. • Anlage einer Sprint-Infoseite, damit alle Stakeholder aus erster Hand über den Stand des Projekts informiert sind und das Aufkommen von Gerüchten und das Treffen von Annahmen möglichst unterbunden werden. • Verwendung eines Releaseplans, um den Fertigstellungstermin, Funktionalität und Kosten leichter miteinander in Einklang bringen zu können.

Tabelle 1: Übersicht der Handlungsempfehlungen

4 Zusammenfassung und Ausblick

Agile Vorgehensmodelle sollen helfen, den Projekterfolg zu erhöhen, wo klassische Modelle an ihre Grenzen stoßen. Dies ist heute insbesondere in der IuK-Industrie von Relevanz. Der vorliegende Beitrag lieferte anhand der Erfahrungen aus einer Fallstudie in der Praxis Empfehlungen bei der Einführung eines agilen Vorgehensmodells in der IuK-Industrie. Im Rahmen dieser Fallstudie wurde eine erstmalige Einführung agiler Methoden für ein Softwareentwicklungsprojekt durchgeführt. Auf Basis einer Literaturstudie wurde SCRUM als für das Projekt geeignete Methode ausgewählt. Die Handlungsempfehlungen wurden induktiv abgeleitet. Im Einzelnen konnten vier allgemein gültige sowie sechs methodenspezifische Empfehlungen für die Einführung herausgearbeitet und vorgestellt werden. Weiterer Forschungsbedarf besteht hinsichtlich der Erprobung in weiteren Branchen, in der Kombination mit klassischen Vorgehensmodellen sowie in der Einführung in einem interkulturellen Umfeld.

Literaturverzeichnis

[Ag01] Agilemanifesto: Manifesto for Agile Software Development. In: http://www.agilemanifesto.org/, 2001, Abruf am 07.04.2010.

[An09] Angermeier, G.: Agiles Projektmanagement. In: http://www.projektmagazin.de/glossar/gl-0775.html, 2009, Abruf am 19.04.2010.

[Ba08] Balzert, H.: Lehrbuch der Softwaretechnik: Softwaremanagement. 2. Aufl., Spektrum Akademischer Verlag, Heidelberg, 2008.

[BK08] Bunse, C.; Knethen v. A.: Vorgehensmodelle kompakt. 2. Aufl., Spektrum Akademischer Verlag, Heidelberg, 2008.

[BW08] Bleek, W.; Wolf, H.: Agile Softwareentwicklung -Werte, Konzepte und Methoden. dpunkt Verlag, Heidelberg, 2008.

[Co02] Coldewey, J.: Multi-Kulti: Ein Überblick über die agile Entwicklung. In: ObjektSpektrum Nr. 1 2002.

[Co05] Coldewey, J.: Management am Rande des Chaos: Die Führung agiler Prozesse. In: ObjektSpektrum Nr. 1 2005.

[DK05] Dogs, C.; Kimmer, T.: Agile Softwareentwicklung kompakt. mitp-Verlag, Bonn, 2005.

[Ec04] Eckstein, J.: Agile Softwareentwicklung im Großen - Ein eintauchen in die Untiefen erfolgreicher Projekte. dpunkt Verlag, Heidelberg, 2004.

[Gc03] Gernert, C.: Agiles Projektmanagement - Risikogesteuerte Softwareentwicklung. Carl Hanser Verlag, München, Wien, 2003.

[Ge09] Gregor E. et. Al.: Sicherstellen der Betrachtung von nicht-funktionalen Anforderungen in SCRUM-Prozessen durch Etablierung von Feedback. In(Fischer, S; Maehle, E.; Reischuk R.: Informatik 2009 - Im Focus das Leben. , LNI, vol. 154, pp. p. 458, 2009.

[Gl10] Gloger, B.: SCRUM Der Paradigmenwechsel im Projekt- und Produktmanagement - Eine Einführung. In: Informatik Spektrum: V. 33, Number 2 / April 2010, S. 195-200.

[Hi04] Highsmith, J.: Agile Project Management - Creating Innovative Products. Addison-Wesley Verlag, Boston u. a. 2004.

[Ho08] Hoffmann, K.: Projektmanagement heute. In: (Hoffmann, K.; Mörike M. Hrsg.): IT-Projektmanagement im Wandel. HMD-Praxis der Wirtschaftsinformatik, Heft 260, dpunkt Verlag, Heidelberg, 2008.

[Ob08] Oestereich, B.: Agiles Projektmanagement. In: (Hoffmann, K.; Mörike M. Hrsg.): IT-Projektmanagement im Wandel. HMD-Praxis der Wirtschaftsinformatik, Heft 260, dpunkt Verlag, Heidelberg, 2008.

[Pi08] Pichler, R.: Scrum - Agiles Projektmanagement erfolgreich einsetzen. dpunkt Verlag, Heidelberg, 2008.

[Sc04] Schröder, R.: Einsatz agiler Verfahren im Management eines Großprojekts. In: (Engelien, M.; Meißner, K. Hrsg.): Virtuelle Organisation und neue Medien 2004, Eul Verlag, Dresden, S. 249-257.

[Scc07] Schwaber, C.: The Truth About Agile Processes - Frank Answers To Frequently Asked Questions. In: Forrester Research, http://www.forrester.com, 2007, Abruf am 01.04.2010.

[Sck07] Schwaber, K.: Agiles Projektmanagement mit SCRUM. Microsoft Press Deutschland Verlag, Unterschleißheim (2007).

[Sc10] Schatten, A et. Al.: Eine praxiserprobte Zusammenstellung von komponentenorientierten Konzepten, Methoden und Werkzeugen. Spektrum-Verlag, Heidelberg, Berlin, 2010.

[St10] Standish Group: CHAOS Report. In: http://www.standishgroup.com/, Abruf am 23.04.2010.

[WH07] Wilde, T., Hess. T.: Forschungsmethoden der Wirtschaftsinformatik – Eine empirische Untersuchung. In. Wirtschaftsinformatik, 49(4), 2007; S. 257-266.

Warum Projekte einen Business Case brauchen

Olaf Lewitz

microTOOL GmbH
Voltastr. 5
13355 Berlin
Olaf.Lewitz@microTOOL.de

Abstract: Ein Business Case macht die Investition in ein Projekt und den erwarteten Nutzen transparent. Obwohl die Herstellung dieser Transparenz eine der wichtigsten Aufgaben des Projektmanagements ist, vernachlässigen einige Vorgehensmodelle wie z.B. das V-Modell XT diesen Aspekt eines Projektes. Nach wie vor scheitern zwei Drittel aller Projekte[1] – eher an Management-Fehlern als an technischen Problemen. In Anbetracht dieser Zahlen ist es fahrlässig, dass wir Vorgehensmodelle verwenden, die für ein Projekt keinen Business Case verlangen. Mithilfe einer Gegenüberstellung von PRINCE2 und V-Modell XT wird die Frage diskutiert, wie die Erfolgswahrscheinlichkeit von Projekten durch bessere Kosten-/Nutzen-Transparenz erhöht werden kann.

„The Stars might lie, but the numbers never do."

– Mary Chapin Carpenter, „I Feel Lucky"

1 Projekte sind Investitionen

Projekte kosten Geld. Jemand investiert dieses Geld, um ein bestimmtes Ziel zu erreichen, eine Vision Realität werden zu lassen. Gutes Projektmanagement bedeutet, den Nutzen dieser Investition so transparent wie möglich zu machen. Dem Investor sollte möglichst klar sein, wie viel Geld bereits investiert wurde und wofür, und wie viel Geld noch investiert werden soll, um welchen weiteren Nutzen zu erreichen. Genau diese Punkte werden in einem Business Case festgehalten.

Ein Kollege hat eine schöne Definition für einen Business Case formuliert: „*The business case is the vision captured in numbers*" – Der Business Case ist die in Zahlen gefasste Vision. Wobei Zahlen für die Vision natürlich nicht ausreichen, ich mich hier aber auf die Zahlen konzentrieren möchte.

[1] Standish Group, CHAOS Summary 2009, http://www1.standishgroup.com/newsroom/chaos_2009.php
Die diesem Bericht zugrunde liegende Definition von Projekterfolg halte ich für fragwürdig, weil sie den ROI nicht berücksichtigt. An der Größenordnung der gescheiterten Projekte muss das jedoch nichts ändern.

Warum Zahlen? Zahlen sind eine *universal language*, jeder im Projekt versteht, was sie bedeuten; vor allem, wenn sie für Geldbeträge stehen. Wenn also der zu schaffende Wert für jeden im Projekt fassbar und transparent ist, ist das Ziel des Business Case erreicht.

Der Business Case soll Handlungsoptionen für den Auftraggeber transparent machen. Es gibt zunächst mindestens zwei: Nichts zu tun – den Status quo erhalten – und die Durchführung des Vorhabens. Aber der Business Case spielt nicht nur bei der „großen" Entscheidung für oder gegen das Projekt eine Rolle. Ein guter Business Case bietet eine Orientierung für alle Entscheidungen, die in einem Projekt getroffen werden müssen, die großen wie die sehr kleinen.

Jeder Projektmitarbeiter muss jeden Tag Entscheidungen treffen. Diese kleinen Entscheidungen bestimmen ganz wesentlich die Qualität des erstellten Produktes. Falls sichergestellt ist, dass diese Entscheidungen immer im Sinne der Wertschöpfung für den Kunden getroffen werden, kann die Qualität im Detail stark zunehmen. Ist dem einzelnen Mitarbeiter sein Beitrag zur Wertschöpfung des Projektes unklar, so bleibt die Qualität seiner Entscheidungen dem Zufall überlassen. Das wäre für mich als Kunden ein untragbares Risiko.

Alle Projektmitarbeiter und Stakeholder können ihre Ideen, ihre Kreativität und vor allem ihre unterschiedlichen Erfahrungen, die den potentiellen Lösungsraum vergrößern, in ein Projekt mit einbringen. Allerdings macht nicht jedes Vorgehen dieses Potential in gleicher Weise nutzbar. Der Business Case muss so formuliert sein und verfügbar gemacht werden, dass alle Projektmitarbeiter auf allen Ebenen jederzeit in der Lage sind, tatsächlich im Sinne des Kunden zu entscheiden. Denn dadurch kann man die zu erreichende Qualität des Produktes bei identischer Investition von Zeit und Geld maximieren.

2 Wann sind Projekte erfolgreich?

Vorgehensmodelle sollen die Durchführung von Projekten erleichtern und den Erfolg wahrscheinlicher machen. Was ist nun aus Sicht eines Sponsors oder Investors erfolgreich? Die klassische Definition für ein erfolgreiches Projekt ist die Lieferung des zugesagten Funktionsumfangs zu den vereinbarten Kosten pünktlich zu einem vorher festgelegten Termin. Ein Beispiel: Motorola hat in den 90ern das Satelliten-Kommunikationssystem Iridium entwickelt, das mit genau den bestellten Funktionen zum vereinbarten Termin unter Einhaltung des Budgets geliefert wurde, sogar in der erwarteten Qualität. Das System kam auf den Markt, wurde aber nicht benutzt, Motorola hat damit kein Geld verdient.[2] Eine andere Art, Erfolg zu messen, ist der ROI: bringt das Projektergebnis die erwartete Wertschöpfung? Die Produktion des Films Titanic war in jeder erdenklichen „klassischen" Projektmanagement-Betrachtung eine pure Katastrophe. Es gab kaum einen Parameter, ob Aufwand, Kosten, Ressourcen, oder

[2] Detaillierte Informationen zum Vergleich der beiden Projekte finden sich unter http://www.insightpo.com/en/resources-articles/project-management-resources/project-success-vs-project-management-success.aspx

Termine, den dieses Vorhaben nicht massiv gerissen hätte. Möglicherweise die Qualität, aber dies ist nicht der Platz für ästhetische Überlegungen. Der Film war einer der erfolgreichsten aller Zeiten, ist dann nicht auch das Projekt als Erfolg zu betrachten?

Vorgehensmodelle fallen (zumindest im Hinblick auf die Thematik dieses Artikels) in zwei Kategorien: solche, die den Aspekt der Investition und Gewinnerwartung eines Investors oder Sponsors durch das Erstellen und Pflegen eines transparenten Business Case ernst nehmen, und solche, die andere, eher technische Aspekte des Projekterfolgs in den Vordergrund stellen, wie z.B. das Festlegen von Anforderungen oder die Nachvollziehbarkeit von Tests.

3 Vorgehensmodelle

Man könnte ein Vorgehensmodell als Definition und Festlegung eines Produktentstehungsprozesses definieren und sich auf den Standpunkt stellen, dass ja während der Entstehung eines Produktes sowieso lediglich Kosten entstehen, und daher eine Nutzenbetrachtung und Wertschöpfung für ein Vorgehensmodell out-of-scope sei. Das ist mir zu einfach. Betrachten wir ein aktuelles Beispiel aus der Mobilfunk-Industrie. Vor einigen Jahren hat eine kalifornische Firma namens Apple ein Telefon auf den Markt gebracht, dass viele Funktionen vermissen ließ, die für Mobiltelefone dieser Preisklasse üblich sind. In einigen Wochen kommt das iPhone 4 auf den Markt, und dies ist das erste Modell dieses Namens, das alle wesentlichen Funktionen, die man in einem Gerät dieses Typs erwartet, enthält: Multitasking, UMTS, Videokamera und so weiter.

Offensichtlich hat man bei Apple also nicht einfach Geld bereitgestellt, Anforderungen aufgeschrieben und gewartet, bis diese alle realisiert und getestet waren. Stattdessen hat man sorgfältig an der Kommunikation eines Business Case gearbeitet, der allen Mitarbeitern der Entwicklung völlig klar gemacht hat, dass es durchaus zulässig ist, Anforderungen nicht zu implementieren, wenn im Gegenzug bei jedem Detail dieses Geräts sichergestellt ist, dass seine Funktion und Bedienung einfach, intuitiv, logisch, verständlich und vor allem sehr elegant und stylisch sind. Jeder, der eines dieser Geräte benutzt, kann bestätigen, dass die positive User Experience für die (wenigen) vorhandenen Features deutlich den „Schmerz“ überwiegt, auf andere Features verzichten zu müssen.

Der bei Apple verwendete Entwicklungsprozess, das Vorgehensmodell, ist also wohl sehr detailliert auf die Geschäftsziele abgestimmt. Wäre die Spezifikation ohne einen Business Case entstanden, der eine detaillierte Trennung erforderlicher von nicht erforderlichen Features zur Erzielung einer früheren Markteinführung bewirkte, wäre sie vermutlich wesentlich umfangreicher ausgefallen. Und Apple hätte sehr viel Geld dadurch nicht verdient, dass die erste Version des iPhone deutlich später auf den Markt gekommen wäre.

Dieses Beispiel zeigt, dass man durchaus während des Prozesses der Produktentwicklung schon sehr viel Geld mit einem Produkt verdienen kann, auch wenn es noch nicht wirklich „fertig“ oder „vollständig“ ist. Darüber hinaus zeigt das Beispiel,

dass eine Vorgehensweise im Projekt durchaus sicherstellen kann, dass die für das Business relevanten Produkteigenschaften realisiert werden – und zwar in der für den geschäftlichen Erfolg richtigen Reihenfolge.

Ich möchte den Begriff Vorgehensmodell daher ganzheitlicher, allgemeiner und vor allem erfolgsorientierter verstanden wissen. Schließlich soll ja die Verwendung von Vorgehensmodellen die Erfolgswahrscheinlichkeit erhöhen. Sofern es um Projekte geht, die in einem Business-Kontext stattfinden – für die also jemand Geld bezahlt und im Gegenzug eine Wertschöpfung dafür erwartet – muss dieser Business-Kontext in die Vorgehensweise des Projektes einbezogen sein.

Ein Vorgehensmodell ist ein Versuch, einen Informationsfluss zu standardisieren. Projekte erzeugen neues Wissen auf zwei Ebenen, Wissen über den Prozess (die Zusammenarbeit von Menschen in einem einmaligen Unternehmen) sowie Wissen über das Produkt (das Ergebnis eben dieses einmaligen Unternehmens). Dieses Wissen ist das Ergebnis eines Austausches von Informationen zwischen den beteiligten Menschen. Vorgehensmodelle definieren Kommunikationswege und -formen für den Transport und die Speicherung dieser Informationen. Weiterhin sind die gewonnenen Erkenntnisse Grundlage für Entscheidungen, die im Projekt getroffen werden müssen, und das auf vielen Ebenen und bei allen Beteiligten: Von den Management-Entscheidungen bis hin zu den dauernden kleinen Entscheidungen aller Projektmitarbeiter bei ihrer täglichen Arbeit.

Man kann diesen Informationsfluss als Kreislauf visualisieren:



Abbildung 1 Kreislauf des Wissens im Projekt

Dies ist in der Praxis kein so regelmäßiger Kreislauf, wie es hier vereinfacht dargestellt ist. Die Tätigkeiten überlappen sich und geschehen – häufig ohne dass es den Handelnden konkret bewusst ist – auf allen Ebenen. Damit dieser Kreislauf auf eine Wertschöpfung hin konvergiert, brauchen alle Beteiligten den Business Case als Richtschnur für ihre Entscheidungen. Ansonsten werden die Möglichkeiten des Einzelnen, zur Qualität des Ergebnisses beizutragen, werden nur unzureichend genutzt.

3.1 V-Modell XT

Das V-Modell XT als vorgeschriebenes Standard-Vorgehensmodell für Bundesbehörden definiert keinen Business Case und auch kein vergleichbares Konzept. Es gibt die IT-WiBe, die Wirtschaftlichkeitsbetrachtung, die z.B. für die Genehmigung des Projektes in den Projektvorschlag einfließen sollte und auch an anderen Stellen des Projektes eingesetzt werden sollte. Meine Erfahrung in vielen Behörden zeigt jedoch, dass diese Verwendung in der Praxis selten überhaupt stattfindet und eher eine Alibi-Funktion hat. Auch wenn das Modell andere Möglichkeiten offen lässt, wird das V-Modell in der Praxis als Wasserfall gelebt. Das ist auch die Interpretation, die durch das Modell selbst nahegelegt und durch die gängige Ausschreibungspraxis meist erzwungen wird. Eine detaillierte Nutzenüberlegung für das zu erstellende Produkt findet in der Regel nicht statt.

Das V-Modell folgt in seinen Managementverfahren sowohl in der Planung als auch in Realisierung und Qualitätssicherung einen klaren *conformance to plan* Ansatz. *Fitness for purpose* ist außerhalb des V-Modell-Weltbildes. Das führt m.E. zu zwei wesentlichen Problemen:

1. hilft das V-Modell nicht bei der Planung des „richtigen“ Produktes und
2. unterstützt es die Projektmitarbeiter nicht bei ihren Entscheidungen.

Daher wird aus wirtschaftlicher Sicht die Erfolgswahrscheinlichkeit eines Projektes durch das V-Modell XT nicht nachhaltig genug unterstützt.

3.2 PRINCE2

PRINCE2 ist ein Business Case Modell. Alle Praktiken und Prozesse in PRINCE2 dienen neben der Produkterstellung zunächst und zumeist dem Zweck, den Business Case transparenter zu machen. Nach der Genehmigung des Projektes wird der Business Case in jeder Management-Phase überarbeitet und aktualisiert, während die immer detaillierter werdende Produktstrukturplanung anhand dieses Business Case die Vision des Auftraggebers Schritt für Schritt in einen Plan umsetzt. In jeder Management-Phase wird dabei ein Regelkreis durchlaufen, der dem oben skizzierten Kreislauf des Wissens recht nahekommt:

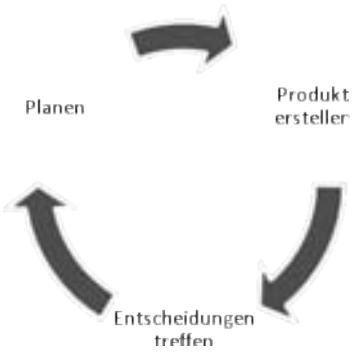


Abbildung 2 Regelkreis einer Managementphase in PRINCE2

Das Business Case Template von PRINCE2 schlägt vor, dabei regelmäßig folgende Fragen zu beantworten: Sind die Erwartungen an die Gewinne gerechtfertigt? Sind Projektplan und Business Case aufeinander abgestimmt? Sind die Gründe für das Projekt mit den Geschäftszielen konsistent?

Dabei regelt PRINCE2 im Gegensatz zum V-Modell überhaupt nicht die Arbeit der Spezialisten. Deren Arbeit, Informationsgewinnung und Entscheidungsfindung wird also ausschließlich über den Business Case und die darauf aufbauende inkrementelle Produktstrukturplanung geregelt. Dass hier ein Feedback-Zyklus eingesetzt wird, um einerseits wirtschaftliche Erwägungen in die Planung einfließen zu lassen, und andererseits möglichst inkrementell auch bereits wirtschaftlichen Nutzen zu erzeugen, lässt den Einsatz von PRINCE2 unter wirtschaftlichen Erwägungen besonders geeignet erscheinen.

5 Fazit

Projekte brauchen einen Business Case. Er führt Unternehmensleitung, Projektmanagement und die Teammitglieder bei den täglichen Entscheidungen und bildet die transparente Grundlage für den Projekterfolg im Detail, da er die gemeinsamen kreativen Anstrengungen der am Projekt beteiligten Menschen auf ein gemeinsames Ziel hin synchronisiert.

PRINCE2 institutionalisiert diesen Einfluss des Business Case auf die Entwicklung des Projektergebnisses. PRINCE2 verwenden Feedback-Schleifen und kontinuierliches Lernen als Mittel zur wertschöpfungsgesteuerten Optimierung des Projektergebnisses. Davon könnte das V-Modell XT noch profitieren.

Einführung eines agilen Vorgehensmodells auf der Basis von Werkzeugen und eines Leitbildes

Andreas Romahn

Bereich System- und Integrationslösungen
InMediasP GmbH
Neuendorfstaße 18a
16761 Hennigsdorf
romahn@inmediasp.de

Abstract: Die Einführung oder Veränderung eines Vorgehensmodells hängt im besonderen Maße von der Akzeptanz der Betroffenen ab. Gerade agile Methoden verzichten auf Formalismen und setzen voraus, dass selbstverständlich Verantwortung durch das Entwicklerteam übernommen wird. In diesem Praxisbericht wird die Einführung eines Vorgehensmodells auf der Basis von XP beschrieben und wie den Herausforderungen eines dynamischen Umfeldes begegnet und das Ziel hoher Akzeptanz und Mitarbeit an der Weiterentwicklung des Modells durch alle Betroffenen erreicht wurde.

1 Einleitung

Die Firma InMediasP GmbH bietet Softwareentwicklungsleistungen und Beratungsleistungen für Firmen in der Automobil-, Schienenfahrzeug-, Luftfahrt-, Maschinenbau-, Hausgeräte- und Zulieferindustrie sowie dem Anlagenbau. Der Schwerpunkt liegt dabei auf den Produktentstehungsprozessen.

Aufgrund von Firmenwachstum und Reorganisation der Firma wurde die Einführung eines firmenweiten Vorgehensmodells für die Softwareentwicklung erforderlich. Bis zu diesem Zeitpunkt gab es kein firmenweites Modell, so dass jedes Projektteam eine eigene Ausprägung von XP vornahm. Die durch die Projektteams vorgenommenen Anpassungen führten zu individuellen Prozessen, die nicht mehr vergleichbar waren. Für die Einführung des firmenweiten Vorgehensmodells wurden zwei Hauptziele definiert. Das erste war, dass den Projektmanagern und der Geschäftsführung ein einheitliches Modell für die Planung und Berichtserstellung geboten werden sollte. Das zweite war, dass die Einführung und Weiterentwicklung des Modells in der Mitarbeiterschaft verankert werden sollte. Dies erforderte zum einen die Benennung eines zentralen Verantwortlichen, der für die Weiterentwicklung und Mitarbeiterschulungen zum Modell zuständig war, und zum anderen aber auch Maßnahmen, durch die sich alle Mitarbeiter dazu aufgefordert fühlen sollten an der Weiterentwicklung teilzunehmen und Ideen einzubringen.

2 Ausgangssituation

2.1 Anforderungen

Um die Anforderungen an ein neues Vorgehensmodell zu ermitteln, wurden alle Projekte, die die Firma aktuell bearbeitet, analysiert. Dazu gehörten Industrieprojekte, Forschungsprojekte sowie interne Projekte, die Software zur Unterstützung interner Prozesse entwickeln. Dabei wurden die Projekte auf die Aspekte Laufzeit, Vertragsart (Dienstleistung/Werksvertrag), Teamgröße sowie Aufgaben und Verantwortlichkeiten, in Rollen zusammengefasst, hin betrachtet. Es hat sich herausgestellt, dass die Verteilung der Rollen zwischen Kunden, Projektpartnern und der InMediasP in den verschiedenen betrachteten Projekten sehr unterschiedlich war. Während die Forschungsprojekte und internen Projekte alle eine sehr ähnliche und stabile, etablierte Rollenverteilung hatten, stellten die Industrieprojekte für die spätere Konzeption des Vorgehensmodells eine Herausforderung dar. Die Ursache ist, dass die Industrieprojekte oft durch Fachabteilungen, nicht IT, beauftragt werden. Von Vorteil ist, dass Fachabteilungen für sich selbst in einem agilen Vorgehensmodell einen Vorteil sehen, was andernfalls erheblichen Überzeugungsaufwand kostet. Daraus ergeben sich allerdings zwei große Herausforderungen. Die erste ist, dass selbst bei wenigen Rollen im Modell für die Verantwortlichkeiten und Aufgaben, die durch den Kunden wahrgenommen werden sollen, erst mit der Zeit Verständnis aufkommt. Die zweite ist, dass agil sehr oft so interpretiert wird, dass zu jeder Zeit neue Anforderungen oder Anforderungsänderungen, selbst in die laufende Iteration, gegeben werden können.

2.2 Die Firma im Wandel

Die interne Organisation der Firma war so, dass jeder der drei Geschäftsführer (GF) einen Teil der Mitarbeiter und Projekte führte und überwachte. Durch kontinuierliches Wachstum sammelten sich allerdings zu viele Aufgaben bei den jeweiligen GF, so dass eine Neuaufteilung der Aufgaben und Verantwortlichkeiten erforderlich wurde. Ergebnis war eine Matrixorganisation bestehend aus den Bereichen Softwareentwicklung und Beratung und einem Bereich Geschäftsentwicklung, unter den alle Aufgaben fielen, die die Firmenentwicklung betrafen. Jeder GF führte einen Bereich. Die Bereiche wurden in Teams aufgeteilt, von denen jedes einen inhaltlichen Schwerpunkt bekam.

Die Aufteilung der Firma in Teams hatte allerdings tiefergehende Konsequenzen für die Praxis des angewendeten Vorgehensmodells. Während schon vor der Einführung von Teams jedes Projekt für sich das Vorgehensmodell unterschiedlich anwendete, stellte die Trennung in Teams psychologisch eine noch größere Hürde für den Austausch und die Harmonisierung der Anwendung dar. Die Identifikation eines Entwicklers mit seinem Team, die zusätzlich zur Identifikation mit dem Projekt hinzukam, führte eher zu einer Trennung zwischen den Projektgruppen, so dass mit der Zeit im Projektalltag festgestellt werden konnte, dass es kein gemeinsames Verständnis zum Vorgehensmodell mehr gab.

3 Leitbild und Vorgehensmodell

3.1 Das Leitbild

Als Basis sowohl für die Konzeption als auch die Einführung und Weiterentwicklung des Vorgehensmodells wurde ein Leitbild entwickelt, an dem alle Aktivitäten ausgerichtet wurden.

Im Zentrum steht der Mensch. Er ist das verbindende Element zwischen dem Entwicklungsprozess, den verwendeten Werkzeugen und den Projekten. Er hält den Prozess am Leben, nutzt die Werkzeuge und bearbeitet die Projekte. Das größte Gewicht in diesem Bild hat der Mensch und soll so verstanden werden wie im ersten Wert des agilen Manifestes „Individuen und Interaktionen haben Vorrang vor Prozessen und Werkzeugen“ [Be04]. Unsere wichtigsten Grundsätze in der Konzeptionierungsphase des Vorgehensmodells und dessen Veränderungsprozesses waren: alle vier Komponenten müssen einander zuarbeiten, um sich gegenseitig zu verstärken, der Mensch ist Motor und Zentrum, er hält alles in Bewegung und an ihm soll alles ausgerichtet sein, der Mensch soll die Möglichkeit zu Änderungen haben.

3.2 Die Entwicklung des Vorgehensmodells

Als Basis für das neue Vorgehensmodell wurde das bereits genutzte XP genommen, da hiermit alle Mitarbeiter Erfahrung hatten und es für die Umsetzung der Anforderungen an ein Vorgehensmodell am besten geeignet schien. Ansätze wie SCRUM oder XP ähneln sich in ihren Grundannahmen und Ausprägungen. Aufgrund der komplexen Rollenverteilung in den Industrieprojekten kam ein Modell mit starren Rollenverteilungen, wie SCRUM, nicht in Frage, da die Rollen hätten aufgebrochen werden müssen, wie bspw. in [Ep10] beschrieben.

Das konzipierte Modell ist eine Sammlung von Rollen, Methoden und Aktivitäten, wie sie in den gängigen agilen Vorgehensmodellen zu finden sind. Am schwierigsten war es, die wenigen in den agilen Methoden definierten Verantwortlichkeiten so zu definieren, dass eine sinnvolle Abbildung auf die Projektteams in Industrieprojekten möglich war. Die Nutzung von Methoden und Aktivitäten aus XP geschah im Sinne eines Baukastenprinzips, indem wir das verwendeten, was von den Entwicklern für sinnvoll erachtet wurde. Über die Konstruktion eines eigenen Vorgehensmodells aus den Methoden und Praktiken bestehender agiler Ansätze gibt es mittlerweile Literatur, die die Idee eines Baukastens in Form eines Kataloges vorstellt [St10]. Ein solcher Baukasten ist auch über agile Vorgehensmodelle hinaus interessant, da laut [GW10] gerade die pragmatische Integration von agilen und nicht-agilen Methoden erfolgreich ist, so dass nicht ein populäres Vorgehensmodell angewendet werden muss, sondern Methoden und Praktiken, die zur Arbeits- und Ergebnisverbesserung führen.

4 Werkzeuge, ihre Einführung und der Weiterentwicklungsprozess

4.1 Die Rolle der Werkzeuge

Parallel zur Konzeptionsphase des Vorgehensmodells hatten Werkzeugevaluationen stattgefunden, die sowohl den Entwicklern als auch den Projektmanagern Unterstützung bieten sollten. Die Erfahrung mit dem bisherigen Modell hatte gezeigt, dass eine Verankerung des Modells nur funktioniert, wenn es durch Werkzeuge unterstützt wird. Wichtig war, dass es einen Konsens zu den Werkzeugen geben sollte, d.h. alle nutzen für gleiche Probleme und Methoden die gleichen Werkzeuge. Die vornehmlichen Ziele sind, dass die Werkzeuglandschaft nicht allzu heterogen wird und, dass Wissen zu Werkzeugen unter den Anwendern geteilt wird und so zu einer effizienteren Nutzung führen soll. Eine wichtige Annahme für die Auswahl der richtigen Werkzeuge war, dass hiervon der Erfolg des Vorgehensmodells abhängt. Die Werkzeuge sollen den spürbaren Rahmen bilden, in dem sich die Entwicklung auf der Basis des Vorgehensmodells abspielt. Ohne einen solchen Rahmen droht wieder die projektspezifische Veränderung des Vorgehensmodells, für das es nicht vorgesehen ist.

Die Auswahl der Werkzeuge konzentrierte sich zum einen auf die Unterstützung des Projektmanagements und zum anderen auf Codeanalysewerkzeuge für Entwickler. Für das Projektmanagement wurden Werkzeuge untersucht, die agile Methoden unterstützen und Funktionalitäten für User Storys, Release- und Iterationsplanung, Testmanagement, sowie Ressourcenmanagement boten. Hier wurden Werkzeuge bevorzugt, die eher eine einfache Benutzeroberfläche hatten und eine überschaubare und schnell zu erlernende Menge an Funktionen haben. Die Codeanalysewerkzeuge wurden daraufhin ausgewählt, dass sie sowohl auf einem lokalen Entwicklerrechner als auch auf einem zentralen Build-Server genutzt werden konnten. Als erstes wurden Codeanalysewerkzeuge angeschafft, die der Anwendung von Software-Metriken und der Codeabdeckungsanalyse dienten.

4.2 Die Einführungsphase

Bereits vor der Einführungsphase wurde ein zentraler Build-Server in Betrieb genommen, in den alle Werkzeuge, sowohl die seit Längerem benutzten, als auch die neu angeschafften, integriert wurden. Dies löste das Problem, dass es nicht für alle Werkzeuge individuelle Entwicklerlizenzen gab und erhöhte auch die Transparenz in der gesamten Firma für alle Entwickler dafür, welche Werkzeuge zur Verfügung stehen und wie man weitere integrieren kann.

Die Einführungsphase für das Vorgehensmodell und die Schulung der Mitarbeiter für die unterstützenden Werkzeuge lief über ein halbes Jahr. Dabei wurden zuerst die Werkzeuge eingeführt, um das bereits vorhandene Modell durch intuitives Nutzen der Werkzeugfunktionen zu unterstützen. Erst danach wurde die Verwendung der Werkzeuge durch Schulung der Entwickler und Konfiguration der Werkzeuglandschaft iterativ so verändert, dass es dem neuen Vorgehensmodell entsprach.

Das anfänglich erwartete Problem, dass die Einführung neuer Werkzeuge bei gleichzeitig erlaubtem alten Modell dazu führt, dass das neue nicht mehr angenommen wird, hat sich nicht bestätigt. Die durch die Werkzeugverwendung spürbar verbesserten Arbeitsabläufe und Arbeitsergebnisse führten dazu, dass der Wunsch nach einer effizienteren Werkzeugnutzung die Einführung des neuen Vorgehensmodells motivierte.

4.2 Der Weiterentwicklungsprozess

Die Anwendung und Weiterentwicklung des Vorgehensmodells ist vornehmlich verbunden mit der Akzeptanz durch die Entwickler, der Methodenkompetenz und der Offenheit des Modells und der Beteiligten für Veränderungen. Um die Offenheit des Modells zu veranschaulichen, wurde das Vorgehensmodell auf ein großes Poster gedruckt und an zentraler Stelle, der Kaffeeküche, aufgehängt. Ausnahmslos jeder kann dort Fragen, Kommentare und Vorschläge entweder als „Post Its“ anbringen oder direkt auf das Poster malen. Damit sind alle aufgerufen das Modell immer wieder anzusehen, zu diskutieren und Änderungen vorzuschlagen. Das Poster wurde auch von Entwicklerteams mitgenommen, die nicht im Hauptstandort, sondern bei Kunden arbeiten, so dass auch hier keine Abgrenzung durch räumliche Entfernung empfunden werden sollte. Regelmäßig werden die Poster eingesammelt, die Kommentare, Vorschläge und Fragen erfasst, und in Entwicklergruppentreffen verarbeitet. Ergebnis ist in der Regel ein neues Poster, das wieder an allen Standorten verteilt wird.

Gemäß dem Leitbild spielt auch im Weiterentwicklungsprozess der Mensch eine zentrale Rolle, in diesem Fall alle Entwickler. Es hat sich herausgestellt, dass die Transparenz und Offenheit des Vorgehensmodells hin zu den Anwendern sehr positiv für die Akzeptanz waren, da sich jeder an der Weiterentwicklung beteiligen kann und gemeinsam Änderungen diskutiert werden. Darüber hinaus bietet ein öffentlich sichtbares Modell in Form eines Posters die Identifikation mit einem Thema über die Grenzen eines Projektes und Teams hinaus, was früher ein Problem war und dazu geführt hatte, dass es praktisch kein gemeinsames Modell mehr gab (Kapitel 2.2). In den ersten Monaten war der überwiegende Teil der auf dem Poster zu findenden Anmerkungen Fragen, die auf einen Klärungsbedarf hindeutete. Als Maßnahme fanden zusätzliche zu den bereits geplanten Schulungen statt, die speziell auf die Anmerkungen hin gestaltet waren. Vorschläge zur Veränderungen des Modell sind so gut wie nicht vorhanden, was vermutlich daran liegt, dass sich viele noch nicht firm darin fühlen.

5 Fazit

Der Erfolg der Einführung des Vorgehensmodells bzw. die iterative Veränderung des bestehenden basiert darauf, dass unterstützende und an das Modell angepasste sowie konfigurierte Werkzeuge die Arbeitsabläufe für alle positiv spürbar unterstützen. Das Mehr an neuen Aktivitäten und Werkzeugen wurde nicht als ein Mehr an Arbeit empfunden, da deren Ergebnisse der Transparenz und Entscheidungsunterstützung dienten. Die von Anfang an kommunizierte Offenheit für Veränderungen, was durch ein Poster, das jeder verändern darf, sichtbar ist, ist Motivation für alle, sich mit dem Modell zu beschäftigen und Verbesserungen vorzuschlagen.

Bei der Einführung des Vorgehensmodells hat sich die vorangegangene Einführung eines Projektmanagementwerkzeuges mit überschaubarer Funktionalität als besonders erfolgreich erwiesen. Dieses Werkzeug wird sowohl von den Entwicklern, dem Projektmanager als auch der Geschäftsführung genutzt. Es bietet für die agile Entwicklung zentrale Funktionen für User Storys, Release- und Iterationsplanung, Ressourcenmanagement, u.v.m. Die wenigen aber wichtigen Funktionen wurden in kurzer Zeit von allen gelernt und angewendet. Daraus erwuchs der Wunsch diese auch basierend auf einem Modell zu verwenden, was die Einführung des neuen Modells erleichterte.

Literaturverzeichnis

[Be04] Beck, K.: Extreme Programming - Das Manifest, Addison-Wesley, 2004

[Ep10] Epping, T.: Projektmanagement durch Scrum-Proxies, 17. Workshop der Fachgruppe WI-VM der Gesellschaft für Informatik e.V., Integration von Vorgehensmodellen und Projektmanagement, Suttgart, 2010

[GW10] Grant, T.; West D.: Mainstream Adoption Has Changed Agility, Forrester Research, Inc., 2010

[St10] Steyer, M.: Agile Muster und Methoden - Agile Softwareentwicklung maßgeschneidert, Entwickler.Press, 2010

Applying Scrum and XP in an Enterprise Context

Dr. Uwe Bombosch, Dr. Stefan Bente, Dr. Margret Hesselmann, Gero Seifert

Tata Consultancy Services Deutschland GmbH
Messe Turm
60308 Frankfurt am Main
uwe.bombosch@tcs.com
stefan.bente@tcs.com
margret.hesselmann@tcs.com
gero.seifert@tcs.com

Abstract: This article describes the challenges in applying the Scrum and eXtreme Programming (XP) software development practices, summarising several years of industry experience. The crucial questions are: How applicable are these practices in an enterprise context, and do they show their well-known advantages under today's characteristic business conditions? The assessment takes the project management, enterprise, human, software engineering, and business viewpoint in order to shed light on applicability gaps, and best practices to fill them.

1 Introduction

In the past years, we consistently applied Scrum as defined in [BS02] and interspersed this approach with elements from XP, namely the requirement capturing by user stories, and practices like pair programming, continuous integration (CI), and test-driven development. We've seen many successful Scrum/XP projects, and do neither question its general applicability, nor hold that its value proposition is invalid under certain conditions. Nevertheless, applying Scrum/XP in an enterprise context poses a number of challenges that are not covered by standard Scrum/XP, and requires complementary approaches.

2 Challenges in the Project Management Dimension

2.1 Large and Distributed Teams

Large-scale projects are generally more demanding than smaller ones, but the following factors are particularly critical for bigger, distributed Scrum/XP projects.

Communication: Minimizing the number of involved sites, collocation per site and frequent mutual visits facilitate the ideal of face-to-face communication.

Visibility of progress: In addition to management-centered instruments like task tracking or status meetings, CI enables stakeholders to inspect the actual product and quite reliably determine the progress of product creation.

2.2 Tooling in Large Enterprise Projects

Adequate tooling and automation help keeping the aforementioned factors at bay.

For *project management* we have successfully used *Atlassian Jira with GreenHopper plug-in*. They support all essential Scrum/XP practices in distributed mode, are stable, very intuitive to use and come with integrated issue tracking.

For *Continuous Integration* there are mature open source tools. Yet, putting these tools together in order to run fully automated end-to-end builds, from SW version control to automated smoke tests in the reference environment, tends to be *far* underestimated.

For CI it is essential that developers integrate their work daily without breaking the end-to-end build. The CI set-up has to support this, e.g. by staging and by fast feedback (end-to-end build < 10 minutes). A "build manager" should instantly react to broken builds.

3 Challenges in the Human Dimension

3.1 Changed Roles, Skills and Mindsets

Scrum/XP enforces working in a cross-functional team. That is why job-descriptions, roles, privileges, and compensation models must be changed. There is, for example, no "Tester" role in Scrum/XP.

Soft skills are gaining in importance. Non-technical stakeholders enter the every-day-life of programmers, and take part in the development lifecycle. The individual's work and its results are frankly exposed to the whole team. Perfectionists have to learn to provide simple solutions, and to enhance the design on stakeholder requests only.

3.2 Cultural Aspects

Cultural factors may impact the introduction of Scrum/XP, which requires a high degree of autonomous, self-responsible thinking and acting. This tends to conflict with hierarchical team structures and autocratic leadership common to Asia. On the other hand, the strong preference for teamwork in Asian cultures is beneficial to Scrum/XP.

4 Challenges in the Software Engineering Dimension

While Scrum/XP on the one hand washes away some idle production of up-front specifications that is inherent to the waterfall approach, the principles and practices for doing architecture, design, programming, and testing must not be thrown overboard under the "we're agile now" pretence.

4.1 Architecture and Design

4.1.1 Large Systems

Architecture work on a large system cannot be done on the Sprint planning day or ad-hoc while doing pair-programming. Large systems or even enterprise application landscapes cannot be developed without applying a suitable architecture framework.

In Scrum/XP, we plan and execute architecture work as ordinary Sprint tasks. The modeling of business processes, for instance, is formulated as a user story, broken down into tasks, and then phased into a Sprint.

The challenge here is to find the appropriate heartbeat for the architecture work. Overspecification must be avoided, but there always has to be sufficient guidance to support the project execution. On the planning day, the architecture must be clear enough to derive concrete implementation tasks.

One of the most wide-spread and complete frameworks is the TOGAF™ Architecture Development Method (ADM) owned by the Open Group consortium [OG09]. Its iterative, incremental nature, and the openness towards tailoring and stripping-down makes ADM a suitable choice for large-scale Scrum/XP projects.

The need for an intentional architecture is a natural limitation of the stakeholder's freedom to choose user stories for implementation in the next Sprint. Picking a user story that is architecture-wise not covered triggers the next cycle of architecture work, and therefore comes with a high price tag.

4.1.2 User-Interface Design

Transforming a user story or a use case into a UI specification requires a dedicated design step, involving intense discussions with the stakeholders. The use of a UI mocking tool like Balsamiq facilitates this process, but a plain paper prototype also works.

A visual representation enables stakeholders to envisage the end-product, and uncovers omissions and ambiguities in the verbal requirement description.

Such UI specifications are dedicated artefacts that should be created during the Sprint. A possible process for doing it is to shape the specification and mockup in Sprint N, review it by a usability test and a customer demo, and phase it into Sprint N + 1 for implementation.

4.2 Testing

Agile has a strong affinity to test automation. Executing unit and end-to-end tests during the build is an integral part of the CI concept. If the quality assurance of an enterprise is not adapted to automated test cases, agile concepts are harder to put into practice.

In some areas of testing, e.g. exhaustive performance tests, continuous testing is not possible in an economic fashion. Therefore, certain specific test phases have to be taken into account during Sprint planning, also with a respect to. the availability of test environments.

5 Challenges in the Enterprise Dimension

5.1 Process Incompatibilities

There are areas where incompatibilities between standard enterprise processes and a Scrum/XP approach can be expected. In some cases, the friction is due to the resistance of a waterfall-centric organization against Scrum/XP methods. In others, Scrum/XP methods just uncover hidden omissions and problems of an organization's processes.

5.1.1 Smuggling Scrum/XP Elements into a Waterfall Process

A pragmatic solution to overcome incompatibilities and the organizational resistance is to stick to a traditional waterfall process, but sneak in some Scrum/XP. Two examples for such elements, that have proved successful in practice, are:

- Dividing the long releases into informal Sprints, with a regular demo session, called a "stakeholder review" if Scrum/XP terminology is to be avoided.
- CI should be implemented on a moderate-effort-level, as a tool for ensuring quality.

5.1.2 Budget Planning

The enterprises budget planning typically has a time horizon of about a year. Promoters of a project candidate must in advance put a quite reliable price tag to it and convince the sponsors about the business value.

Proposals saying that „we'll start off agile with N developers, and let's see where we end" are rejected by the planning board. Scrum/XP projects need a roadmap in advance that at least on the granularity of Epics stipulates the target of the endeavour. Likewise, efforts must be estimated and approved.

As with the need for an intentional architecture that we discussed in Section 4.1.1, this limits the freedom to pick, change, or add user stories for implementation, too.

5.2 Specifics of Scrum/XP Offshoring

Running Scrum/XP projects with a contribution from offshore teams poses special challenges. The predominant benefit and the major risk stem from the business model of offshore companies of minimizing cost per head, which implicitly leads to temporarily minimizing productivity per head.

Lower productivity per head implies a larger project team, which in turn leads to a more difficult communication. Moreover, the general communication issues for large-scale Scrum/XP projects remain (cf. previous Section), and the solution we propose further increases cost.

The importance of customer involvement can hardly be overemphasized for Scrum/XP projects. Scrum/XP suppliers must be present onsite with a significant number of project members for a substantial amount of time. This is a major cost driver, which again conflicts with the cost-saving goal.

6 Challenges in the Business Dimension

6.1 Project Acquisition Phase

The vast majority of requests for proposals (RFP) favour or even enforce a waterfall methodology. Yet, when the authors tried to acquire such projects and proposed Scrum/XP, many purchasers valued its benefits of flexibility and transparency. However, if substantial effort had gone into a waterfall-oriented requirements specification already most purchasers were not willing to directly invest more into discussing and prioritizing requirements.

The actors on the ground typically acknowledge that such a "fire-and-forget" project ramp-up will not work well. They experienced that the collaboration with the supplier on user stories and an architecture vision shows better results in the end. But applying the Scrum/XP paradigm of "collaboration over contract negotiation" before the first contract has been signed is rejected by most purchasers as it binds them to the supplier too early.

A possible mitigation of this conflict is "smuggling in" Scrum/XP elements; cf. section 5.1.1.

6.2 Agile Contract Models

The agile conception of a project scope gives purchasers a high flexibility in changing requirements or re-prioritizing them. But under business conditions, the service provider and the purchaser have to agree on a contract nailing down the pricing conditions.

The challenge is to choose a contract model that does not undermine the agile flexibility. "Time and Material" (T&M) is such a model, but unpopular with purchasers as it leaves most risk with them and does not force providers to work efficiently. An example of a better model is "Fixed Price per Iteration with Story Point Guarantee", which provides flexibility but also enforces that all deliveries have a certain comparable complexity.

A lot of contract models have been invented in order to mitigate the tension between agile flexibility and calculable prices [St09]. However, none of these models have gained market-acceptance yet.

7 Summary

As we've seen, the introduction of Scrum/XP to an enterprise context shows challenges in several aspects. The hot spots are probably in the engineering and the business dimension.

The authors have seen large projects in trouble because they neglected the need for an architecture framework, and silently assumed that the target architecture would show up by magic while "being agile".

However, the most demanding friction, at least from the perspective of a service delivery company, is the RFP acquisition process. Nevertheless it is observable that companies start considering this friction also as an opportunity to come to a different mode of operation, and a better collaboration with their service providers. Time will tell whether Scrum/XP must be enhanced to support fixed prices and RFPs, or vice versa.

References

[BS02] Beele, M.; Schwaber, K.: Agile Software Development with Scrum. Prentince Hall 2002.

[Co04] Cohn, M: User Stories Applied. Addision-Wesley 2004.

[HMR09] Hastreiter, G.; Mathis, C.; Roberts, S.: Leuchtfeuer Entzünden: Einführung von Enterprise Scrum bei der Allianz Deutschland AG. OBJECTSpektrum 2009; S. 54-60.

[LV10] Larman, C.; Vodde, B: Practices for Scaling Lean and Agile Development: Large, Multisite, and Offshore Product Development with Large-Scale Scrum. Addison-Wesley Longman, Amsterdam, 2010.

[OG09] The Open Group: TOGAF Version 9, http://www.opengroup.org/togaf/, 2009.

[Sc10] Schwaber, K.: The Home of Scrum; http://scrum.org/scrumdeveloper, 2010

[St09] Stevenson, P.: 10 Contracts for your Next Agile Software Product, April 2009, http://agilesoftwaredevelopment.com/blog/peterstev/10-agile-contracts

Workshop

Risk Management, Compliance und Governance für widerstandsfähige Informationssysteme

Dr. Michael Schermann, Prof. Dr. Helmut Krcmar

Risk Management, Compliance und Governance für widerstandsfähige Informationssysteme

Vorwort

Widerstandsfähige Informationssysteme sind in der Lage, unvorhergesehene Ereignisse abzufedern. Auf Grund der Komplexität, Dynamik und Interkonnektivität moderner Informationssysteme ist es weder möglich, alle potenziellen Risiken zu identifizieren noch erkannte Risiken in ihrem vollen Umfang zu steuern. Wettbewerbsfähige Informationssysteme müssen die oftmals hohen Anforderungen der Verfügbarkeit, Integrität und Vertraulichkeit auch nach Eintritt unerwarteter Ereignisse noch erfüllen können. CIOs sind daher gefordert, die Leistungsfähigkeit ihrer Informationssysteme trotz unerwarteter Ereignisse zu sichern.

Die aktuelle Wirtschaftslage zeigt anschaulich, dass zurzeit mit Risiken noch nicht effektiv genug umgegangen wird. Das Ziel muss es daher sein, zukünftig das Ausmaß von Schäden durch effektive Maßnahmen zur Risikoeliminierung und -reduktion zu verringern. Problematisch ist hier die ganzheitliche Integration in die vorhandene Organisationsstruktur, da isoliertes Risikomanagement keine Verknüpfungen und Auswirkungen zwischen verschiedenen Risikobereichen im Unternehmen aufzeigen kann.

Regulatorische Anforderungen verpflichten Unternehmen zur Einhaltung von Verhaltensgrundsätzen, Gesetzen und Richtlinien. So müssen einerseits Auflagen durch Vorschriften und Gesetze berücksichtigt werden, andererseits spielen freiwillige, durch Wertvorstellungen, Moral und Ethik festgelegte, interne Grundsätze eine besondere Rolle. Neben Maßnahmen zur Überwachung der Einhaltung der aufgestellten Vorgaben muss ein generelles Verständnis für Risiken unternehmensweit etabliert werden.

Im Rahmen der IT-Governance müssen Compliance-Vorgaben, Risikosteuerung und Risikoüberwachung umgesetzt und in die Unternehmensstruktur und -prozesse integriert werden. Herausforderungen an das Risikomanagement umfassen die Identifikation und Quantifizierung von Risiken für Informationssysteme. Hierfür müssen Risikoinformationen zeitnah und detailliert bereitgestellt werden, aus denen adäquate Steuerungsmaßnahmen abgeleitet werden können.

Die für diesen Workshop akzeptierten Beiträge können in zwei Kategorien unterteilt werden. Die erste Kategorie umfasst Beiträge, die eine integrative Perspektive auf das Thema GRC vermitteln. Benjamin Schwering und Alexander Pellengahr untersuchen den stark regulierten Bankensektor und beurteilen die Relevanz von Prüfungsstandards für IT-Auslagerung. Patrick Wolf und Matthias Goeken analysieren die noch unzureichende Integration des IT-Risikomanagements in ein unternehmensweites Risikomanagement. Nicolas Racz, Edgar Weippl und Andreas Seufert untersuchen den Einsatz von Risikomanagement-Frameworks auf Gemeinsamkeiten und Rationierungspotenziale.

Die zweite Kategorie umfasst die organisatorische Einbettung und Anwendung von Governance, Risk Management und Compliance. Konrad Walser untersucht die Beziehungen zwischen IT-Governance und dem Unternehmensarchitekturmanagement und fokussiert auf deren Nutzen für die Sicherstellung des Business-IT-Alignment. Manfred Pauli, Michael Schermann und Helmut Krcmar schlagen einen Ansatz für die Verzahnung von IT-Risikomanagement und Unternehmensarchitekturmanagement zur Verbesserung der Informationsgrundlage des IT-Risikomanagements vor. Silvia Knittl und Wolfgang Hommel verdeutlichen den Einsatz von Cloud-basierten Diensten im Umfeld von Hochschulen und stellen die damit verbundenen Risiken vor.

Die vorliegenden Beiträge zeigen die aktuelle Forschung in den Bereichen Risk Management, Compliance und Governance. Unser Dank gebührt den Mitgliedern des Programmkomitees, Herr Prof. Dr. Matthias Goeken, Herr Prof. Dr. Hannes Federrath, Herr Prof. Dr. Knut Hildebrand, Herr Christian Martini und Herr Robert Kamrau für die Unterstützung des Workshops. Ebenso danken wir den vielen Gutachtern, die durch kritische, aber konstruktive und umfangreiche Begutachtung der Beiträge den Autoren wertvolle Hilfestellungen bei der Weiterentwicklung ihrer Ansätze geliefert haben. Wir bedanken uns ebenso bei Frau Alexandra Gerstner, Herrn Prof. Dr. Ing. habil. Klaus-Peter Fähnrich und Herrn Prof. Dr. Bogdan Franczyk sehr herzlich für ihre Unterstützung und die Organisation der Informatik 2010.

Es bleibt zu hoffen, dass die Ideen und Konzepte zu Governance, Risk Management und Compliance für widerstandsfähige Informationssysteme von Wirtschaft und Wissenschaft aufgegriffen werden und neue Potenziale und Forschungsfelder erschlossen werden können.

Dr. Michael Schermann
Prof. Dr. Helmut Krcmar

Programmkomitee

Christian Martini, Siemens AG
Robert Kamrau, IBM Deutschland GmbH
Prof. Dr. Matthias Goeken, Frankfurt School of Finance & Management
Prof. Dr. Hannes Federrath, Lehrstuhl Wirtschaftsinformatik 4, Universität Regensburg
Prof. Dr. Knut Hildebrand, Fachbereich Wirtschaft, Hochschule Darmstadt

Die Relevanz von IT-Compliance Prüfungsstandards bei IT-Auslagerungen im deutschen Bankensektor

Benjamin Schwering

Technische Universität München
Chair for Information Systems, Prof. Dr. Krcmar
Boltzmannstr. 3
85748 Garching
benjamin.schwering@in.tum.de

Abstract: Dieser Forschungsbeitrag soll die Relevanz von Prüfungsstandards auf die Gestaltung von Informationssystemen und der IT Landschaft von IT-Dienstleistern im deutschen Bankensektor aufzeigen. Zu diesem Zweck wurde ein empirisch-qualitativer Forschungsansatz gewählt. Es wurde ein Fragebogen konstruiert und an zirka 600 deutsche Banken versendet. Aus den Ergebnissen kann geschlossen werden, dass Banken tatsächlich eine große Nachfrage nach entsprechenden Prüfungsleistungen besitzen und damit weitreichende Anforderungen an die Gestaltung der IT Landschaft ihrer IT-Dienstleister stellen. In diesem Beitrag werden damit die Anforderungen deutscher Banken dargestellt und belegt. Aufbauend auf diesen Beitrag sollte sich weitere Forschung mit den genauen Konsequenzen für IT-Dienstleister und deren IT Landschaft beschäftigen.

1 Einleitung

Im deutschen Bankenumfeld ist eine Auslagerung der IT-Landschaft nicht unüblich und hat in mehreren Beispielen zu großen Erfolgen geführt ([SS08] S. 120 f.). So hat die Deutsche Bank 2007 ihr Rechenzentrum erfolgreich an Dritte ausgelagert. Auch die Stadtsparkassen und Genossenschaften haben seit den siebziger Jahren bereits ihre EDV ausgelagert. Dass das bis heute Bestand hat, spricht ebenfalls dafür, dass die Auslagerung im deutschen Bankenbereich durchaus üblich und erfolgversprechend ist. In diesem Umfeld stellen Banken umfangreiche Anforderungen an ihre IT-Dienstleister. Diese sind insbesondere durch gesetzliche Anforderungen begründet, denen die Banken selbst unterworfen sind. Denn der Bankensektor ist wie kein anderer Wirtschaftszweig einer umfassenden staatlichen Regulierung unterworfen ([HPW07] S. 355; [Ge96])

In der Literatur lassen sich relevante Gesetze finden, die besondere Anforderungen an operationelle Risiken definieren und damit Auswirkung auf die IT im deutschen Bankensystem stellen. (Vgl. dazu im Folgenden [LS06], [GRB04] Kapitel 4.5 und [HPW07] Kapitel L2). Gesetze mit allgemeiner Bedeutung sind im Aktiengesetz (AktG), dem Handelsgesetzbuch (HGB) und den Grundsätze ordnungsmäßiger DV-gestützter Buchführungssysteme (GoBS) zu finden. In vielen Gesetzen findet sich die Forderung nach einem internen Kontrollsystem, um operationellen Risiken entgegen zu wirken. So fordern § 91 Abs. 2 AktG und § 317 Abs. 4 HGB die Einrichtung und Prüfung eines internen Kontroll- bzw. Überwachungssystems. Außerdem ist der Abschlussprüfer laut § 289 Abs. 1 HGB (für Konzerne gilt § 317 Abs. 2 HGB) dazu verpflichtet, im Lagebericht des Unternehmens, auf die Risiken der künftigen Entwicklung einzugehen.

Für Banken sind zudem das Kreditwesengesetz (KGW) und die Mindestanforderungen an das Risikomanagement (MaRisk) von besonderer Bedeutung. Insbesondere werden restriktive Anforderungen an ein internes Kontollsystem in den MaRisk (MaRisk BT1) definiert. So wird von dem internen Kontrollsystem insbesondere verlangt, dass es „Regelungen zur Aufbau- und Ablauforganisation und; Prozesse zur Identifizierung, Beurteilung, Steuerung, Überwachung sowie Kommunikation der Risiken (Risikosteuerungs- und -controllingprozesse)“ (MaRisk AT1) umfasst. Sollten Teile der Unternehmens-IT, die für das interne Kontrollsystems eines Unternehmens relevant sind, an externe IT-Dienstleister ausgelagert sein, darf das auslagernde Unternehmen unter keinen Umständen die Verantwortung für die internen Kontrollen der ausgelagerten Bereiche abgeben (MaRisk AT 9). Das Unternehmen muss entweder selber die Prozesse in dem Dienstleistungsunternehmen prüfen oder von dem Dienstleistungsunternehmen eine Prüfung der ausgelagerten Prozessteile bzw. der implementierten internen Kontrollen durch einen unabhängigen Wirtschaftsprüfer verlangen. Abbildung 1 verdeutlicht den dargestellten Zusammenhang zwischen Bank und IT-Dienstleister.

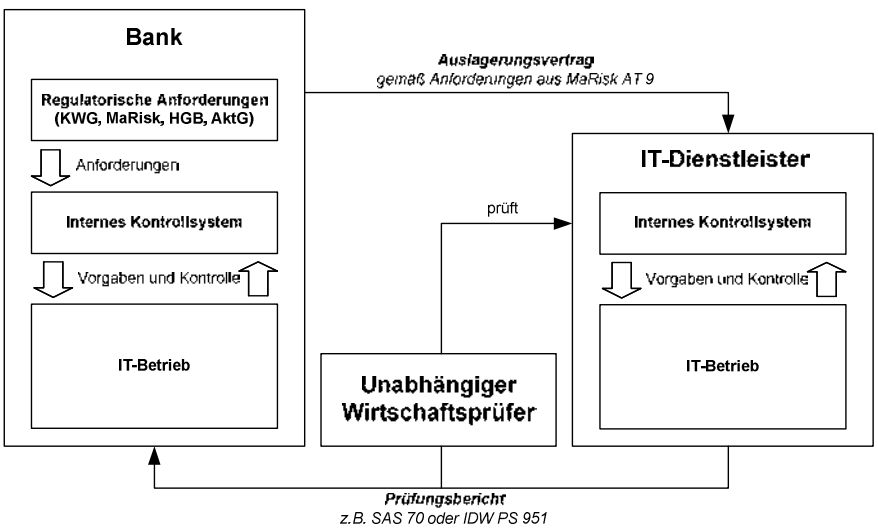


Abbildung 1 Beziehung zwischen Bank und IT-Dienstleister

Für den Zweck gibt es den internationale Prüfungsstandard SAS 70 sowie den deutschen Prüfungsstandard IDW PS 951. Beide beschreiben nicht etwa eine Checkliste mit Kontrollen vor, die der Wirtschaftsprüfer zu testen hat. Der Prüfumfang muss je nach Art der ausgelagerten Dienstleistungen für jeden Bericht definiert werden. Orientieren können sich IT-Dienstleister bei der Einführung der Kontrollen an weit verbreiteten Kontrollziel-Frameworks, wie COSO oder CobiT. Sowohl der IDW PS 951 als auch der SAS 70 definieren zwei unterschiedliche Typen von Prüfberichten, von denen der Typ B bzw. Type II deutlich restriktiver ist. Dieser bewertet im Gegensatz zum einfachen Bericht, der lediglich eine Zeitpunktbetrachtung darstellt, eine Zeitraumbetrachtung. Im Gegensatz zum internationalen Standard SAS 70, der einen hohen Freiheitsgrad bei der Gestaltung zulässt, werden in einem IDW PS 951 Bericht Hinweise auf nicht durchgeführte Kontrollen gegeben.

Das Ziel dieses Forschungsbeitrags ist es, vor diesem Hintergrund, den tatsächlichen Bedarf deutscher Banken an solchen Prüfungsleistungen zu ermitteln, um eine Aussage über deren Relevanz in diesem Bereich treffen z können.

2 Forschungsmethode

Zum beantworten der Forschungsfrage wurde ein empirisch quantitativer Ansatz gewählt. Es wurde ein Fragebogen entwickelt. Dieser Fragebogen wurde erstellt, um die oben erwähnte Fragestellung nach den Anforderungen deutscher Banken an IT-Dienstleister zu beantworten. Es sollte damit auf empirisch-quantitativer Weise belegt werden, wie relevant IT-Compliance Prüfungsstandards im deutschen Bankensektor sind und welche Prüfungsleistungen bzw. Zertifikate besonders wichtig sind. Folgende Fragestellungen sollen daher überprüft werden:

- Ist der Umgang mit operationellen Risiken für Banken besonders wichtig?
- Welche Prüfungsstandards für das interne Kontrollsystem des IT-Dienstleisters sind für deutsche Banken besonders wichtig?

Der Fragebogen ist in drei Bereiche untergliedert: Der erste Teil enthält Fragen zu dem allgemeinen Umgang mit operationellen Risiken in der Bank. Der zweite Teil beinhaltet Fragen zu den Prüfungen des internen Kontrollsystems von IT-Dienstleistern. Abschließend sollte im letzten Teil eine Selbsteinordnung der Banken in verschiedene Kategorien geschehen.

Um die Rücklaufquote der Fragebögen, die an die ca. 200 nicht-Kunden des IT-Dienstleisters gesendet wurden, zu erhöhen, wurden personalisierte Anfragen gesendet soweit es die vorhergehende Recherche zuließ. Adressiert wurde der Fragebogen an Personen innerhalb der Banken, die mit der Prüfung des Informationssystems betraut sind (Interne Revision), oder an der Auslagerung von IT-Aktivitäten mitwirken.

Der Fragebogen wurde so konzipiert, dass er im Internet ausgefüllt werden konnte. Da die Empfänger des Fragebogens von ihrem Arbeitsplatz aus generell Zugang zum Internet besitzen, erschien dieses Medium naheliegend. Zur Realisierung wurde die Freeware-Software Limesurvey® verwendet, die eine komfortable Konstruktion des Online-Fragebogens ermöglichte.

3 Ergebnisse der Umfrage

3.1 Allgemeine Anmerkungen zu der Studie

Von den insgesamt ca. 630 angeschrieben Finanzdienstleistern wurden 62 Fragebögen vollständig ausgefüllt. Damit beläuft sich die Rücklaufquote der Studie auf fast 10%

Obwohl der Fragebogen mit der Unterstützung des IT-Dienstleisters an mehr Banken aus seinen Kundenkreis versandt wurde als an sonstige deutsche Banken, lässt sich im Rücklauf der Fragebögen dieser Einfluss kaum erkennen. Es kann festgestellt werden, dass die Bankenverteilung in der Studie erstaunlich nahe an der tatsächlichen Verteilung von deutschen Banken heranreicht. Durch einen Homogenitätstest kann statistisch nachgewiesen werden, dass die empirische Verteilung der Banken in dieser Studie, repräsentativ für die tatsächliche Verteilung deutscher Banken ist. [1] So kann mit einem äußerst kleinen α-Fehler von 0,01% die Hypothese beibehalten werden, dass beide Verteilungen derselben Grundgesamtheit entstammen. Damit kann diese Studie Aussagen treffen, die auf alle deutschen Bankenarten selber maßen zutreffen.

Die Verteilung der Funktionen der Personen, die diesen Fragebogen ausgefüllt haben, lässt darauf schließen, dass der Fragebogen durchaus kompetent ausgefüllt wurde. So wurde der Großteil der Fragebögen entweder von Mitgliedern der Internen Revision, IT-Managern oder sogar von der Geschäftsleitung ausgefüllt. Nur 10% der Beteiligten erfüllten sonstige Aufgaben in Verbindung mit der Auslagerung von IT-Aufgaben. Davon gaben allerdings einige an, dass sie im Risikomanagement arbeiten oder direkt für die IT-Compliance in der Bank verantwortlich sind.

Um bei der Auswertung die Banken ihrer Größe nach kategorisieren zu können, wurde nach der Bilanzgröße gefragt. Diese wurde in 25%-Perzentile gruppiert. Es wurden also vier Gruppen gebildet, die annähernd die gleiche Anzahl an Banken enthalten. Daraus sind folgende Gruppen entstanden:

- kleiner als 600.000
- von 600.000 bis 7,25 Mio.
- größer als 7,25 Mio. bis 480 Mio.
- größer als 480 Mio.

[1] Die Grundlage für die Verteilung von Universalbanken in Deutschland wurde Hartmann-Wendels, Pfingsten, Weber (2007), S. 28 entnommen.

3.2 Ist der Umgang mit operationellen Risiken für Banken besonders wichtig?

Auf die Frage nach den gesetzlichen Regelwerken, die für die Bank relevant sind, gaben alle Beteiligten an, dass sie den Regeln des KWG unterworfen sind. Für zirka 18% der beteiligten Banken ist ebenfalls das AktG von Bedeutung. Lediglich 2% fallen unter die Jurisdiktion des Sarbanes-Oxley Acts (SOA).

Wie bereits dargelegt, kommt dem verantwortungsvollen Umgang mit operationellen Risiken, aufgrund des KWG im Bankensektor, eine besondere Bedeutung zu. Den tatsächlichen Stellenwert dieses internen Kontrollsystems und damit die tatsächliche Relevanz von operationellen Risiken in den befragten Banken, wird im Folgenden dargestellt. So geben immerhin 5% der befragten Banken an, dass das Management operationeller Risiken in Ihrem Unternehmen nicht so relevant sei. Die Unternehmen, die das angegeben haben, fallen sogar alle in die Größenkategorie „größer als 7,25 Mio. bis 480 Mio. “. Allerdings lässt sich allgemein festhalten, dass zumindest bei 93% aller befragten Banken, das interne Kontrollsystem einen durchaus hohen Stellenwert besitzt. Anzumerken ist ebenfalls, dass es keinen besonderen Zusammenhang zwischen dem Stellenwert des internen Kontrollsystems und der Unternehmensgröße gibt. So weisen die beiden Variablen eine Spearman-Korrelation nahe 0 auf.

Auf die Frage nach der Einschätzung, wie sich regulatorische Anforderungen im Bezug auf operationelle Risiken in der Zukunft entwickeln werden, gaben 94% der Banken an, dass die Anforderungen steigen werden. 3% gaben an, dass die Anforderungen gleich bleiben und die restlichen 3% gaben an, dass die Anforderungen sinken werden. Es lässt sich daher feststellen, dass die Banken sich eher auf eine verstärkte Regulierung der operationellen Risiken einstellen.

3.3 Sind die einzelnen Prüfungsstandards für das interne Kontrollsystem des IT Dienstleisters für deutsche Banken wichtig?

Wie zuvor beschrieben, sind Banken bei der Auslagerung von wichtigen IT-Prozessen dazu verpflichtet, einen Nachweis der internen Kontrollen in diesen Prozessen zu erbringen (MaRisk AT9). Welche Prüfungsstandards, vor diesem Hintergrund, für Banken relevant sind, wird im Folgenden beschrieben. Dazu wurde erst grundlegend nach der Wichtigkeit einer solchen Prüfungsleistung gefragt. Danach werden 60% der Banken bei zukünftigen Ausschreibungen auf solche Prüfungsberichte bestehen. 45% der Banken werden solche Prüfungsleistungen auch für die bisher ausgelagerten Prozesse verlangen. Immerhin sehen 24%der Banken eine solche Prüfungsleistung zwar als wichtig an, jedoch nicht als Voraussetzung für eine Auslagerung. Den restlichen 16% ist eine solche Prüfung sogar vollkommen unwichtig oder nicht so wichtig, wie etwa andere Konditionen bei der Auslagerung. Auch an dieser Stelle besteht keine signifikante Korrelation zur Unternehmensgröße der Bank (eine Spearman-Korrelation von 0,161 deutet nicht auf einen signifikanten Zusammenhang hin). Was jedoch überrascht, das ebenfalls keine nennenswerte Korrelation zu dem, zuvor dargestellten Stellenwert des internen Kontrollsystems im eigenen Unternehmen besteht (auch an dieser Stelle deutet eine Spearman-Korrelation von 0,246 nicht auf einen signifikanten Zusammenhang hin). Allgemein lässt sich festhalten, dass Prüfungen des internen Kontrollsystems des IT-Dienstleisters, für Banken durchaus relevant sind.

Um herauszufinden, welche Prüfungsstandards für die Banken besonders relevant sind, wurde zunächst nach der Bekanntheit einzelner Prüfungsstandards gefragt. So ist der IDW PS 330 mit 76% am bekanntesten. Danach folgt der IDW PS 951 mit 53% Bekanntheitsgrad. Bei nur etwa 11% der Banken war der internationale Prüfungsstandard SAS 70 bekannt. Die besondere Bekanntheit des IDW PS 330 ist sicher damit zu erklären, dass die Banken ihr internes Kontrollsystem selbst danach prüfen lassen. Im Gegensatz zum IDW 330, der allgemein das interne Kontrollsystem in einem Unternehmen prüft, ist der IDW PS 951 speziell auf die Prüfung des internen Kontrollsystems bei Auslagerung ausgerichtet. An dieser Stelle sei noch darauf hingewiesen, dass nur ca. 58%der Banken, die den IDW PS 951 kennen, auch über den Unterschied zwischen dem einfachen Typ A und restriktiveren Typ B Bescheid wussten. Bei den wenigen Banken, bei denen der SAS 70 bekannt ist, ist zu 88% auch der Unterschied zwischen Typ I und Typ II des SAS 70 Berichts bekannt.

Weiterhin wurde gefragt, welcher Prüfungsstandard sich aus Sicht der Bank am besten eignet, um ihre Anforderungen, bezüglich der Testierung des internen Kontrollsystems, zu erfüllen. Dazu konnten die Banken auf einer Likert-Skalar von eins bis fünf wählen, für wie geeignet sie die einzelnen Prüfungsstandards halten. Dabei bedeutete eins komplett geeignet und fünf nicht geeignet. Für die Auswertung wird auf den Median zurückgegriffen, der als Maß für den Mittelwert bei einer ordinal-skalierten Variabel verwendet werden kann.[2] Im Ergebnis ist festzustellen, dass sowohl der IDW PS 951 Typ B als auch der IDW PS 330, aus Sicht der befragten Banken, als durchaus geeignet erscheint, ihre Anforderungen zu erfüllen. So besaßen beide Prüfungsstandards einen Median von 2 („gut geeignet“). Der internationale Prüfungsstandard SAS 70 besitzt selbst unter den Banken, die ihn überhaupt kennen, eine eher untergeordnete Bedeutung. Der Typ I besaß einen Median von 5 (“ungeeignet“) und der Typ II mit einem Median von 3 (“bedingt geeignet“).

4 Zusammenfassung und Ausblick

In der obigen Auswertung wurden die Anforderungen deutscher Banken an Prüfungsleistungen, nach denen sich IT-Dienstleister ausrichten können, vorgestellt. Es wurde gezeigt, dass das interne Kontrollsystem, unabhängig von der Unternehmensgröße, schon heute einen hohen Stellenwert für Banken besitzt. Auch in Zukunft werden entsprechende Regulierungen aus Sicht der Banken immer restriktiver werden. Die Prüfung des internen Kontrollsystems bei dem IT-Dienstleister ist für Banken durchaus relevant und viele Banken werden entsprechende Zertifikate bei zukünftigen Ausschreibungen verlangen. Unter den Prüfungsstandards sind insbesondere die Prüfung nach IDW PS 330 und IDW PS 951 Typ B von großer Bedeutung.

Als Konsequenz ergeben sich damit besondere Anforderungen an die Unternehmens-IT von IT-Dienstleistern im Bankensektor. Diese müssen ihre IT-Landschaft compliant zu den Anforderungen aus den entsprechenden Prüfungsstandards ausrichten. So wird z.B. im IDW PS 951 auf den COSO Framework (Committee of Sponsoring Organizations of the Treadway Commission) zum Aufbau eines internen Kontrollsystems verwiesen. Für die IT-Landschaft hat COSO allerdings eine geringere Bedeutung. So werden zwar Kontrollen für die Finanzberichtserstattung (die im allgemeine durch IT-Systeme realisiert wird) und die allgemeine Regeleinhaltung definiert, direkte IT bezogene Kontrollen gibt es jedoch nicht ([FG07] S. 76.). Ein gutes Framework für effektive IT-Kontrollen bietet dagegen CobiT (Control Objectives for Information and related Technology), das ein Referenzmodellrahmen für die Planung und Steuerung des Einsatzes von Informationstechnik ist ([Me04] S. 79). Damit stellt es eine gute Ergänzung zu COSO dar ([FG07] S. 77).

[2] Auf die Angabe eines Streuungsmaßes wird an dieser Stelle verzichtet, da es für ordinal-skalierte Daten kaum ein aussagekräftiges Streuungsmaß gibt. So bezieht die Spannweite z.B. nur die größten und die kleinste Ausprägung mit ein.

Um sich nach diesen Standards compliant auszurichten wird auf Seiten des IT-Dienstleisters ein nicht geringer Aufwand entstehen. Weitere Forschung sollte sich der Frage widmen wie dieser Aufwand genau aussieht bzw. aus welchen Bestandteilen dieser besteht und welchen Umfang er besitzt. Es sollte eine Abwägung zu den Nutzen solcher Prüfungen des internen Kontrollsystems stattfinden.

Literaturverzeichnis

[FG07] Fröhlich, M.; Glasner, K.: IT Governance: Leitfaden für eine praxisgerechte Implementierung. 1. Aufl., Gabler Verlag, Wiesbaden 2007.

[Ge96] Gerlach, R.: Deutsche Regulierungslust - Segen oder Fluch für das Kreditgewerbe? . In: Sparkasse, 09 (1996); S. 404 - 408.

[GRB04] Grob, H. L.; Reepmeyer, J.-A.; Bensberg, F.: Einführung in die Wirtschaftsinformatik. 5. Aufl., Vahlen Verlag, München 2004.

[HPW07] Hartmann-Wendels, T.; Pfingsten, A.; Weber, M.: Bankbetriebslehre. 4. Aufl., Springer-Verlag, Berlin Heidelberg New York 2007.

[LS06] Lensdorf, L.; Steger, U.: IT-Compliance im Unternehmen. In: Der IT-Rechtsberater, H. 8 (2006), S. 206-210.

[Me04] Menzies, C.: Sarbanes-Oxley Act: Professionelles Management interner Kontrollen. 1. Aufl., Schäffer-Poeschel Verlag, Stuttgart 2004.

[SS08] Schüller, B.; Simon, A.: Verbesserung des Unternehmensertrags durch Sourcing. In: Outsourcing in Banken: Mit zahlreichen aktuellen Beispielen. 2. Aufl., Gabler Verlag, Wiesbaden 2008.

Integration von IT- und unternehmensweitem Risikomanagement auf Konzeptebene

Patrick Wolf, Matthias Goeken
Frankfurt School of Finance & Management
p.wolf@fs.de
m.goeken@fs.de

Abstract: Obwohl das IT-Risikomanagement (IT-RM) ein wichtiges Teilsystem des unternehmensweiten Risikomanagements (Enterprise Risk Management, ERM) darstellt, ist in der Praxis oft nur eine schwache Kopplung beider festzustellen. Es mangelt an Gestaltungsempfehlungen, um die Integration zu verbessern. Der Beitrag thematisiert die Integration von IT-RM und ERM auf Ebene ihrer Begrifflichkeiten (Konzepte). Mit Techniken der Grounded Theory werden aus bestehenden Standards und Normen des ERM relevante Konzepte für die Integration mit dem IT-RM herausgearbeitet. Diese werden verwendet, um IT-RM-Ontologien zu evaluieren und Weiterentwicklungspotenziale aufzuzeigen.

1 Einleitung

Informationstechnologie (IT) erfüllt sowohl eine Enabler- als auch eine Unterstützerfunktion [JK01], in dem sie IT-gestützte Geschäftsmodelle ermöglicht und eine effizientere Abwicklung von Aktivitäten erlaubt. Daher können sich aus der mangelnden Ausrichtung der IT an den Unternehmenszielen oder der Beeinträchtigung des IT-Betriebs schwerwiegende Konsequenzen für ein Unternehmen ergeben [Ju05] [Pr08]. Diese durch vorausschauende Identifikation und angemessene Steuerung von Risiken zu verhindern bzw. zu begrenzen, ist Aufgabe des IT-Risikomanagements (IT-RM) [Ju05].

Obwohl das IT-RM ein Teilsystem des unternehmensweiten Risikomanagements (Enterprise-RM, ERM) darstellt, ist in der Praxis oftmals nur eine schwache Verknüpfung zwischen IT-RM und ERM festzustellen [JK04]. Unter ERM wird die "Gesamtheit aller organisatorischen Regelungen und Maßnahmen zur Risikoerkennung und zum Umgang mit den Risiken unternehmerischer Betätigung [verstanden]" [IDW00]. IT-Risiken stellen dabei eine Teilmenge der Unternehmensrisiken dar. Wissenschaftliche Arbeiten thematisieren die Integration von IT-RM und ERM bislang nur selten [Ju05] und es fehlen fundierte Gestaltungsempfehlungen bezüglich der Integration. Diese sind nötig, damit im ERM aussagekräftige Informationen zu wesentlichen IT-Risiken vorhanden sind und das ERM ein zutreffendes Bild der Gesamtrisikosituation vermitteln kann.

Der Artikel thematisiert die Integration von IT-RM und ERM auf Ebene der Begrifflichkeiten (Konzeptebene). Die Forschungsfrage lautet: *„Welche Konzepte des ERM sind für die Gestaltung des IT-RM grundlegend, damit es dem ERM aussagekräftige und relevante Informationen zu kurz-, mittel- und langfristigen IT-Risiken bereitstellen kann?“*

2 Forschungsdesign

Das Forschungsdesign umfasst die Abgrenzung des Untersuchungsobjekts und die Explikation des Vorgehens mit Blick auf die Forschungsfrage [BE06]. Untersuchungsobjekt sind wissenschaftliche Arbeiten und Standards bzw. Normen zum ERM. Eine gute Übersicht findet sich u.a. bei [Wi07]. Aus Praktikabilitätsgründen wird nur eine Teilmenge der Grundgesamtheit betrachtet. Die bewusste Auswahl orientiert sich an folgender Anforderung: *„Die jeweilige Konzeption erfreut sich einer weiten Verbreitung bzw. hiermit ist aufgrund ihres Status zukünftig zu rechnen. Sie dient z.B. als Referenz für Prüfungen oder ist Ergebnis der Interpretation gesetzlicher bzw. regulatorischer Anforderungen."* Tabelle 1 zeigt die ausgewählten Konzeptionen.

IDW Prüfungsstandard 340 (IDW PS 340) – Die Prüfung des Risikofrüherkennungssystems nach § 317 Abs. 4 HGB	September 2000
Deutscher Rechnungslegungs-Standard Nr. 5 (DRS Nr. 5) – Risikoberichterstattung	Februar 2010
ISO 31000 - Risk management — Principles and guidelines	November 2009
Mindestanforderungen an das Risikomanagement (MaRisk)	August 2009
COSO Enterprise Risk Management (COSO ERM)	September 2004

Tabelle 1: Übersicht der selektierten Konzeptionen

Die Extraktion der für das IT-RM relevanten Konzepte aus den o.g. Normen und Standards erfordert die Deutung ihrer natürlichsprachigen Inhalte. Um hierbei die intersubjektive Nachvollziehbarkeit zu gewährleisten, wurde auf Techniken der Grounded Theory zurückgegriffen, eine durch Glaser und Strauss entwickelte Methode für die qualitative Feldforschung [GS67] [BHO08][Ku07]. Sie umfasst u.a. die systematische Auswertung von Textmaterial mit dem Ziel der Theoriebildung. Zentrales Element des Forschungsprozesses ist die Zuordnung von Textbestandteilen zu Kategorien (offenes Kodieren), das Herausarbeiten der Beziehungen zwischen diesen Kategorien (axiales Kodieren) und die Verdichtung der Kategorien zu Kernkategorien, auf deren Grundlage dann die Theorie formuliert wird (selektives Kodieren) [BHO08].

Die Aktivitäten beschränkten sich auf die Phase des offenen Kodierens: Es wurden die Inhalte der Untersuchungsobjekte analysiert, anschließend die relevanten ERM-Konzepte extrahiert und zu Kategorien verdichtet (z.B. "Risikotragfähigkeit"). Die Kategorien wurden dann in Anforderungen überführt, um eine Evaluierung bestehender IT-RM-Ontologien hinsichtlich der Vollständigkeit der berücksichtigten Konzepte zu ermöglichen. Ontologien stellen formale, explizite Modelle der von mehreren Individuen geteilten Sicht auf eine bestimmte Domäne (z.B. das IT-RM) dar [Gr93] [Us96]. Sie eignen sich besonders für die Analyse und Integration von Systemen auf Konzeptebene, da sie die Semantik der Konzepte einer Domäne eindeutig definieren. Die Evaluierung wird in Kapitel 4 beispielhaft für eine IT-RM-Ontologie vorgenommen.

3 Relevante Konzepte für die Integration von ERM und IT-RM

Tabelle 2 stellt die extrahierten Konzepte für die Integration von ERM und IT-RM dar. Nachfolgend wird die Relevanz der Konzepte begründet.

(1): Gan zheitliche ERM-Ansätze, wie ISO 31 000 od er COSO ERM, b asieren au f einem spek ulativen R isikoverständnis, de mgemäß unt er dem R isikobegriff so wohl Chancen (p ositive Ab weichungen) als au ch Gefah ren (negative Abweichun gen) subsumiert werden. Entsprechende Informationen müssen von den ERM-Subsystemen, wie d em IT-RM (z.B. wesentliche Ch ancen und Gefahren n euer Technologien), bereitgestellt werd en. Kon zept (1) sorg t für ein e au sgeglichene Betrach tung des IT-Einsatzes (Chancen/Gefahren) und die Verfügbarkeit dieser Informationen für das ERM.

ERM-Konzept	ERM-Quellen	Abgeleitete Anforderung an das IT-RM
(1) Spekulatives Risikoverständnis	ISO 31000, COSO ERM	Das IT-RM soll sowohl positive als auch negative Konsequenzen von IT-Risiken betrachten und dem ERM hierüber Informationen bereitstellen.
(2) Ganzheitlicher Organisationsfokus	IDW PS 340, ISO 31000, COSO ERM	Das IT-RM soll sämtliche betrieblichen Prozesse und Funktionsbereiche (einschließlich aller Hiearchiestufen und Stabsfunktionen), Strategien, Projekte, Produkte, Services und Vermögenswerte eines Unternehmens darauf untersuchen, ob und in welchem Umfang Risiken aus dem gegenwärtigen oder zukünftigen Einsatz von IT bestehen.
(3) Berücksichtigung von Wechselwirkungen	IDW PS 340, DRS 05, MaRisk, COSO ERM	Das IT-RM soll Wechselwirkungen zwischen IT-Risiken untereinander und bzgl. anderer Risikoarten betrachten. Es soll Auskunft darüber geben, ob IT-Risiken im Zusammenspiel mit anderen Risiken zu einer Bestandsgefährdung führen können.
(4) Zielbezogene Risikobetrachtung	IDW PS 340, DRS 05, ISO 31000, COSO ERM	Das IT-RM soll die möglichen kurz-, mittel- und langfristigen Konsequenzen der erkannten IT-Risiken für das Unternehmen bzw. für dessen Ziele aufzeigen.
(5) Risikotragfähigkeit	MaRisk	Das IT-RM soll gewährleisten, dass die wesentlichen IT-Risiken, ggf. unter Berücksichtigung von Wechselwirkungen, laufend durch die vorhandenen Risikodeckungsmassen des Unternehmens, also das zur Abfederung schlagend werdender Risiken verfügbare Kapital, gedeckt sind und somit die Risikotragfähigkeit jederzeit gegeben ist.
(6) Klare Verantwortlichkeiten	IDW PS 340, ISO 31000, COSO ERM	Das IT-RM soll klar festlegen, wer im Unternehmen für die Durchführung von RM-Aktivitäten und für einzelne IT-Risikofelder bzw. IT-Einzelrisiken, einschließlich der Risikobewältigung, verantwortlich ist und diesbezüglich welche Kompetenzen besitzt.
(7) Szenariobetrachtung / Stresstests	ISO 31000, MaRisk	Das IT-RM soll im Rahmen der Risikobewertung verschiedene Szenarien betrachten und Sensitivitäten zu den Einschätzungen angeben, um die Bandbreite an möglichen Konsequenzen sowie die Unsicherheit der Einschätzungen transparent zu machen.
(8) Einheitliche, geschäftsorientierte Bewertungskriterien	ISO 31000	Die Bewertungen im IT-RM sollen unter Verwendung einheitlicher Kriterien erfolgen, die sich an den unternehmerischen Zielen und weiteren relevanten Kontextfaktoren ausrichten (z.B. Risikoneigung, Risikotragfähigkeit). Die Kriterien sollen eine angemessene Priorisierung der Risiken ermöglichen.

Tabelle 2: Relevante Konzepte für die Integration von IT-RM und ERM

(2), (4) und (8): Das ERM m uss alle wesentlichen Chancen u nd Gefah ren aufzeigen, die Einfluss auf die kurz-, mittel- oder l angfristige Zielerreichung des Gesamtunternehmens haben können. Die Konzepte (2), (4) und (8) gewährleisten, dass sich das IT-RM an den Unternehmenszielen ausrichtet, Chancen und Gefahren der IT-Nutzung auf Basis einheitlicher Kriterien beurteilt und alle releva nten Unt ernehmensteile einbezieht. Ei n mittel- b is la ngfristiger Zeith orizont ist notwendig, weil w esentliche Ch ancen und Gefahren d es IT-Ei nsatzes mit st rategischen Ent scheidungen (z .B. Ge staltung der I T-Architektur) bzw. Umweltveränderungen (z.B. neue Technologien) einhergehen können.

(3) und (5): Die aktuelle W irtschaftskrise hat g ezeigt, dass eine isolierte Betrachtung von Ei nzelrisiken häufig z u kei ner a däquaten Ei nschätzung de r t atsächlichen R isikosituation führt [HO08], denn es bl eiben Wechselwirkungen zwi schen den Ei nzelrisiken unberücksichtigt, o bwohl si e ei ne Verä nderung des Gesamtrisikos zu r Fol ge ha ben können [HO08]. Standards (bspw. IDW PS 340 und DRS 5) empfehlen die Berücksichti-

gung von Wechselwirkungen bei der Aggregation von Einzelrisiken. Die aggregierten Risiken sollen den im Unternehmen vorhandenen Deckungsmassen (z.B. Eigenkapital) gegenübergestellt werden, um Transparenz über die Risikotragfähigkeit zu schaffen. Für Finanzinstitute ist dies gemäß MaRisk verpflichtend. Im IT-RM stellen die Konzepte (3) und (5) sicher, dass Wechselwirkungen zwischen IT-Risiken berücksichtigt und Steuerungsentscheidungen mit Blick auf die Risikotragfähigkeit getroffen werden.

(6): Ein wirksames ERM erfordert die Definition von Verantwortlichkeiten und Kompetenzen bzgl. der Erhebung, Analyse, Bewertung, Steuerung, Kommunikation und Überwachung der Risiken im Gesamtunternehmen bzw. in seinen Teilen. Konzept (6) soll gewährleisten, dass im Rahmen des IT-RM ebenfalls Verantwortlichkeiten und Kompetenzen klar geregelt und dem ERM hierüber Informationen bereitgestellt werden.

(7): Je weiter in die Zukunft geblickt wird, desto größer ist die Unsicherheit bzgl. des Eintretens oder Ausbleibens von Ereignissen. Das ERM trägt diesem Umstand Rechnung, indem verschiedene Szenarien betrachtet werden (z.B. Best-Case, Worst-Case). Hierdurch soll verhindert werden, dass das Unternehmen von zukünftigen Entwicklungen überrascht wird. Konzept (7) etabliert die Szenariobetrachtung im IT-RM, so dass verschiedene mögliche Entwicklungen bei der Bewertung berücksichtigt werden.

4 Berücksichtigung der Konzepte in bestehenden IT-RM-Ontologien

In den letzten Jahren sind verschiedene Ontologien mit Relevanz für die Domäne IT-RM entwickelt worden, wobei zumeist das IT-Sicherheitsmanagement im Mittelpunkt stand (siehe [Bl08] für einen Überblick). Es mangelt an Ontologien, die der Vielschichtigkeit der IT-Risiken sowie der Integrationsnotwendigkeit mit dem ERM adäquat Rechnung tragen. Tabelle 3 zeigt unter Berücksichtigung der ERM-Konzepte aus Kapitel 3 bestehende Defizite verfügbarer Ontologien am Beispiel der Security-Ontologie von [FE09] auf. [FE09] wurde gewählt, da sie Grundlage zahlreicher Veröffentlichungen war und praktisch erprobt ist. Abbildung 1 stellt die Kernelemente der Ontologie dar.

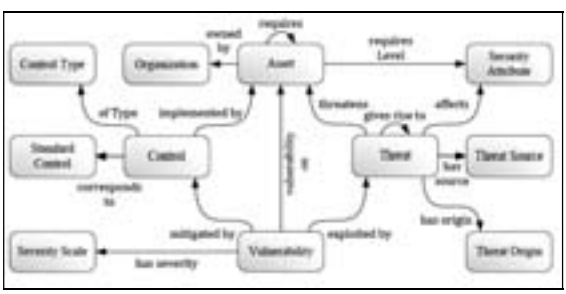


Abbildung 1: Kernkonzepte und -beziehungen der Ontologie von [FE09]

Konzept und IT-RM-Anforderung	Berücksichtigt in der Ontologie von [FE09]
(1) Spekulatives Risikoverständnis: Das IT-RM soll sowohl positive als auch negative Konsequenzen von IT-Risiken betrachten und dem ERM hierüber Informationen bereitstellen.	**Nein.** Die Ontologie umfasst keine Konzepte zur Abbildung der positiven Konsequenzen von IT-Risiken

(2) Ganzheitlicher Organisationsfokus: Das IT-RM soll sämtliche betrieblichen Prozesse und Funktionsbereiche, Strategien, Projekte, Produkte, Services und Vermögenswerte eines Unternehmens darauf untersuchen, ob und in welchem Umfang Risiken aus dem gegenwärtigen oder zukünftigen Einsatz von IT bestehen.	**Teilweise.** Die Ontologie beinhaltet das Konzept "Organization" auf oberster Ebene. Eine Verfeinerung nach Prozessen, Funktionsbereichen, Projekten, Produkten, etc. ist möglich, aber bislang nicht umgesetzt.
(3) Berücksichtigung von Wechselwirkungen: Das IT-RM soll Wechselwirkungen zwischen IT-Risiken untereinander und bzgl. anderer Risikoarten betrachten. Es soll Auskunft geben, ob IT-Risiken im Zusammenspiel mit anderen Risiken zu einer Bestandsgefährdung führen können.	**Ja.** Die Beziehungen zwischen Bedrohungen ("Threats") können mit Hilfe des Konzepts "Interrelations" abgebildet werden (kein Kernkonzept, daher nicht in Abb. 1 enthalten).
(4) Zielbezogene Risikobetrachtung: Das IT-RM soll die möglichen kurz-, mittel- und langfristigen Konsequenzen der erkannten IT-Risiken für das Unternehmen bzw. für dessen Ziele aufzeigen.	**Teilweise.** Bedrohungen werden mit den klassischen Schutzzielen verbunden (z.B. Verfügbarkeit, Vertraulichkeit, Integrität). Eine Verknüpfung mit geschäftlichen Zielen (z.B. Umsatz, Compliance) wird nicht thematisiert.
(5) Risikotragfähigkeit: Das IT-RM soll gewährleisten, dass die wesentlichen IT-Risiken, ggf. unter Berücksichtigung von Wechselwirkungen, laufend durch die vorhandenen Risikodeckungsmassen des Unternehmens gedeckt sind und so die Risikotragfähigkeit gegeben ist.	**Nein.** Das Konzept ist in der Security Ontologie nicht berücksichtigt.
(6) Klare Verantwortlichkeiten: Das IT-RM soll klar festlegen, wer im Unternehmen für die Durchführung von RM-Aktivitäten und für einzelne IT-Risikofelder bzw. IT-Einzelrisiken, einschließlich der Risikobewältigung, verantwortlich ist und welche Kompetenzen besitzt.	**Nein.** Das Konzept ist in der Security Ontologie nicht berücksichtigt.
(7) Szenariobetrachtung / Stresstests: Das IT-RM soll im Rahmen der Risikobewertung verschiedene Szenarien betrachten und Sensitivitäten zu den Einschätzungen angeben, um die Bandbreite an mögliche Konsequenzen und die Unsicherheit der Einschätzungen transparent zu machen.	**Nein.** Das Konzept ist in der Security Ontologie nicht berücksichtigt.
(8) Einheitliche, geschäftsorientierte Bewertungskriterien: Die Bewertungen im IT-RM sollen unter Verwendung einheitlicher Kriterien erfolgen, die sich an den unternehmerischen Zielen und weiteren relevanten Kontextfaktoren ausrichten (z.B. Risikoneigung, Risikotragfähigkeit). Die Kriterien sollen eine angemessene Priorisierung der Risiken ermöglichen.	**Teilweise.** Die Ontologie sieht eine einfache Bewertung von Bedrohungen und Maßnahmen-Effekten mittels Likert-Skala (hoch, mittel, gering) vor. Eine quantitative oder semi-qualitative Bewertung wird ebenso wenig thematisiert wie relevante Kontextfaktoren der Beurteilung (z.B. Zeithorizont, Risikoneigung).

Tabelle 3: Evaluierung der Security-Ontologie von [FE09]

Die Evaluierung zeigt, dass viele ERM-Konzepte nur unzureichend in der Ontologie von [FE09] berücksichtigt sind. Für ein ganzheitliches, geschäftsorientiertes und integriertes IT-RM wäre somit eine Weiterentwicklung erforderlich. Angesichts des wachsenden Kostendrucks und der zunehmend engeren Verzahnung von Geschäftsmodellen, Abläufen und IT ist zu erwarten, dass einem ganzheitlichen, geschäftsorientierten und integrierten IT-RM eine immer gewichtigere Rolle zukommen wird. Auch die ISACA unterstreicht diese Einschätzung mit dem jüngst veröffentlichten "Risk IT"-Referenzmodell, das eine ganzheitliche Betrachtung von Chancen und Gefahren der IT fordert und der Integration von IT-RM und ERM einen eigenen Prozess widmet [Is09].

5 Limitationen der Analyse und Ausblick

In diesem Artikel wurden relevante Konzepte für die Integration von ERM und IT-RM herausgearbeitet. Ihre Berücksichtigung in IT-RM-Ontologien und IT-RM-Ansätzen soll gewährleisten, dass dem ERM aussagekräftigere Informationen zu relevanten IT-Risiken bereitgestellt werden. Wichtige Limitationen stellen einerseits die Beschränkung auf einzelne ausgewählte Standards und Normen dar, andererseits die subjektive Beurteilung der Relevanz der Konzepte. Durch Offenlegung des Auswahlkriteriums und der Begründungen für die Relevanz wurde versucht, die Nachvollziehbarkeit zu erhöhen. Eine weitere Limitation besteht darin, dass die Analyse nur an einem als repräsentativ erach-

teten Beispiel diskutiert wurde. Eine Analyse weiterer Ontologien zur Fundierung soll erfolgen. Weitere Forschungsaktivitäten zielen auf die Konstruktion einer Ontologie für das geschäftsorientierte, ganzheitliche und integrierte IT-RM. Bestehende Security-Ontologien sowie "Risk IT" der ISACA sollen als Ausgangspunkt dienen. Bislang mangelt es innerhalb der Domäne des IT-RM an entsprechenden Ontologien, die der Vielschichtigkeit der IT-Risiken angemessen Rechnung tragen, nicht nur auf einzelne Teilaspekte fokussieren und die Integration mit dem ERM unterstützen.

Literaturverzeichnis

[BE06] Braun, R.; Esswein, W.: Eine Methode zur Konzeption von Forschungsdesigns in der konzeptuellen Modellierungsforschung. In (Schelp, J. et. al. Hrsg.): Integration, Informationslogistik und Architektur, Proc. DW2006. Bonn, 2006; S. 143-171.

[BHO08] Brühl, R.; Horch, N.; Orth, M.: Grounded Theory und ihre bisherige Anwendung in der empirischen Controlling- und Rechnungswesenforschung. In: Zeitschrift für Planung & Unternehmenssteuerung, 19. Jg., Heft 3, 2008; S. 299-323.

[Bl08] Blanco, C. et. al.: A systematic Review and Comparison of Security Ontologies. In: Proceedings of the 2008 Third International Conference on Availability, Reliability and Security (ARES). IEEE Computer Society, Barcelona, Spain, 2008; S. 813-820.

[FE09] Fenz, S.; Ekelhart, A.: Formalizing Information Security Knowledge. In: Proceedings of the 4th ACM Symposium on Information, Computer, and Communications Security (ASIACCS); ACM, Sydney, Australia, 2009; S. 183-194.

[Gr93] Gruber, T. R.: A translation approach to portable ontology specifications. In: Knowledge Acquisition, Jg. 5, Heft 2, 1993; S. 199-220.

[GS67] Glaser, B. G.; Strauss, A. L.: The Discovery of Grounded Theory: Strategies for Qualitative Research. Aldine Pub. Co, Chicago, 1967.

[HO08] Hempel, M.; Offerhaus, J.: Risikoaggregation als wichtiger Aspekt des Risikomanagements. In (Deutsche Gesellschaft für Risikomanagement e.V. Hrsg): Risikoaggregation in der Praxis – Beispiele und Verfahren aus dem Risikomanagement von Unternehmen, Springer, Berlin, Heidelberg, 2008; S. 3 – 13.

[Is09] ISACA (Hrsg.): The Risk IT Framework. ISACA, Rolling Meadows, 2009.

[IDW00] Institut der Wirtschaftsprüfer (IDW): IDW Prüfungsstandard: Die Prüfung des Risikofrüherkennungssystems nach § 317 Abs. 4 HGB (IDW PS 340) (Stand: 11.09.2000). In: IDW-Fachnachrichten, 1999, Nr. 8, S. 350-357.

[JK01] Junginger, M.; Krcmar, H.: IT Risk Management – Fit für E-Business?. In (Buhl, H. U.; Huther, A.; Reiwieser, B. Hrsg.): Information Age Economy; 5. Internationale Tagung Wirtschaftsinformatik, Heidelberg, 2001. Physica-Verlag, Heidelberg, 2001.

[JK04] Junginger, M.; Krcmar, H.: Wahrnehmung und Steuerung von Risiken im Informationsmanagement – Eine Befragung deutscher IT-Führungskräfte. Studie, Lehrstuhl für Wirtschaftsinformatik, Technische Universität München, Garching, 2004.

[Ju05] Junginger, M.: Wertorientierte Steuerung von Risiken im Informationsmanagement. DUV/GWV, Wiesbaden, 2005.

[Ku07] Kuckartz, U.: Einführung in die computergestützte Analyse qualitativer Daten.2., aktualisierte und erweiterte Auflage. VS/GWV, Wiesbaden, 2007.

[Pr08] Prokein, O.: IT-Risikomanagement – Identifikation, Quantifizierung und wirtschaftliche Steuerung. Gabler, Wiesbaden, 2008.

[Us96] Uschold, M.: Building Ontologies: Towards a unified methodology.In: Proceedings of the 16th Annual Conference of the British Society Specialist Group on Expert Systems, Cambridge, 1996.

[Wi07] Winter, P.: Risikocontrolling in Nicht-Finanzunternehmen. Lohmar-Köln, 2007.

Questioning the need for separate IT risk management frameworks

Nicolas Racz[1], Edgar Weippl[1], Andreas Seufert[2]

[1]Institut für Softwaretechnik und Interaktive Systeme
TU Wien
Favoritenstr. 9-11/188
1040 Wien, Austria
[2]Institut für Business Intelligence, Steinbeis Hochschule Berlin
{racz, eweippl}@ifs.tuwien.ac.at; andreas.seufert@i-bi.de

Abstract: The growing importance of enterprise risk management and the resulting integration efforts put the need for separate IT risk management frameworks in question. In this research we analyse common and distinct elements of the COSO enterprise risk management and ISACA Risk IT frameworks. The analysis affirms the hypothesis that separate IT risk management frameworks are redundant.

1 Motivation

The alignment of IT with business objectives is an important part of contemporary IT management. Ever since the creation of the terms "enterprise risk management" (ERM) and "governance, risk, and compliance" (GRC) the search for integration possibilities within these disciplines has been ongoing. However as of today different frameworks are used for the management of business risks and IT risks. The emergence of horizontal integration (across disciplines and across departments) and vertical integration (across the organisational hierarchy and across process levels) has helped to realise that formerly separate approaches are often redundant [Mi07], which provokes the authors to establish the hypothesis that a separate management of IT risks might not be justified.

2 Status quo in research and practice

A quick scan of the ACM, SpringerLink and EmeraldInsight databases shows that in research as of today enterprise risk management and IT risk management (IT-RM) have hardly ever crossed paths. Foley uses ERM processes to manage security risks [Fo09]. In their high-level process model for IT GRC management, Racz et al. [RWS10] use COSO ERM [CO04] as risk management standard, assuming that the selection of a framework that does not focus on IT would facilitate integration with non-IT GRC in future research.

In practice enterprise risk management and IT risk management are also treated as separate topics. With ISO 31000:2009 [ISO09] (superseding AS/NZS 4360:2004 [AS04]) and ISO/IEC 27005:2008 [ISO08] the International Organization for Standardization treats ERM and information security risk management (including IT-RM) in two distinct standards. ISO 31000 does not even reference ISO/IEC 27005. The alignment of IT with business in practice is mainly done through the IT governance and management frameworks COBIT (Control Objectives for Information and related Technology [ITG07]) and ITIL (IT Infrastructure Library [OGC07]). These frameworks suggest enabling alignment through deriving IT goals from business goals.

We can conclude that while the connection of IT risks with business objectives is enforced at present, the merger of IT-RM with ERM on a process level is hardly looked at. The frame of reference for research of integrated GRC [Ra10] recommends identifying integration possibilities on the strategic, process, organisational and technology level. Strategically, through the alignment of IT goals with business goals, the integration is already ongoing. We suggest to take the next step and to review potential synergies of ERM and IT-RM on the process level. Following the claim of ERM to cover all risks of an enterprise, IT-RM should either be completely covered by ERM and therefore be redundant; or it might enhance the broader ERM through detailed consideration of IT specifics in the risk management process.

3 Methodology

As a first step in evaluating our hypothesis we decided to carry out an exemplary comparison of an ERM framework with an IT-RM framework. Of course a comparison of two frameworks is not representative, but we selected widely-used frameworks (see below) that therefore suffice to provide a first indication about the hypotheses' validity. The results should then be discussed with other experts at the Informatik 2010 GRC workshop before taking further action. The methodology applied consists of four steps. First, we selected a framework for ERM and one for IT-RM. Second, the frameworks' commonalities were identified. Third, we analysed the references of the ERM framework to IT risk and vice versa. Finally we discussed and summed up the results.

In the selection process for an ERM framework we considered ISO 31000:2009 and COSO ERM, two well-known standards for ERM. Their process models are very similar. On a high level they only differ in their wording. "Establishing the context" in the ISO standard corresponds to the "internal environment" of COSO ERM, "risk evaluation" and "risk treatment" equal "risk response" and "control activities", etc. Eventually we opted for COSO ERM, as it is the successor of the widely implemented COSO framework for internal control [CO92], a de-facto standard explicitly acknowledged in the US Public Company Accounting Oversight Board Auditing Standard No. 5 for financial reporting [PCA07]. The standard is referenced in the Sarbanes Oxley Act of 2002, which of all regulations passed in the new millennium probably has the strongest impact on risk management and internal control systems.

For IT risk management we chose the ISACA Risk IT Framework because it complements COBIT, which is arguably the most appropriate control and governance framework used by many organisations world-wide to ensure alignment of IT and business goals [RYC04]. The framework's importance is expected to grow since the new COBIT version 5, which is currently in development, plans to consolidate and integrate the Risk IT framework [ISA10]. ISO/IEC 27005:2008 was also considered. As it includes all aspects of information security (including non-IT aspects), its scope surpasses the ISACA framework, which is limited to information technology. In our opinion Risk IT is more detailed, and it draws out the specifics of IT-RM more clearly.

The identification of the frameworks' commonalities in the second phase of our research was done through a mapping of the described processes of ISACA Risk IT to those of COSO ERM. The documentation of COSO ERM proved to be a hurdle. On the highest level the framework consists of seven processes and the "internal environment" component. Unfortunately the processes are not broken down. Instead COSO just names the basic sub-components, such as "risk tolerance" or "inherent and residual risk". In order to map the processes of ISACA Risk IT, we had to go through the complete description of the COSO components to find if the same processes were included.

The qualitative analysis of references from ERM to IT-RM and vice versa in the third research step was followed by a descriptive discussion and summary of the insights gained in the research process.

4 Results and discussion

4.1 Mapping of ISACA Risk IT to COSO ERM

Our comparison of risk management frameworks is based on the assumption that "risk" in ERM has the same characteristics as "risk" in IT-RM. In COSO ERM, risk is "the possibility that an event will occur and adversely affect the achievement of objectives; events with a potentially positive impact may offset negative impacts or they may represent opportunities" [Co04]. Throughout the framework "risk" then also refers to upside risk (opportunities). According to ISACA Risk-IT, IT risk is "a component of the overall risk universe of the enterprise [...]. IT risk is business risk [...]. It consists of IT-related events and conditions that could potentially impact the business" [ISA09]. The two frameworks consequently share a common understanding of the term "risk".

ISACA Risk-IT consists of the three processes risk governance, risk evaluation, and risk response on level one, with three sub-processes each on level two. Level three comprises 43 processes. COSO ERM on the other hand describes 8 high level processes with 41 sub-components. While the ERM framework is more profound on the internal environment component and on risk aggregation, Risk IT is more specific when it comes to IT specific and communication processes. Still, all but seven of the IT-RM processes can easily be mapped to COSO components (see appendix A).

Two of the exceptions deal with ERM integration: “RG2.2: Co-ordinate IT risk strategy and business risk strategy”, and “RG2.3: Adapt IT risk practices to enterprise risk practices”. They treat the alignment of IT and business risks on a strategic and on a process level; we will analyse them later on in the section about ERM references in the IT-RM framework. Three other processes that could not be mapped belong to the process group “RG3: Make risk-aware business decisions”: “RG3.1: Gain management buy-in for the IT risk analysis approach”, “RG3.2: Approve IT risk analysis”, and “RG3.5: Prioritise IT risk response activities”. Management buy-in for risk analysis approaches and their approval is not explicitly mentioned in COSO ERM, but it could seamlessly be integrated with the “internal environment” component. Prioritisation of response activities is probably so self-evident that COSO ERM does not highlight it; in COSO the prioritisation could be part of risk response. Furthermore the processes “RE2.4: Perform a peer review of IT risk analysis” and “RE3.3: Understand IT capabilities” do not exist in COSO ERM. Peer reviews are a control mechanism that can be seamlessly included in the ERM process. Understanding IT capabilities is an extremely general “process” that is a prerequisite for any kind of IT activity, therefore suitable to be added to the “internal environment” component of COSO ERM.

As we can see, drawing from the standards IT risks may be treated like any other risk, as the IT-RM framework is completely absorbed in COSO ERM, apart from the ERM integration specifics (RG2.2, RG2.3) analysed below. The ISACA framework does not explain why an IT-specific risk management framework in the hierarchical relationship to ERM would be necessary. It even disposes of the distinction by stating that “IT risk is business risk”, consisting of “IT-related events that could potentially impact the business” [ISA09]. Thus the need for separate IT risk frameworks is questionable. It seems to be owed more to the complexity of IT, to habits and to the separation of IT and business responsibilities in modern organisations than to a real business reason.

4.2 References of COSO ERM to ISACA Risk IT and vice versa

In fact the Risk IT Framework (RG1.1) recommends taking a top-down, end-to-end look at business services and processes and identifying the major points of IT support. However it does little to support this advice. The relation to ERM is explicitly treated in the framework. “Integrate with ERM” as a sub-process of “risk governance” states as goal to integrate the IT risk strategy and operations with the business strategic risk decisions that have been made at the enterprise level. Five key activities shall help achieve this goal. Three of them are governance processes indispensable for any risk domain: establishing and maintaining accountability for IT-RM (RG2.1), providing adequate resources for IT-RM (RG2.4) and providing independent assurance over IT-RM (RG2.5). RG2.1 involves business with IT risk through risk ownership and the ability to address IT risk issues. RG2.4 weighs investing resources for IT risks with investments in competing business risk issues, thus surpassing the IT risk domain and respecting all risk domains of ERM. RG2.5 actually is not ERM-specific at all.

Consequently we are left with the two other processes allegedly dealing with ERM integration that could not be mapped to COSO ERM before: “co-ordinate IT risk

strategy and business risk strategy" (RG2.2) and "adapt IT risk practices to enterprise risk practices" (RG2.3). RG2.2 requires to "integrate any IT specifics into one enterprise approach" and to define the IT department's role in operational risk management. Existing ERM principles and views of risk should be used wherever possible. How this integration works is not explained. RG2.3 demands that the business context for IT, and ERM expectations, activities and methods relevant to IT-RM be understood. IT-RM should be enhanced with useful ERM activities, ERM expectations should be met, and methods of other functions should be identified. The gaps between IT risk and ERM shall be closed – but the framework owes a clear explanation of how this could be done.

The COSO ERM framework on the other hand gives even less advice on IT-RM. It is only high-level guidance as far as IT is concerned, but specifics of IT risk management may still be considered on lower process levels [Mo07]. It mentions the importance of information systems controls due to the "widespread reliance on information systems" [CO04]. General controls shall ensure the continued, proper operation of information systems, while application controls ensure completeness, accuracy and validity of information. General controls are further subdivided into controls for information technology management, information technology infrastructure, security management and software acquisition, development and maintenance. Apart from these control-related hints there is no detailed reference in COSO ERM to information technology. IT risks are not even mentioned. Thus the COSO ERM document remains on a very high level, not helping practitioners deal with IT risks in the ERM context.

4.3 Discussion

Drawing on the results we see the hypotheses that a separate framework for IT-RM might not be necessary preliminarily affirmed. ISACA implies a hierarchical structure between ERM and IT-RM, but our research rather suggests that the IT-RM framework might inhibit the integration with ERM through introduction of a redundant framework into the process. Certainly the comparison of two frameworks is not sufficient to prove the hypothesis, but it is a hint that further efforts to confirm the assertion are worthwhile. Future research would have to provide real case study examples to prove the point.

In practice today IT-RM is started within the IT organisation and it is aligned with business mainly through business objectives. ERM is a top-down approach, and IT-RM is top-down within IT, but bottom-up on the enterprise level, as IT risks are analysed and subsequently linked to business objectives and quantifications from operational risk management. For example an IT risk manager might look at a database and identify the data therein, then find out which applications it is used in, next look at which business processes they support and, eventually, what the (financial) impact on these processes would be if the data lost its integrity, validity, privacy or availability [RS09]. Historically the coexistence of ERM and IT-RM can be explained because enterprise-wide approaches to risk have only emerged over the last decade (COSO ERM as the first ERM framework was only published in 2004). IT-RM meanwhile has been around for much longer due to ever-present IT security and operational issues.

We argue that the more reasonable way to manage risks would be to follow a business process top-down to all its enabling resources, be they human or natural resources, technology or information. Starting at the process level, business would have to consult IT as part of the ERM exercise to deliver the IT resources linked to a specific process on the application, data and infrastructure level. Then the events and risks (e.g. data loss due to a virus) could be analysed hand-in-hand by business and IT. The main advantage of this end-to-end approach is that only relevant, value-creating business processes would be considered, and that they could be prioritised early-on.

5 Conclusion and future research

The analysis of the COSO ERM and ISACA Risk IT frameworks has shown that the need for a separate IT-RM framework indeed is questionable. The majority of IT-RM processes match the ERM components; the few remaining processes can be integrated with ERM. We recommend future research to evaluate the possibility of creating an integrated approach to IT risks within enterprise risk management that makes the application of separate IT-RM frameworks redundant.

Bibliography

[AS04] AS/NZS 4360:2004. Risk management. AS/NZS, 2004.

[CO92] COSO: Internal control – integrated framework. 1992. www.coso.org

[CO04] COSO: Enterprise risk management framework. 2004. www.coso.org

[Fo09] Foley, S.: Security Risk Management using Internal Controls. WISG, 2009.

[ISA09] ISACA: The Risk IT Framework. ISACA, Rolling Meadows, 2009.

[ISA10] ISACA: COBIT 5 Design Paper Exposure Draft. 2010. www.isaca.org

[ISO08] ISO/IEC 27005:2008. Information technology – Security techniques – Information security risk management. ISO/IEC, 2008.

[ISO09] ISO 31000:2009. Risk management – principles and guidelines. ISO, 2009.

[ITG07] IT Governance Institute: COBIT 4.1. ISACA, Rolling Meadows, 2007.

[Mi07] Mitchell S.L.: GRC360: A framework to help organisations drive principled performance. In: Int. Journal of Disclosure and Governance, 4:4, 2007; S. 279-296

[Mo07] Moeller, R.R.: COSO Enterprise Risk Management. Wiley, New Jersey, 2007.

[OGC07] Office of Government Commerce: ITIL v3, 2007. http://www.itil-officialsite.com

[PCA07] Public Company Accounting Oversight Board: Auditing Standard No. 5. http://www.pcaobus.org/Rules/Rules_of_the_Board/Auditing_Standard_5.pdf

[Ra10] Racz, N.; Weippl, E.; Seufert, A.: A frame of reference for research of integrated governance, risk & compliance (GRC). In (De Decker, B.; Schaumüller-Bichl, I., Hrsg.): Communications and Multimedia Security. Springer, Berlin, 2010; S. 106-117

[RS09] Rath, M.; Sponholz, R.: IT-Compliance: Erfolgreiches Management regulatorischer Anforderungen, ESV Erich Schmidt Verlag, Berlin, 2009

[RWS10] Racz, N.; Weippl, E.; Seufert, A.: A Process Model for Integrated IT Governance, Risk & Compliance Management. Databases and Information Systems VI. Selected Papers from the Ninth International Baltic Conference, DB&IS 2010.

[RYC04] Ridley, G.; Young, J.; Carroll, P.: COBIT and its utilization. A framework from the literature. 37th Hawaii International Conference on System Sciences Proceedings, 2004.

Appendix A

Mapping of ISACA Risk IT processes to COSO ERM components. "Risk communication" and "Risk culture" in the RITF are not part of the process model, but they are separately described in the framework document and have therefore been added. The wording of two mapped components might be very different, especially since the COSO components have very general names and sometimes include a variety of processes in their description. Each of the three authors first did the mapping on his own using the COSO ERM and ISACA Risk-IT process descriptions. Results were then merged and discrepancies were discussed until a joint decision could be taken.

COSO ERM Framework	ISACA Risk IT Framework
01 Internal environment	
01.01 Risk management philosophy	
01.02 Risk appetite	
01.03 Risk culture	RG1.5 Promote IT risk-aware culture
	Risk Culture
01.04 Board of directors	
01.05 Integrity and ethical values	
01.06 Commitment to competence	
01.07 Management philosophy and operating style	
01.08 Organisational structure	
01.09 Assignment of authority and responsibility	RG2.1 Establish and maintain accountability for IT risk management
	RG2.4 Provide adequate resources for IT risk management
01.10 Human resource policies and practices	
01.11 Differences in environment	
02 Objective setting	RE2.1 Define IT risk analysis scope
02.01 Strategic objectives	
02.02 Related objectives	RG2.4 Provide adequate resources for IT risk management
02.03 Selected objectives	
02.04 Risk appetite	RG3.3 Embed IT risk considerations in strategic business decision making
02.05 Risk tolerance	RG1.2 Propose IT risk tolerance thresholds
	RG1.3 Approve IT risk tolerance
03 Event identification	
03.01 Events	RE3.4 Update IT risk scenario components
03.02 Factors influencing strategy and objectives	RE3.5 Maintain the IT risk register and IT risk map
03.03 Methodology and techniques	RE3.6 Develop IT risk indicators
03.04 Event interdependencies	RE1.3 Collect data on risk events
03.05 Event categories	RE1.4 Identify risk factors
03.06 Risks and opportunities	RR1.4 Identify IT-related opportunities
04 Risk assessment	RG1.1 Perform enterprise IT risk assessment
04.01 Inherent and residual risk	RG3.4 Accept IT risk (= accept residual risk)
04.02 Likelihood and impact	RE2.2 Estimate IT risk

	RE3.1 Map IT resources to business processes
	RE3.2 Determine business criticality of IT resources
04.03 Qualitative and quantitative methodologies and techniques	RE1.1 Establish and maintain a model for data collection
	RE1.2 Collect data on the operating environment
04.04 Correlation	
05 Risk response	
05.01 Identify risk responses	RE2.3 Identify risk response options
05.02 Evaluate possible risk responses	RR1.3 Interpret independent IT assessment findings
05.03 Select response	RR3.1 Maintain incident response plans
	RR3.3 Initiate incident response
05.04 Portfolio view	
06 Control activities	
06.01 Integration with risk response	RR2.1 Inventory controls
06.02 Types of control activities	RR2.3 Respond to discovered risk exposure and opportunity
06.03 General controls	RR2.4 Implement controls
06.04 Application controls	
06.05 Entity-specific	
	RR3.2 Monitor IT risk
	RR2.2 Monitor operational alignment with risk tolerance thresholds
08 Monitoring	
08.01 Ongoing	
08.02 Separate evaluations	RG2.5 Provide independent assurance over IT risk management
08.03 Reporting deficiencies	
07 Information and communication	
07.01 Information	
07.02 Strategic and integrated systems	
07.03 Communication	RR2.5 Report IT risk action plan progress
	RR3.4 Communicate lessons learned from risk events
	RR1.1 Communicate IT risk analysis results
	RR1.2 Report IT risk management activities and state of compliance
	RG1.6 Encourage effective communication of IT risk
	RG1.4 Align IT risk policy
	Risk Communication

Unternehmensarchitekturmanagement als Mittler zwischen IT-Governance und IT-Management

Konrad Walser, Reinhard Riedl

Berner Fachhochschule
PEG – Kompetenzzentrum für Public Management und E-Government
Morgartenstrasse 2a/Postrach 305
CH-3000 Bern 22
konrad.walser@bfh.ch; reinhard.riedl@bfh.ch

Abstract: Im vorliegenden Beitrag werden Zusammenhänge zwischen IT-Governance und IT-Management und dem Unternehmensarchitekturmanagement differenziert und analysiert. Dadurch soll einerseits die aktuell unterdotierte Rolle der Unternehmensarchitektur in Unternehmen dargestellt werden. Andererseits wird im Beitrag dargestellt, wie Unternehmensarchitektur als Mittlerobjekt zwischen IT-Governance und -Management positioniert werden kann. Durch die integrative Sicht steigt der Nutzen des Unternehmensarchitekturmanagements als Prozess. Es erhält dadurch eine neue zentrale Drehscheibenfunktion zur Zusammenschau verschiedener IT-Governance- und -Management-Teilbereiche und bietet Anknüpfungspunkte für diese. Das Unternehmensarchitekturmanagement bildet eine adäquate gemeinsame Gestaltungs- und Arbeitsplattform für die Integration von IT-Governance und –Management, lautet das Fazit dieses Beitrags.

1 Einleitung

1.1 Begriffliche Klärungen

Zunächst sind einige begriffliche Klärungen erforderlich, um die folgenden Ausführungen zu verstehen. Das Management der IT-Strategie, abgeleitet aus der Unternehmensstrategie, umfasst u.a. die folgenden Themenbereiche ([Kr04; 284 ff.], [WW02]). Ausgehend von einer definierten IT-Strategie sind Fragen des IT-Sourcing und der IT-Leistungserbringung zu klären. Es sind Klärungen der Nutzenbeiträge der IT für das Geschäft erforderlich [ITG06]. Weiter sind Klärungen bezüglich der IT-Aufbau- und -Ablauforganisation erforderlich. Zu den Governance- und -Compliance-Aufgaben, die zu definieren sind, gehören Überwachungsaufgaben der IT-Leistungserbringung sowie Vertragsmanagement, IT-Controlling, IT-Portfolio- sowie IT-Risikomanagement. Klärungen zu den Themenbereichen erfolgen im Rahmen laufender IT-Strategiebildungs- und deren taktischer Umsetzungsprozesse. Output daraus sind Unternehmens-IT-Strategiepapiere. [DV04] definieren IT-Governance wie folgt: „IT governance is the organizational capacity exercised by the Board, executive management and IT management to control the

formulation and implementation of IT strategy and in the way ensure the fusion of business and IT“. IT Management ist definierbar als strategische und operative Führung der IT-Serviceerbringung. Die IT Strategie determiniert die IT-Governance und das IT-Management [HV93]. Nach TOGAF/ANSI/IEEE ist das Unternehmensarchitekturmanagement (UAM) wie folgt definiert:[1] „The fundamental organization of a system, embodied in its components, their relationships to each other and the environment, and the principles governing its design and evolution“ [BHH07]. Das UAM ist als einer unter verschiedenen zu führenden IT-Prozessen zu verstehen, die strukturierend auch für IT-Governance und -Management wirken. In der Unternehmensarchitektur nach TOGAF werden vier Sichten auf die Architektur unterschieden, nämlich Geschäfts-, Anwendungs-, System- und Informationssicht, die in ihrem Zusammenhang Mittlerobjekte zwischen organisatorischen Rollen von IT-Governance und -Management darstellen.[2]

1.2 Problemstellung

Ein UAM bringt als alleinstehendes Instrument für IT-Governance und -Management keinen Nutzen. Dieser entsteht für Unternehmen erst dann, wenn der intensive Abgleich mit den thematisierten IT-Governance-Instrumenten konkretisiert und deren für das Unternehmen relevante Beziehungen herausgearbeitet werden. Das UAM ist in Unternehmen vielfach unterentwickelt. Noch stärker trifft dies auf die dringend notwendige Zusammenschau mit IT-Governance- und -Management-Instrumenten und Referenzmodellen zu.[3] Das UAM stellt in seinen unterschiedlichen Darstellungsformen das zwingend notwendige Bindeglied für ein erfolgreich geführtes und umfassend implementiertes IT-Management dar.[4] Das UAM umfasst alle erforderlichen Eigenschaften und Sichten auf das Informationssystem Unternehmen, um zur Überwindung der Grenzen zwischen Geschäfts- und IT-Seite (IT-Leistungserbringer (LE) und -bezüger, (LB)) beizutragen. Die verschiedenen Sichten auf die grafische Darstellung der UAM (Geschäfts-, Anwendungs-, System- und Informationssicht) und das intensive Bemühen um ein Verständnis für die Zusammenhänge zwischen denselben in grafischer Ausprägungsform stellen das ideale Mittlerobjekt zwischen verschiedenen IT-Governance- und -Ma-nagement-Instrumenten und –Rollen dar, die in Unternehmen direkt oder indirekt mit IT befasst sind (Business-IT-Alignment [Ni05]). Die Umsetzung der Konzepte einer Unternehmens-IT-Governance und -IT-Managements in Verbindung mit der UA ist bislang wenig konkret. Um die Zusammenhänge zwischen den im Beitrag fokussierten Themen besser zu verstehen, ist im komplexen System Unternehmen zunächst ein Bewusstsein für die Verknüpfung der Sachverhalte aktiv zu fördern.

---

[1] Unter dem UAM kann nach TOGAF verstanden werden [BHH07]: Design, Pflege, Weiterentwicklung und Zusammenschau von Geschäfts-, Anwendungs-, System- und Informationsarchitektur eines Unternehmens.
[2] Vgl. zur IT-Governance [VV04]. IT-Governance wird darin als “bridge between disciplines” verstanden, so wie hier die Unternehmensarchitektur als Mittler zwischen IT-Governance und IT-Management gesehen wird.
[3] Im Bereich der IT-Governance kommt heute vielfach das Referenzmodell COBIT, aktuell in der Version 4.1, zum Einsatz. Im IT-Management kommt heute vielfach das Referenzmodell ITIL, aktuell verfügbar in der Version 3, oder die ISO/IEC 20000 zum Einsatz. Letztere entspricht in deren Ausprägung leider noch immer der ITIL Version 2.
[4] Es existiert beispielsweise ein Mapping zwischen dem erwähnten TOGAF und COBIT [ITG07].

1.3 Zielsetzungen des Beitrags und methodisches Vorgehen

Die folgende Zielsetzung wird mit dem vorliegenden Beitrag verfolgt. Es geht um eine Konkretisierung der Zusammenhänge zwischen IT-Governance (Umsetzungsprüfung der IT- und der Unternehmensstrategie), dem IT-Management (eigentliche Umsetzung der IT) und der UA (abgeleitet aus der Unternehmensstrategie). Im vorliegenden Beitrag erfolgt die Themenbearbeitung basierend auf der Literatur und der Kenntnis bestehender IT-Governance- und –Referenzmodelle (Abbildung 1 ist durch den Vergleich der IT-Governance- und -Management-Referenzmodelle entstanden). Die Analyse der Beziehungen zwischen UA und den IT-Governance- und -Management-Instrumenten erfolgt vor dem Hintergrund, dass das UAM in Unternehmen zu selten in Konjunktion mit IT-Governance- und -Management-Instrumenten betrachtet wird. Das UAM wird leider zu oft aus der Perspektive der LE's betrieben und stiftet damit wenig Nutzen für das Geschäft.

2. Zusammenhänge zwischen UA, IT-Governance und IT-Management

Im Folgenden wird generisch auf die Zusammenhänge zwischen IT-Governance-Instrumenten und UAM eingegangen. Dabei sind verschiedene IT-Stakeholder als am UAM-Prozess Beteiligte zu sehen. Im Kern sind es die LE's und LB's und darin unterschiedliche Rollen, welche in der Bezugnahme auf das UAM zu einer gemeinsamen Sprache finden (können; vgl. zu Boundary Objects [Kü06]). Für die IT-Führung des LB stehen in Relation zum UAM und im Sinne des Business-IT-Alignments u.a. die folgenden IT-Governance-Instrumente im Vordergrund: IT-Controlling, IT-Risiko-, IT-Sicherheits- sowie IT-Portfoliomanagement. Seitens der IT- oder IT-Service-Lieferanten können u.a. folgende IT-Management-Prozesse genannt werden, die einen direkten Bezug zum UAM haben: Configuration Management, Change- und Release-Management (Vgl. zu Prozessen [VVP06]), in der ITIL V3 spielt überdies das UAM im ganzen Band Service Design eine wesentliche Rolle sowie im Band Service Operation (Funktionen IT Operations und IT Application Management). Der erwähnte UAM-Prozess erbringt die regelmäßige und rollierende Überprüfung des Ist-Architekturzustands im Hinblick auf den künftigen Soll-Zustand. Ausgehend von aktuellen und künftigen Geschäftsbedürfnissen und deren Unterstützung mit IT-Anwendungen kann eine Transformation von der Ist- zur Soll-Architektur definiert werden. Diese Transformation mündet in die zu planende Definition von Machbarkeitsstudien- (Plan), Projekt- (Build), Ist- und Soll-Anwendungs-Portfolios (Run). Auf diese Portfolios einigen sich Geschäft und IT gemeinsam. Dadurch ist ein idealer Abgleich zwischen beiden möglich (Anforderungs- und Lieferantendefinition, etc.). Die Überwachung von Zielen und finanziellen Aspekten der Portfolioentwicklungen erfolgt typischerweise durch ein IT-Controlling. Im Rahmen des IT-Portfoliomanagements gelangen durch das UAM unterstützt IT-Sicherheits- und –Risikoaspekte zur Anwendung, die laufend an die sich verändernde Ist-Landschaft in Richtung Soll-Zustand anzupassen sind [BHH07: 71].

Tatsächlich ist das UAM meist eine technisch und weniger geschäftlich getriebene Disziplin. Wünschbar ist ein bidirektionaler, mehrheitlich durchs Geschäft (Stakeholder) und seine Informationsbedürfnisse getriebener oder angestoßener UAM-Prozess, der mit den IT-Stakeholdern laufend abzustimmen ist. Die IT-Strategie ist idealerweise aus der Unternehmensstrategie abzuleiten. Die Ziele von IT-Governance und UAM lauten: Aus strategischer, taktischer und operativer Sicht zum Business-IT-Alignment beizutragen [HV93]. Das Ziel beider Aktivitätsbereiche ist es, dass Anspruchsgruppen seitens Geschäft und IT über das Management der IT-Governance und das UAM aus proaktiver und reaktiver Sicht eine adäquate gemeinsame Gestaltungs- und Arbeitsplattform finden. Dies unterstützen die verschiedenen UA-Sichten.

3 Zusammenhänge zwischen Teildisziplinen

In Abbildung 1 werden in tabellarischer Form Beziehungen zwischen den IT-Governance- und -Management-Instrumenten und dem UAM differenziert und im Hinblick auf deren Nutzen für das Business-IT-Alignment analysiert.

Teildisziplin	Bezug zum UAM	Nutzen für das Business-IT-Alignment
IT-Portfoliomanagement	• Ableitung von Portfolios aus Ist-Architektur richtung Soll-Architektur • Resultat: Machbarkeitsstudien, Projekte, Soll-Anwendungen und entsprechende Portfolios[5] • UAM-Prozess als Ableitungsvehikel • Über UAM-Prozess Realisierung von UA-Domänenstrategien für Geschäfts-, Anwendungs- und Systemarchitekturen möglich.	• Vereinfachung der Analyse von Abhängigkeiten und Zusammenhängen zwischen Portfolio-Elementen, Architektur-Modulen oder -Domänen • Ableitbarkeit von Integrationsanforderungen • Definition systematischer Interoperabilitäts- und Infrastruktur-Strategien • Priorisierbarkeit von Studien, Projekten und Anwendungen in Portfolios und deren Ausrichtung auf Geschäfts- und IT-Strategien • Vereinfachung der Mittelallokation und des IT-Controllings (Portfolio-Elemente als Controlling-Objekte) • Ressourcenallokation vereinfachend: Personell, zeitlich, finanziell auf LE- und LB-Seite.
IT-Controlling und IT Financial Management	• Architekturelemente stellen Bezugspunkte für das IT-Controlling dar • Denkbar aus statischer Sicht: Domänen einer Architektur als Bestandteile einer Kosten-Leistungs-Rechnungssicht auf der LE- oder der LB-Sicht zu verstehen • Dynamische Zuordenbarkeit der finanziellen Ressourcen zu Portfolios aus UAM als erweiternde Sicht des Kostenmanagements im IT-Bereich.	• Vereinfachung finanzieller Bewertbarkeit und Ressourcenallokation über Konkretisierung von IT-Portfolios und Architekturdarstellung • Budgetierung auf Architekturbasis und UA-Entwicklung in Zeitverlauf eruierbar • Planung auf Architekturbasis konkretisierbar und finanzielle Implikationen ableitbar • Kostenleistungsrechnung operationalisierbar • Asset Management ermöglichend • Grundlage für IT-Servicekosten-Verrechnung • Vereinfachung Zurechenbarkeit von Kosten auf LE- und LB-Seite • Business IT Alignment und Service Level Management unterstützend • Vereinfachung der Klärung von Kostenfolgen von Organisationsänderungen und deren Auswirkungen auf Anwendungs- und Systemlandschaft.
IT-Risikomanage-	• UAM als Basisinstrument eines umfassenden (IT-)Risikomanagements	• Risiken aus Geschäfts-, Anwendungs- oder Systemperspektive visualisierbar, lokalisierbar und adressierbar,

[5] Vgl. zu IT-(Service)Portfoliomanagement [VDK08: 188 ff.].

Teildisziplin	Bezug zum UAM	Nutzen für das Business-IT-Alignment
ment	• Differenzierung verschiedenster Risiken über unterschiedlichste UAM-Sichten aus Business- und IT-Sicht: Zweck differenzierte Risikoevaluation (Eruierung Assets, Schwachstellen und Bedrohungen im Rahmen von CRAMM [VVP06: 117 f.]).	ebenso Behebung von Risiken über entsprechende UAM-Maßnahmen und Mittelallokationen[6] • Vereinfachung Risikoprävention und Chancen-Analyse • Durch die Modularisierung der UA-Sichten (zur Reduktion der Komplexität) Konkretisierung, Ortung und die zusammenhängende Sicht von IT-Risiken über die verschiedenen Sichten ermöglichend • Zentrales Instrument des IT-Risikomanagements im Rahmen des Business- und IT Continuity Managements.
Interoperabilitäts-Management	• Architektonische Artefakte als Elemente des Informationssystems verstehbar: Modularisierung von Geschäfts-, Anwendungs- und Systemarchitektur • Analyse der Elementen-Beziehungen führt zu Interoperabilitätsmanagement.	• Strukturiertes Interoperabilitätsmanagement in unterschiedlichem Detaillierungsgrad möglich: Ableitung von Geschäfts-, Anwendungs- und Systemdomänen und Ableitung von Beziehungen semantischer, syntaktischer und technischer Art; Visuell dar- und unterstützbare Interoperabilitäts-Diskussion • Vereinfachte Ableitbarkeit Domänen-interner und -externer Interoperabilitätsbedarfe.
Configuration Management	• Configuration Items als kleinste Elemente des UAM, welche zu Zwecken des IT Service Managements geführt werden • Darin stellen sie die Grundlage einer ganzen Reihe von Prozessen wie Change Management, Incident Management, Problem Management, Release Management dar • Eine Verknüpfung von Configuration Items mit der UAM ist in einer zu definierenden Weise erforderlich.	• Ableitbarkeit von Changes und Releases aus UAM, entsprechenden genannten Portfolios und finanziellen Konsequenzen auf der effektiven Implementierungsbasis • Problems und Incidents können Hinweise auf Adaption der Unternehmensarchitektur geben • Problems und Incidents können mit Portfolios von Machbarkeitsstudien, Projekten und Ist- sowie Soll-Anwendungen abgestimmt werden • Sicherstellung des Austauschs und Alignments zwischen UAM und Configuration Management.
Change Management und Release Management	• Ableitungen von Soll-Architekturen aus dem UAM haben immer Veränderungen der Ist-Architektur und -Infrastruktur zur Folge • Damit werden Changes auf der IT-Management-Ebene abgeleitet, über welche ein prozedural determiniertes Vorgehen für die umfangreiche und systematische Veränderung der Konfiguration resultiert • Change Management Prozess als Planungs, Bewilligungs- und Überwachungsprozess des Release Managements als effektiver Umsetzungsprozesse.	• Systematisches Change Management unterstützend • Ideale Basis bietend für Releases und darin z.B. des Testings etwa von Interoperabilitätsabhängigkeiten, etc. • Releases können mit IT-Portfoliomanagement und mit Soll-Architektur im UAM orchestriert werden; Changes und Releasewechsel erfolgen somit klar in einem Abhängigkeitsverhältnis zum UAM • Releasemanagement wird durch UAM gesteuert • Lead des UAM für das Change und Release Management.

Abbildung 1: IT-Governance- und -Management-Teildisziplinen und deren Nutzen für das Busines-IT-Alignment.

4 Zusammenfassung und Ausblick

Das UAM stellt ein zentrales Bindeglied zwischen Geschäft/LB und IT/LE sowie zwischen den verschiedenen IT-Governance- und IT-Management-Rollen und -Instrumenten dar. Diese weisen weitreichende Bezüge auf. Dies ist allen Beteiligten entsprechend zu kommunizieren.

[6] Dies trifft z.B. zu für redundant ausgelegte Standorte im Rahmen von Kontinuitätsstrategien und Wiederherstellungsplänen im Business- und IT Service Continuity Management (Vgl. [BHJ06: 196 ff.]). Vgl. hierzu auch die Aussagen zum Thema IT-Sicherheit.

Die aufgearbeiteten Bezüge sind so weit möglich zu stärken und Abhängigkeiten dahingehend auszuarbeiten, dass es auf Basis des UAM zu einer integrierten Sicht, Führung und einem entsprechenden IT-Management kommt. Das UAM wird heute in Unternehmen selten in Relation zu IT-Governance- und -management-Instrumenten und entsprechenden Referenzmodellen gesehen und auch der UAM-Prozess ist vielfach unterbewertet. Die Maturität von Unternehmen ist, was Verständnis, Nutzen und Zusammenhänge zwischen IT-Governance-, -Management-Instrumenten sowie UAM betrifft, gering. Dies ist bei der IT-Governance und beim IT-Management in den Vordergrund zu rücken, da Ar-chitekturen Boundary Objects zwischen Geschäft und IT darstellen. Konsistenz und integrale Sicht auf die IT-Führung werden durch das UAM gefördert. Die Vernetzung teilweise disparater IT-Governance-Instrumente, des IT-Managements und des UAM ist eine dringlich anzugehende neue Disziplin im Bereich der Wirtschaftsinformatik. Die Gefahr entsprechender Silo-Bestrebungen insbesondere im IT-Governance-Bereich ist zu eliminieren. Die Zusammenführung von IT-Governance, IT-Management und UAM ist in der Forschung weiter zu forcieren, damit Unternehmen angesichts der IT-Herausforderungen neuen Durchblick gewinnen. Ferner ist ausgehend von den hier gemachten Aussagen die Untersuchung der IT-Governance- und -Management-Rahmenwerke im Hinblick, auf das UAM zu vertiefen.

Literaturverzeichnis

[BHJ06] Bartlett, J.; Hinley, D.; Johnson, B.; Johnston, D.; Keeling, C.; Lloyd, V.; McDonald, I.; Mather, J. (2006): Service Delivery, OGC/TSO, London.

[BHH07] Blevins, T.; Harrison, R.; Homan, P.; Josey, A.; Rouse, M.F. (2007): TOGAF Version 8.1.1 Enterprise Edition – A Pocket Guide, The Open Group/Van Haren Publishing, Zaltbommel.

[DV04] De Haes, S.; Van Grembergen, W. (2004): IT Governance and Its Mechanisms, Information Systems Control Journal 1 (2004).

[HV93] Henderson, J.C.; Venkatraman, N. (1993): Strategic Alignment – Leveraging Information Technology For Transforming Organizations, in: IBM Systems Journal 32 (1993) 1, S. 472-484 sowie auf URL: http://fag.grm.hia.no/ikt4100/2001951.pdf (Aufruf per 2009-06-18; erstellt 1993).

[ITG06] ITGI (2006): Enterprise Value – Governance of IT Investments – The ValIT Framework.

[ITG07] ITGI (2007): TOGAF and COBIT – Mapping of TOGAF 8.1 with COBIT 4.0 – Part I, ITGI, Rolling Meadows.

[Kü06] Kühn, A. (2006): Boundary Objects for E-Government, auf URL: http://wirtschaft.bfh.ch/uploads/tx_frppublikationen/Kuehn_Andreas.pdf (Aufruf per 2009-09-02; erstellt per 2006-04-30).

[Kr04] Krcmar, H. (2004): Informationsmanagement, Springer, Berlin et al.

[La08] Lallana, E.C. (2008): e-Government Interoperability, URL: http://unpan1.un.org/intradoc/groups/public/documents/UN-OTHER/UNPAN032094.pdf (Aufruf per 2009-08-14; erstellt per 2008).

[Ni05] Niemann, K.D. (2005): Von der Unternehmensarchitektur zur IT-Governance, Vieweg, Braunschweig.

[VDK08] Van Bon, J.; De Jong, A.; Kolthof, A.; Pieper, M.; Tjassing, R.; Van der Veen, A.; Verheijen, T. (2008): Foundations in IT Service Management basierend auf ITIL V3, ITSMF-Library, Van Haren Publishing, Zaltbommel.

[VVP06] Van Bon, J.; Van der Veen, A.; Pieper, M. (2006): Foundations in IT Service Management basierend auf ITIL, Van Haren Publishing, Zaltbommel.

[VV04] Van Kersberg, K.; Van Waarden, F. (2004): ‚Governance' as a bridge between disciplines: Cross-disciplinary inspiration regarding shifts in governance and problems of governability, accountability, and legitimacy, in: European Journal of Political Research 43 (2004), S. 143-171.

[WW02] Weill, P.; Woodham, R. (2002): Don't just lead, govern: Implementing Effective Governance (Working Paper 236), Sloan School of Management, Cambridge.

The Risk-Aware Enterprise Architecture: Towards a Transparent Inventory of IT Risk Management Artifacts

Manfred Pauli, Michael Schermann, Helmut Krcmar

iteratec GmbH
Inselkammerstraße 4
82008 München
manfred.pauli@iteratec.de

Chair for Information Systems
Technische Universität München
Boltzmannstraße 3
85748 Garching
{michael.schermann|krcmar}in.tum.de

Abstract: A risk inventory provides an integrated view on risk management artifacts, e.g., risks, risk controls, and performance indicators. In this paper, we show how adapting the enterprise architecture management processes (EAM) may provide a foundation for an integrated IT risk inventory. Based on a design research approach, we develop a systematic approach for integrating the disciplines of risk management and enterprise architecture management. We demonstrate the utility of our approach by evaluating an identity management solution in a large bank.

1 Introduction

Although risk management is commonly named as one of the top challenges in information management, risk managers and CISOs struggle with establishing and maintaining transparency over sources for risks, implemented risk controls, and their effectiveness [Pa07, SS08]. In particular, little data is available for forecasting the future impact and probability of an *IT risk* compared to other risks, e.g. in the insurance industry [SS08]. Thus, risk management commonly relies on a very laborious process to extract experts' estimations, which often results in risk assessments of uncertain quality. However, organizations may tap a source of relatively objective data: their enterprise architecture as the formal description of an organizations' IT. [TOG07]. Hence, our research question is: *What are the benefits of integrating risk management and enterprise architecture management?*

The paper follows the design research methodology as suggested by [He04]. Hence, we first argue that risk management lacks of high quality data and that enterprise architecture management (EAM) may provide this foundation. Subsequently, we present a process model that shows how to integrate the disciplines of EAM and IT risk management. We demonstrate the utility of our approach by evaluating an identity management solution in a large bank. The paper closes with a critical appraisal of our approach and points out future avenues of worthwhile research.

2 The Enterprise Architecture as a Foundation for Decision-Making in IT Risk Management

The IT risk management process generally consists of identifying and analyzing risks, plan and implement risk responses and risk monitoring [La06]. There is a plethora of techniques to manage IT risks. When assessing risks, the risk management team has to look at every risk and estimate its probability and its potential impact. The team relies on the experiences of the team member or other expert's estimations. However, experience and estimations can be very erroneous [Sc03]. Furthermore, information on the effectiveness of risk controls is required to justify investments and monitoring risk management efforts.

Risk indicators …

One way to handle this problem is using risk indicators [SS08, He07]. A risk indicator is a variable which has a causal relation to one or more risks [He07]. Usually, a risk indicator exhibits a tolerance levels as well as rules for exception handling [Wi08]. Both, researchers and practitioners, argue that risk indicators will play an important role in risk management, since they provide direct information on the current risk situation as well as on the effectiveness of the risk management process and [Wi04, Ja07].

… and where to find them

Since the objective of enterprise architecture managements (EAM) is to document, plan, control, and monitor the strategic and operational aspects of an organizations' enterprise architecture, we argue that both causes and effects of IT risks manifest in the components of an enterprise architecture.

Commonly, enterprise architectures (EA) follow a hierarchical multi-layer approach [Ha09]. The upmost layer describes the business processes within an organization. The next layer contains the application systems. The following layer specifies the components of the application systems: the technical architecture and infrastructural elements. Projects are cross-cutting concerns across all three layers. Hence, the EA illustrates the relationships between business objectives, business processes, and the components of deployed and planned application systems. Commonly, each element of an EA is called building block and is of a particular building block type.

Furthermore, common EAM methodologies revolve around identifying and maintaining useful indicators to represent the important aspects of enterprise architecture. We argue that these indicators may serve as foundation of developing risk indicators. Hence, EAM provides a high quality data base for risk analysis and justifying decisions in risk management.

3 Towards a Risk-Aware Enterprise Architecture

A typical EAM process begins with the documentation of the existing IT infrastructure, the business processes and their relationships. The next step is to characterize the elements of the IT infrastructure by representing them with different attributes and assigning current and planned values. As soon as a transparent view on the IT infrastructure exists, planning the target shape of the IT infrastructure begins. Finally, the steps of the plan will be addressed by specific projects. Since business requirements will change over time, EAM is an iterative task.

Figure 1 shows our process model to tap EAM as a specific information provider for IT risk management. As discussed above, the first step (1) of the IT risk management process is to identify IT risks resulting in a list of IT risks. Then, each risk is represented as a set of risk indicators (2). However, the values of such risk indicators must relate to building blocks of the enterprise architecture (3). For instance, password security metrics apply to systems with access control only. Subsequently, risk managers deduce a meta-model of identified IT risks, risk indicators and associated building blocks. Next, the risk indicators are being associated with values from the EA (4). For instance, the availability of system A is 95 percent whereas system B has 98 percent availability. Subsequently, each building block should be analyzed whether it may provide information on a particular risk indicator. Since there will be a lot of values to a single risk indicator, they have to be extracted (5) and presented to the risk management experts for aggregation and focus (6). Finally, risk controls are being reflected by setting up particular project in the process of EAM (7). Again, it is possible to attach specific indicators that provide information on the performance and goal congruence of the decisions.

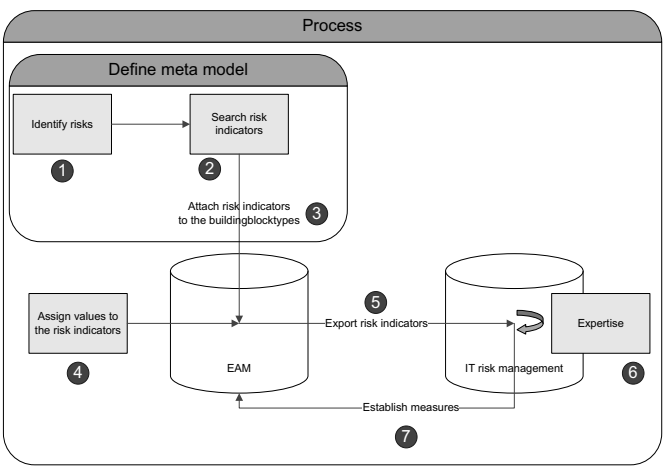


Figure 1: Integrating EAM with IT Risk Management

The presented process provides a generic integration of EAM and IT risk management. In the following, we demonstrate the utility of our approach by evaluating an identity management solution in a large bank.

4 Case Study: Does identity management improve the risk situation?

The following example illustrates the above presented process. We conducted a case study with the IT infrastructure of the bank 'RiskBank[1]' provided by an EAM tool provider. The EA of the RiskBank represents a large IT infrastructure with more than one hundred different application systems. Usually, new risk controls are recommended by specialized application providers or consulting companies. For instance, the fictive company IdentitySolutions offers the CISO of RiskBank a new identity management system called IMS. The system claims to reduce the risk of identity theft resulting from the complex IT landscape [Al08]. IMS consists of training courses for the staff, single-sign-on functionality, password self management, user account management, authorization management and user activity logging.

In the following, we evaluate IMS regarding its effects on the IT infrastructure and the risk situation of RiskBank. As discussed above, the first step is to identify IT risks. The heading of Table 1 shows an extract of the list of considered risks, following the pattern "risk category:risk origin". The categories follow the guidance of Basel II [Ba03]. After having identified the IT risks within the company, the IT risk manager attaches risk indicators to the different risks. One indicator can also be applied to different IT risks. Table 1 shows the assignment of risk indicators to the identified IT risks. The next step is to link the risk indicators to building blocks.

Table 1: Extract of RiskBank's risk indicators [adapted from Wi04, KPI08]

Risk	**Security:Staff**	**Security:Syst**	**Secur Influe**	**Security:F**	
Risk indicators	# Authorized employees	# Password recoveries p employee p	# Criminal intrusions per	# Parti person accour	Logging employe activitie
	# Password loss per employee pe	Password	-	Average tim create an acc	-
	# Security traini	Average tim	-	System h separate	-

Each of the risk indicators listed in Table 1 is attached to building blocks of the type 'application system'. The risk indicators "# Authorized employees" and "# Criminal intrusions per year" are also attached to the infrastructural elements. This step results in a meta-model that contains IT risks, risk indicators, building block types and the relationships among them.

[1] RiskBank is a pseudonym to provide anonymity.

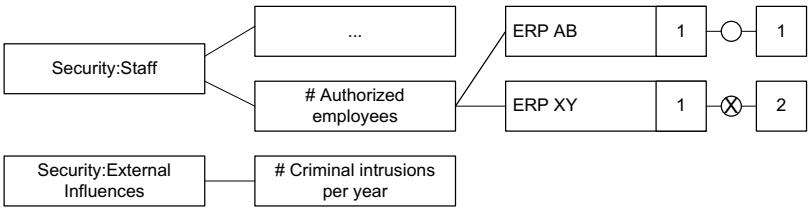


Figure 2: Number of outliers per risk indicator

The next step is to instantiate this meta-model by assigning values to the risk indicators. To do so, the risk management team goes through the enterprise architecture and assigns values from different building blocks (see Figure 2). For example, the indicator '# Authorized employees' for 'ERP XY' has the value '2', i.e. more than system owner that has full access control exist (see Figure 3). The sources of the values can vary. The most confident sources are any results of logging, pinging, and similar actions. The result is a comprehensive and detailed representation of the IT risk situation.

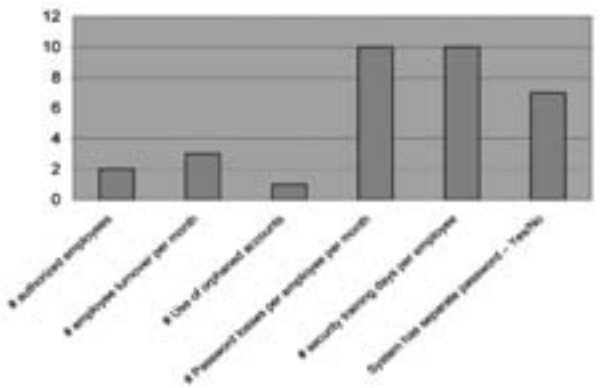


Figure 3: Number of outliers per risk indicator

However, preparing decision making requires aggregating the indicators. At 'RiskBank', we summed up the number of outliers per risk indicator. For instance, we found that the indicator "# Password losses per employee per month" was beyond its tolerance level. In particular, the underlying building blocks were crucial systems at 'RiskBank'. However, these systems were used for particular accounting purposes only and were commonly with separated access control ("System has separate password"). Hence, employees tended to forget the passwords. Thus, the functionality of 'single-sign-on' of the IMS proposal would provide appropriate means to control the associated risks. Although this example is of simplified nature, it shows the benefits of integrating IT risk management with enterprise architecture management. For instance, the IMS could be instantiated as a project in the enterprise architecture and the risk indicators would serve as key performance indicators of the project. Hence, risk managers are able to actually show the benefits of risk controls.

5 Conclusion & Outlook

This paper argues for the enterprise architecture as a solid and reliable source of information for IT risk management. We suggest risk indicators as an approach to establish an integrated information source. The results of the case study show that quantitatively grounded decision-making in risk managements does not require estimating probabilities and potential impacts. The demonstration shows that our approach focuses on the potential effects of risk controls and thus enable risk managers and decision makers critically appraising new trends, products, and services from the IT security industry. In sum, this paper presents an effective approach for grounding decision-making in IT risk management in available data [SS08]. Future research focuses on enhancing the visual representation of risk indicators [Te99].

References

[Pa07] Parker, D.B., Risks of risk-based security. Communications of the ACM, 2007. 50(3): p. 120.

[SS08] Shostack, A. and A. Stewart, The New School of Information Security. 2008, Upper Saddle River, NJ, USA: Addision-Wesley.

[TOG07] The Open Group, The Open Group Architecture Framework (TOGAF). 2007, The Open Group,: San Francisco, CA, USA.

[He04] Hevner, A.R., et al., Design Science in Information Systems Research. MIS Quarterly, 2004. 28(1): p. 77-105.

[La06] Landoll, D.J., The security risk assessment handbook: a complete guide for performing security risk assessments. 2006, Boca Raton, FL, USA: Auerbach Publications.

[Sc03] Schneier, B., Beyond Fear: Thinking Sensibly About Security in an Uncertain World. 2003, Berlin: Springer.

[He07] Herrmann, D.S., Complete Guide to Security and Privacy Metrics: Measuring Regulatory Compliance, Operational Resilience, and ROI. 2007, Boca Raton, FL, USA: Auerbach Publications.

[Wi04] Witty, R.J., K. Brittain, and A. Allan, Justify Identity Management Investment With Metrics. 2004, Gartner, Inc.: Stamford, CT, USA.

[Ja07] Jaquith, A., Security Metrics: Replacing Fear, Uncertainty, and Doubt. 2007, Upper Saddle River, NJ, USA: Addison-Wesley.

[Ha09] Hanschke, I., Strategisches Management der IT-Landschaft: Ein praktischer Leitfaden für das Enterprise Architecture Management. 2009, München: Hanser.

[Al08] Allan, A., Identity and Access Management Technologies Defined. 2008, Gartner, Inc.: Stamford, CT, USA.

[Ba03] Basel Committee on Banking Supervision, Sound Practices for the Management and Supervision of Operational Risk. 2003, Bank for International Settlements: Basel, Switzerland.

[KPI08]. KPI Library. Information Technology. 2008 [cited October 06, 2008]; Available from: http://kpilibrary.com/category/itman/.

[Te99] Tegarden, D.P., Business information visualization. Communications of the AIS, 1999. 1(1): p. 1-38.

Umgang mit Risiken für IT-Dienste im Hochschulumfeld am Beispiel des Münchner Wissenschaftsnetzes

Silvia Knittl
Technische Universität München
knittl@tum.de

Wolfgang Hommel
Leibniz-Rechenzentrum
hommel@lrz.de

Abstract:
Die steigenden Benutzererwartungen und Kosten zur Erbringung der komplexen IT-Dienste, die zur technischen Umsetzung und Unterstützung der Geschäftsprozesse mit hoher Servicequalität erforderlich sind, machen Ansätze wie Infrastructure, Platform bzw. Software as a Service auch für Hochschulen zunehmend attraktiv. Im Zusammenspiel zwischen Hochschulverwaltungen, zentralen Einrichtungen, Fachbereichen und Hochschulrechenzentren wurden entsprechende Dienstleistungsansätze bereits lange genutzt, bevor sie unter dem Stichwort Cloud Computing neu aufgegriffen wurden. Auf Basis des Cloud Computings sind jedoch jüngst neue Methoden und Werkzeuge entstanden, um diese Formen der Diensterbringung systematisch auf ihre Risiken hin zu analysieren; auf dieser Basis können fundierte, risikogetriebene Entscheidungen darüber getroffen werden, welche Dienste lokal erbracht bzw. zu welchen hochschulinternen oder auch externen Dienstleistern sie ausgelagert werden können. In diesem Artikel stellen wir unsere Erfahrungen mit Hochschuldiensten in diesen drei Kategorien und die mit ihnen verbundenen Risiken am Beispiel des Münchner Wissenschaftsnetzes vor. Wir stellen sie mit den an vielen europäischen Hochschulen zu beobachtenden Entwicklungen gegenüber und geben einen Ausblick auf die sich abzeichnenden Herausforderungen und Möglichkeiten in diesem Bereich.

1 Motivation

Die strategische Ausrichtung der IT-Dienste an den Kern-Geschäftsprozessen nimmt seit einigen Jahren auch auf die Hochschulinfrastrukturen massiven Einfluss. Hochgradig integrierte Campus-Management-Systeme, Groupware-Lösungen und die nahtlose Integration mobiler Endgeräte dominieren die aktuelle Weiterentwicklung der technischen IuK-Infrastrukturen und weisen an vielen Stellen Parallelen zu Entwicklungen in privatwirtschaftlichen Unternehmen auf. Mit der Nutzung der neuen technischen Möglichkeiten steigt jedoch zwangsweise auch die Komplexität der Dienste und Komponenten. Diese wiederum führt zu einer Reihe von Herausforderungen und Risiken, die sich von der Betreibbarkeit mit den vorhandenen Personalressourcen bis hin zur erforderlichen Hardwareredundanz zur Sicherstellung der prozesszielkritischen Systeme erstreckt.

Die Verlagerung des Betriebs von IT-Diensten unter Beibehaltung der vollständigen Prozesskontrolle wird deshalb auch für viele Hochschulen nicht nur finanziell attraktiv, sondern mehr und mehr auch zwingend erforderlich, um die benötigten Dienste flexibel und in der notwendigen Servicequalität zur Verfügung stellen zu können. Im Hochschulum-

feld ist – im Unterschied zu vielen privatwirtschaftlichen Szenarien – häufig eine Dreiteilung zu beobachten: Zunächst werden IT-Systeme, die die Kernaufgaben der Hochschulverwaltung direkt unterstützen, beispielsweise Studenten-, Personal-, Lehrveranstaltungs- und Prüfungsverwaltungssysteme, überwiegend von einer EDV-Abteilung der Hochschulverwaltung selbst betrieben. Ferner werden klassische netzbasierte, aber nicht zwingend hochschulspezifische Dienste wie WLAN, VPN, Mailserver, Fileserver und Arbeitsplatzsysteme vom jeweiligen Hochschulrechenzentrum bereitgestellt. Eine Kopplung der Systeme findet dabei in der Regel mindestens auf Ebene des Identity-Managements statt, um die Account-Erstellung und -Terminierung mit den Prozessen der Personal- und Studentenverwaltung zu koppeln. Schließlich existieren eine Reihe von Diensten, die bei externen Dritten in Auftrag gegeben werden. Dabei handelt es sich einerseits um Dienste, die nur von einem zu geringen Teil aller Hochschulmitglieder genutzt werden, so dass eine zentrale, hochschulweite Bereitstellung durch das Rechenzentrum nicht ökonomisch sinnvoll wäre; andererseits können für Standarddienste auch die erzielbaren Kosteneinsparungen ausschlaggebend sein, wobei sich durch den externen Ort der Datenspeicherung und -verarbeitung neue, zu beachtende Risiken ergeben.

In diesem Artikel stellen wir den Umgang mit ausgewählten Risiken für essenzielle Hochschul-IT-Dienste am Beispiel des Münchner Wissenschaftsnetzes (MWN) vor. Wir differenzieren dabei in Anlehnung an aktuelle Diskussionen unter dem Stichwort *Cloud Computing* zwischen den Kategorien *Infrastructure*, *Platform* und *Software as a Service* (IaaS, PaaS und SaaS), um eine Einordnung und Risikoeinschätzung der Hochschul-IT-Dienste vorzunehmen [ENI09]. Wir stellen deshalb im nächsten Abschnitt zunächst das MWN im Allgemeinen vor und gehen im Anschluss auf die drei Dienstkategorien und den Umgang mit den jeweiligen Risiken ein. In Abschnitt 3 stellen wir diese konkreten Ergebnisse den allgemeinen Entwicklungen an deutschen und internationalen Hochschulen gegenüber und schließen mit einem Ausblick auf die sich aktuell abzeichnenden Möglichkeiten und Herausforderungen in diesem Bereich.

2 Dienste im MWN und deren Risikomanagement

In diesem Abschnitt werden typische IT-Dienste im Hochschulumfeld am Beispiel des Münchner Wissenschaftsnetzes dargestellt. Es wird gezeigt, dass es durchaus üblich ist, Dienste auch an externe Dienstleister auszulagern – teils sogar im Sinne eines „Cloud Computings“. Risiken, die durch die steigende Komplexität beim Outsourcing entstehen, müssen entsprechend offensiv gemanagt werden. Unsere Ansätze hierfür werden ebenfalls in diesem Abschnitt dargelegt.

Das MWN ist das technische Rückgrat der Vernetzung der Hochschulen im Münchner Umkreis und wird vom Leibniz-Rechenzentrum der Bayerischen Akademie der Wissenschaften (LRZ) betrieben. Neben diesen Netzdiensten stellt das LRZ als gemeinsamer Dienstleister der Münchner Hochschulen auch weitere Services bereit. Abbildung 1 zeigt einen kleinen Ausschnitt typischer Dienste, welche die Mitglieder der Hochschulen in ihrer täglichen Arbeit verwenden.

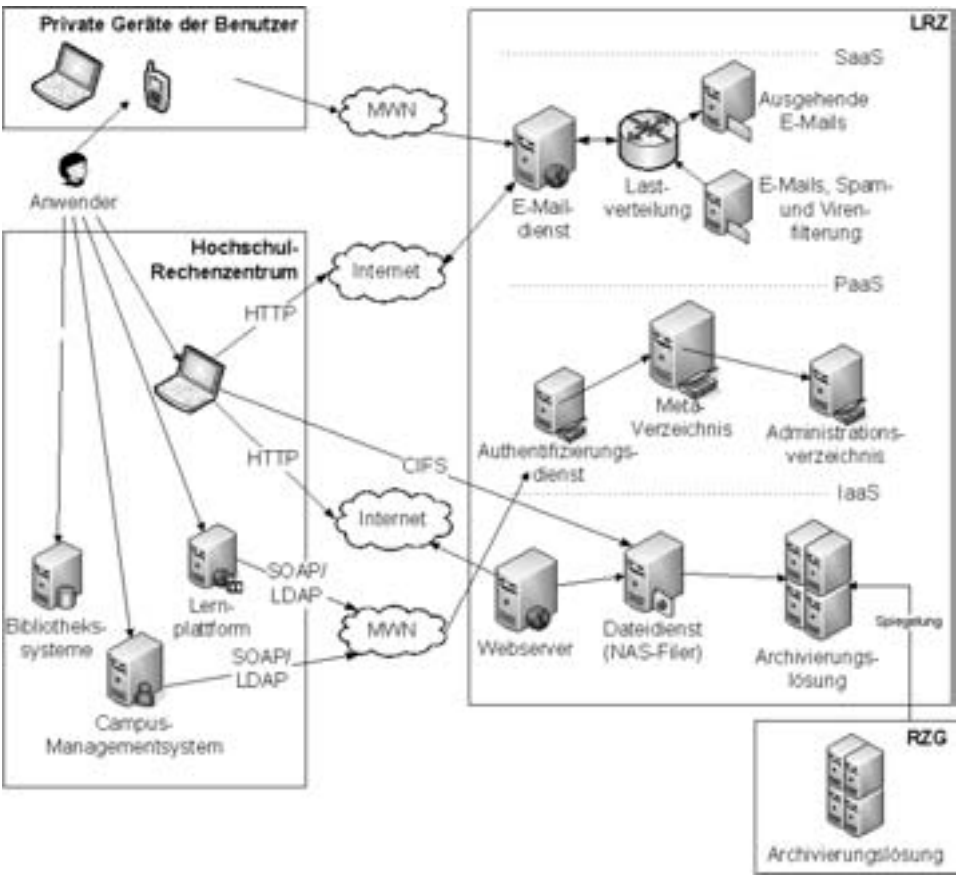


Abbildung 1: Auszug verwendeter Dienste der Benutzer im Münchner Wissenschaftsnetz

Die vom LRZ bereitgestellten Dienste in der Abbildung sind auf unterster Ebene dem Bereich IaaS zuzuordnen. Hierzu zählen flexibel zugängliche Speicherbereiche für jeden Hochschulbenutzer im Sinne eines „Storage as a Service“, aber auch Netzinfrastrukturkomponenten wie die WLAN-Zugänge. Zur Sicherstellung eines kontinuierlichen Betriebs – auch im Falle größerer Störungen – werden die Daten des Speicherdienstes zusätzlich an einen weiteren externen Dienstleister ausgelagert [HKP09]. Die nächste Ebene zeigt vereinfacht die Identity- und Access-Management-Infrastruktur (IAM), welche bedarfsgerecht in hochschulinterne Anwendungen integriert werden kann. Auf oberster Ebene stehen den Hochschulen als SaaS zu bezeichnende Dienste wie E-Mail-Lösungen oder virtuelle Webserver zur Verfügung. Die jeweiligen Grenzen zwischen IaaS, PaaS und SaaS sind dabei nicht immer scharf definiert und unterliegen der Dynamik der Weiterentwicklung des Dienstportfolios. Neben den vom LRZ zur Verfügung gestellten Diensten betreiben die Hochschulen dedizierte Lösungen für den Bereich der Studentenverwaltung (Campus-Managementsysteme) und Bibliothekssysteme. Zunehmend erwarten die Benut-

zer jedoch auch die Möglichkeit der nahtlosen Integration eigener (mobiler) Geräte wie etwa Smartphones oder Laptops in ihre Arbeitsumgebung.

Ein strukturiertes Vorgehen im Risikomanagement einer von zunehmender Komplexität gekennzeichneten IT-Infrastruktur ist hierbei der wesentliche Erfolgsfaktor. Unsere Aktivitäten folgen dem in [SGF+04] beschriebenen Vorgehen und der von ENISA (European Network and Information Security Agency) eingeführten Risikokategorisierung in organisationsbedingte, technische und allgemeine Risiken [ENI09]. Die allgemein notwendigen Schritte sind die Charakterisierung der Systeme, Identifikation von Schwachstellen und Gefährdungspotential, Bewertung der Wahrscheinlichkeit des Auftretens und die resultierenden Auswirkungen, um damit schließlich Empfehlungen zur Steuerung und Kontrolle von Risiken zu geben. Unser Vorgehen bei der Durchführung der Risikoanalyse haben wir in [Kni10] ausführlich dargelegt. Als Grundlage der Identifikation möglichst aller potentiellen Risiken bei der Verwendung von Cloud-Diensten dienten die im ENISA-Bericht aufgelisteten Risiken [ENI09]. Je höher man sich auf der Dienstestruktur (Infrastruktur–Plattform–Software) bewegt, umso mehr verlagern sich die notwendigen Risikosteuerungsaktivitäten vom Kunden bzw. Benutzer zum Provider. So stellen IaaS-Dienstleister lediglich die physische Infrastruktur und deren Grundsicherungsmechanismen zur Verfügung, wie etwa Strom, Kühlung, Firewall, während es den Kunden überlassen bleibt, für die Zugangssysteme oder das Patch-Management zu sorgen. Bei SaaS jedoch ist es auch die Aufgabe der Dienstleister, u.a. für das Patch-Management zu sorgen. Private Geräte stellen eine Besonderheit dar, da deren Management i.d.R. komplett vom Benutzer selbst übernommen wird. Tabelle 1 zeigt Auszüge der wesentlichen Ergebnisse unserer Risikoanalyse. In der linken Spalte werden die identifizierten Schwachstellen und Gefährdungen, in der rechten Spalte unsere gewählten Risikosteuerungsstrategien beschrieben.

Organisationsbedingte Risiken	**Strategie und Maßnahmen**
Obige Infrastruktur zeigt interorganisationale Abhängigkeiten. Diese sind im Falle von Störungen schwierig zu erfassen, was eine schnelle Störungsbehebung behindert.	Eine Reduzierung des Risikos kann hierbei durch eine bessere Transparenz mittels Methoden insbesondere der Softwarekartographie erfolgen.
Eine Lock-in-Situation ergibt sich durch den Einsatz dienstleisterspezifischer Entwicklungen anstelle von Standardlösungen. In unserem Fall betrifft das die IAM-Plattform. Ein Wechsel zu einem anderen Dienstleister würde mit extremem Aufwand verbunden sein.	Dieses Risiko wurde speziell im Fall des LRZs bewusst eingegangen, nachdem es die übergreifende Governance-Struktur den Hochschulen erlaubt, Einfluss auf die Leitung durch Beteiligung am Direktorium des LRZs zu nehmen.
Der Einsatz privater Endgeräte birgt Sicherheitsrisiken, da diese nicht in der Hoheit des eigenen IT-Managements sind, u.a. durch ungesicherten Zugriff auf Daten oder wenn die Geräte aufgrund mangelnder Wartung technisch nicht auf dem neuesten Stand sind.	Risikominimierung erfolgt durch Beschränkung der Zugangswege. Zugriff auf E-Maildienste mit privaten mobilen Geräten ist nur via Zertifikaten und aktivierter Datenverkehrsverschlüsselung möglich. Hierdurch werden die Vertraulichkeit, Authentizität und Integrität von Daten adäquat garantiert.

Technische Risiken	**Strategie und Maßnahmen**
Das Risiko der Überbuchung von Ressourcen kann in unserem Beispiel beim Speicherdienst auftreten, welcher von den Angehörigen der Münchner Universitäten genutzt werden kann. Ursache hierfür kann ein schlechtes Kapazitätsmanagement sein.	Splitten der Verantwortlichkeiten und des Risikos ermöglichen ein bedarfsgerechtes Kapazitätsmanagement. Speicherverantwortliche in den Hochschulorganisationseinheiten erhalten vorkontigentierte Speicherbereiche zur flexiblen Zuteilung; die globale Verantwortung des Kapazitätsmanagements liegt weiterhin beim LRZ.
Eine mangelnde Ressourcenisolation von Komponenten verschiedener Kunden, die in der gleichen Dienstleisterumgebung betrieben werden, kann zu Fehlern oder Angriffen führen. Dieses Risiko wird häufig auch in Verbindung mit virtualisierten Ressourcen gebracht.	Das Auftreten dieses Risikos wird aufgrund des Kooperationscharakters im MWN-Umfeld in unserer Umgebung als niedrig erachtet, weshalb es neben üblichen Sicherheitsverfahren keine weiteren Maßnahmen hierfür gibt.
Allgemeine Risiken	**Strategie und Maßnahmen**
Allgemeine Risiken ergeben sich durch die zunehmende Abhängigkeit von Internettechnologien. Netzstörungen können zu erheblichen Beeinträchtigungen des Normalbetriebes führen, nicht zuletzt, da die Hochschulen durch die Einführung digitaler Verwaltungsprozesse immer mehr vom Internet abhängig sind.	Aufgrund der enormen Abhängigkeit von Netzdiensten wird im Hochschulumfeld auf eine professionell gemanagte Dienstleistung wert gelegt. Diese stellt das LRZ als Netzprovider der Münchner Hochschulen durch umfassende eigene Risikomaßnahmen, wie etwa redundante Anbindung der wichtigsten Komponenten und Standorte sicher.
Der Diebstahl von Equipment ist ein Risiko, welches in jedem Unternehmen auftreten kann.	Maßnahmen sind restriktive Zugangsmechanismen des Dienstleisters durch den Betrieb in einem so genanntes „Dark Center“. Verlust persönlicher Daten kann durch Vermeidung lokaler Speichermöglichkeiten und dem Einsatz oben dargestellter Speicherlösung verhindert werden.

Tabelle 1: Auszug von Risiken bei der Verwendung von Cloud-Diensten im MWN und durchgeführte Risikosteuerungsstrategien [Kni10]

3 Ausblick

Die Auslagerung von IT-Dienstleistungen, auch in Form von Cloud Computing, eröffnet Unternehmen wie auch Hochschulen ein breites Feld an Möglichkeiten zur Flexibilisierung und Kosteneinsparung. In diesem Beitrag haben wir am Beispiel des Münchner Wis-

senschaftsnetzes, einem Verbund der Münchner Hochschulen mit dem LRZ als zentralem IT-Dienstleister, dargestellt, welche Risiken komplexe IT-Umgebungen bergen und wie sie durch ein umfassendes Risikomanagement antizipiert werden können.

Zukünftig nach [Cla08] zu erwartende Entwicklungen, wie etwa die weitere Verwendung (privater) mobiler Endgeräte, die steigende Vernetzung von Anwendungen durch Web 2.0 Technologien, die damit verbundene Zunahme von zugehörigen Content und daraus ggf. resultierende Notwendigkeit der digitalen (Langzeit-)Archivierung, die Konkurrenz durch freie Anwendungen wie Google, Wikipedia oder Disney [Kat10] erfordern, dass ein erfolgreiches wissenschaftliches Unternehmen im digitalen Zeitalter eine Versorgungsstruktur entwickeln muss, welche Flexibilität als Schlüsselmerkmal versteht. Diese Trends und auch das verstärkte Auslagern von Diensten in Richtung Cloud führen zu einer Verwässerung der traditionellen institutionellen Infrastrukturen und werden leichtgewichtigere Architekturen hervorrufen, bei denen die zunehmend mobileren akademischen Angehörigen einfach ihre Umgebung selbst zusammenstellen oder herunterladen können sollen wie in den so genannten „App Stores“ mancher Mobilfunkgerätehersteller. Eine solche Umgebung erlaubt es den Hochschulen, sich auf ihre Kernkompetenzen Forschung und Lehre zu fokussieren. Diese Entwicklungen benötigen jedoch auch eine stärkere Konzentration des Hochschulmanagements auf Governance-, Risiko- und Compliance-Themen, speziell da die im universitären Bereich notwendige offene Umgebung im Konflikt mit allgemeinen Sicherheitsansprüchen steht. So wird sich mit zunehmendem IT-Outsourcing die Rolle des CIOs verändern, indem die Gestaltung von Vertragsverhältnisse mit formalen und informellen Dienstleistern eine steigende Bedeutung bekommt.

Literatur

[Cla08] M. J. Clark. The CIO world of Higher Education in 2015. In *14th International Conference of European University Information Systems (EUNIS 2008)*, Arhus, Denmark, Juni 2008.

[ENI09] European Network and information Security Agency ENISA. Cloud Computing - Benefits, risks and recommendations for information security. Verfügbar online unter http://www.enisa.europa.eu , Zugriff am 12.01.2010, November 2009.

[HKP09] W. Hommel, S. Knittl und D. Pluta. Availability and Continuity Management at Technische Universität München and the Leibniz Supercomputing Centre. In *15th International Conference of European University Information Systems (EUNIS 2009)*, Santiago de Compostela, Spanien, Juni 2009.

[Kat10] Richard N. Katz. Scholars, Scholarship, and the Scholarly Enterprise in the Digital Age. *EDUCAUSE Review*, 45(2):44–56, 2010.

[Kni10] Silvia Knittl. Addressing Risk Management Efforts for Cloud Services at the Technische Universität München. In *6th International Conference of European University Information Systems (EUNIS 2010)*, Warschau, Polen, Juni 2010.

[SGF+04] Gary. Stoneburner, Alice. Goguen, Alexis. Feringa, National Institute of Standards und Technology (U.S.). *Risk management guide for information technology systems : Recommendations of the National Institute of Standards and Technology*. U.S. Dept. of Commerce, National Institute of Standards and Technology, Gaithersburg, 2004.

Workshop

Service Engineer, multidisziplinärer Allrounder oder Erfüllungsgehilfe auf Abruf? Der „Chief Social Media Officer“

Karsten Wendland, Joachim Heinz, Frank Wolf

Service Engineer, multidisziplinärer Allrounder oder Erfüllungsgehilfe auf Abruf? Der „Chief Social Media Officer“

Karsten Wendland[1], Joachim Heinz[2], Frank Wolf[2]

[1] Hochschule Aalen, Fakultät Optik und Mechatronik
Anton Huber Straße 21, 73430 Aalen
karsten.wendland@htw-aalen.de

[2] T-Systems Multimedia Solutions GmbH
Riesaer Straße 5, 01129 Dresden
F.Wolf@t-systems.com, Joachim.Heinz@t-systems.com

Abstract: Im Kontext der Informatik und ihrer Technologien sind immer wieder neue Tätigkeitsfelder und Aufgabenbereiche entstanden, die mit einschlägigen Bezeichnungen die Verantwortungsbereiche und das Aufgabengebiet der handelnden Akteure umreißen. Aktuelle Beispiele für diese Profilbildungen sind der „Webmaster“, der „User Interface Designer“, der „Chief Information Officer“ sowie der am Horizont der aktuellen Begriffsentwicklungen identifizierbare „Chief Social Media Officer“. Dessen Bedeutung, Rolle und Profil stehen im Zentrum des vorgeschlagenen Workshops.

Thema das Workshops

Im Anwendungsbereich „Web“ sind in den letzten Jahren zahlreiche neue Angebote und Infrastrukturen entstanden (Skype, Twitter, Corporate Wikis, Blogs, Facebook, XING usw.), die eines gemeinsam haben: einen Trend zu deutlich mehr direkter Kommunikation zwischen Beteiligten über Abteilungs-, Hierarchie- und Unternehmensgrenzen hinweg.

Für versierte Anwender ist es leicht geworden, sich unter Zugriff auf die Vielfalt neuer Möglichkeiten vom passiven Rezipienten zum aktiven Kommunikator zu wandeln, also vom Leser zum Schreiber im „Mitmach-Web“ fortzuentwickeln und neue Technologien gezielt zur Kommunikation und der Organisation von Zusammenarbeit einzusetzen. Die hierzu „funktionierenden“ Rollenkonzepte hinken allerdings der technologischen Fortentwicklung hinterher:

- Auf der Seite der Anwender vermischen sich bei der Nutzung von „Social Media“ jene Rollenkonzepte, in denen einst bewusst auf eine Trennung von beruflichen und privaten Lebensbereichen Wert gelegt wurde. In der spielerisch-leichten Überwindung etablierter organisatorischer Settings, Arbeits- und

Kommunikationsformen stellt sich derzeit nicht selten die Frage, ob z.B. ein Blogger aus seiner beruflichen Mitarbeiterrolle heraus, als Mitglied einer übergreifenden Fachgruppe oder als Privatperson seine Postings verkündet.

- Auf der Seite der Systembereitstellung im Unternehmenskontext – also an Schnittstellen z.B. zwischen den Fachabteilungen IT und Interner Kommunikation – besteht typischerweise wenig Vorerfahrung mit dem recht jungen Phänomen, dass sich Rollen, Kommunikationsformen und Arbeitsprozesse „im Virtuellen" plötzlich verändern und Eigendynamiken entstehen, die unmittelbar auf die Unternehmenskultur zurückwirken.

Die klassischen Verteilungen im Unternehmen wie Marketing, IT, Interne Kommunikation etc. stehen vor der Herausforderung,

- das technische und soziale Potenzial von Social Media zu verstehen,
- zu begreifen,
- und es strategisch und konzeptionell aufzugreifen und auszugestalten.

Ein seit einigen Monaten diskutierter Ansatz ist, die hierzu nötigen Kompetenzen in der Rolle eines „Chief Social Media Officers" zu bündeln. Ein solcher Akteur müsste über profunde Kenntnisse in den verschiedenen beteiligten Disziplinen verfügen und könnte vor allem – so die These für den Workshop – aus dem fachlichen und methodischen Fundus der Informatik schöpfen.

Der Workshop soll ein Forum dazu bieten, konstruktive Ansätze, Erfahrungswerte und kritische Interventionen zum Rollenkonzept „Chief Social Media Officer" zusammen zu führen, gegenüber zu stellen und erkenntnisorientiert zu diskutieren. Ein Schwerpunkt der Diskussion sollte auf den Anforderungsprofilen an die Rolle des „Chief Social Media Officer" selbst sowie auf den Anforderungen an dessen Rahmenbedingungen liegen.

Programmkomitee

- Conny Becker, Südwestrundfunk
- Stephan Grabmeier, Deutsche Telekom AG, Bonn
- Dorina Gumm, effective WEBWORK GmbH, Hamburg
- Jochen Günther, Fraunhofer IAO Stuttgart
- Frank Hamm, Aareal Bank, Wiesbaden
- Lutz Hirsch, Hirschtec Hamburg
- Georg Horntrich, Akademisches Zentrum Frankfurt
- Udo Seelmeyer, Institut für Sozialinformatik, Bielefeld
- Martin Seibert, //SEIBERT/MEDIA GmbH, Wiesbaden
- Gottfried Vossen, Westfälische Wilhelms-Universität Münster
- Hans-Dieter Zimmermann, FHS St. Gallen

Workshop

IT-Governance in verteilten Systemen (GVS)

Nicolas Repp, Stefan Schulte, Ulrike Steffens

Workshop

IT-Governance in verteilten Systemen (GVS) 2010

Nicolas Repp[1], Stefan Schulte[1], Ulrike Steffens[2]

[1]Technische Universität Darmstadt
contact@nrepp.de, schulte@kom.tu-darmstadt.de

[2]Softwaretechnik in betrieblichen Informationssystemen
OFFIS Institut für Informatik
ulrike.steffens@offis.uni-oldenburg.de

1 Einleitung

Ansätze und Strategien zur IT-Governance sind für viele Unternehmen heutzutage nahezu unverzichtbar geworden. Dazu trägt die Notwendigkeit bei, die IT des Unternehmens an den aktuellen Trends der Gestaltung und Entwicklung komplexer IT-Systeme auszurichten. Hier liegt das Gewicht hauptsächlich auf verteilten System- und Softwarearchitekturen, die beispielsweise dem Paradigma der Service-orientierten Architekturen (SOA) folgen.

In diesem Umfeld führt insbesondere die Steuerung und Kontrolle der Systemlandschaft und deren Zusammenwirken mit der fachlichen Seite zu neuen Herausforderungen, welche durch ein ganzheitliches IT-Management adressiert werden müssen. Um diesen Herausforderungen ganzheitlich begegnen zu können, ist die Entwicklung zeitgemäßer Governance-Ansätze sowie unterstützender Applikationen notwendig. Im Rahmen des Workshops „IT-Governance in verteilten Systemen (GVS)“ werden aktuelle Forschungsarbeiten einschließlich „Work in Progress“ zum Themenkomplex IT-Governance in verteilten Systemen vorgestellt und diskutiert.

Die Themenschwerpunkte des diesjährigen Workshops lagen auf verschiedenen Aspekten der IT-Governance, insbesondere der SOA-Governance. Wir möchten uns bei allen Autoren für die jeweiligen Beiträge bedanken. Aus den eingereichten Papieren wurden sechs zur Präsentation angenommen: Sabine Buckl et al. stellen eine konzeptionelle Methode zur Dokumentation und Analyse von Zugriffsrechten in einer unternehmensweiten IT-Landschaft vor („A method for constructing enterprise-wide access views on business objects“), Stefanie Looso beschreibt ein Verfahren zum systematischen Einsatz verschiedener Best-Practice-Referenzmodelle („Multi-Modell-Umgebung IT-Governance: Einsatz mehrerer Best-Practice-Referenzmodelle“) und Martin Sommer stellt eine Befragung bzgl. des Managements von Mitarbeiter-Fähigkeiten („Skills“) im Rahmen der Einführung und des Betriebs einer SOA vor („Skills-Management – Schattendasein der SOA-Governance).

Des Weiteren beschäftigen sich Michael Niemann et al. mit dem semantischen Vergleich von Ist-Prozessen mit Best Practices aus Governance-Frameworks („Semantische Analyse zur Unterstützung von SOA-Governance“), Danijel Milicevic und Matthias Goeken untersuchen die Metamodellierung des ISO 27001-Standards („Konzepte der Informationssicherheit in Informationssicherheitsstandards am Beispiel ISO 27001“) und Wolfgang Johannsen behandelt in seinem Beitrag Fragestellungen der Information Governance und ihre Herausforderungen in verteilten Umgebungen („Information-Governance – Herausforderungen in verteilten Umgebungen“).

2 Programmkomitee

Der Workshop IT-Governance in verteilten Systemen findet in diesem Jahr zum 2. Mal statt. Die ursprüngliche Idee zur Organisation des Workshops entstand im Rahmen der Zusammenarbeit der Organisatoren und von Mitgliedern des Programmkomitees im Arbeitskreis SOA der Fachgruppe Software-Architektur der Gesellschaft für Informatik (http://www.ak-soa.de/). Der Arbeitskreis ist eine stetig wachsende Kommunikations- und Arbeitsplattform für Interessierte aus Wissenschaft und Industrie. Im Rahmen regelmäßiger Workshops, bei denen Teilnehmer ihre Forschungsideen und Projektarbeiten zum Thema SOA vorstellen und diskutieren, ist mittlerweile eine Community entstanden, die Serviceorientierung nicht als bloßes Hype-Thema versteht, sondern als echte Chance zur Flexibilisierung und damit zur Zukunftsfähigkeit der IT in Unternehmen. Der Arbeitskreis steht neuen Teilnehmern jederzeit offen und lädt auch interessierte Leser dieses Tagungsbandes herzlich zur Mitwirkung ein.

Die Organisatoren möchten sich an dieser Stelle für die Unterstützung der Mitglieder des Programmkomitees bedanken. Ohne ihre Mithilfe wäre die Organisation des Workshops nicht möglich gewesen. Dem Komitee des diesjährigen Workshops gehörten die folgenden Personen an:

- Dr. Markus Böhm, PricewaterhouseCoopers WPG AG
- Prof. Dr. Matthias Goeken, Frankfurt School of Finance & Management
- Jens Happe, SAP Research Karlsruhe
- Prof. Dr. Wilhelm Hasselbring, Christian-Albrechts-Universität zu Kiel
- Prof. Dr. Bernhard Humm, Hochschule Darmstadt
- Dr. Christian Janiesch, SAP Research Brisbane
- Dr. Wolfgang Johannsen, it’s okay Ltd. & Co. KG
- Prof. Dr.-Ing. Arne Koschel, Fachhochschule Hannover
- Dr. Samuel Kounev, KIT Karlsruhe
- Gabriela Loosli, Universität Bern
- Michael Niemann, Technische Universität Darmstadt
- Dr.-Ing. Stefan Pühl, Dell Services
- Prof. Dr.-Ing. Ralf Steinmetz, Technische Universität Darmstadt
- Johannes H. Willkomm, Capgemini sd&m

A Method for Constructing Enterprise-wide Access Views on Business Objects

Sabine Buckl, Florian Matthes, Sascha Roth,
Christopher Schulz, Christian M. Schweda

Lehrstuhl für Informatik 19 (sebis)
Technische Universität München
Boltzmannstr. 3, 85748 Garching
{sabine.buckl, matthes, sascha.roth, christopher.schulz, schweda}@in.tum.de

Abstract: Modern application landscapes consist of a multitude of inter-connected business applications exchanging data in manifold ways. These business applications are used by employees who take on several organizational roles. However, when broadening the scope to an enterprise-wide perspective, lack of clarity and uncertainty prevail regarding the questions which roles have access on which business applications as well as the business objects managed by them. In contrast to alternative approaches, e.g. single sign on, this paper focuses on the enterprise-wide availability of business objects. Motivated by a case study from the car manufacturing industry, this article describes a method to analyze and justify an enterprise-wide access matrix. Respective viewpoints are presented facilitating a business object access management on enterprise level. The method and concepts are further substantiated by means of the case study.

1 Introduction and motivating example

Nowadays, an enterprise's typical application landscape is characterized by a plethora of *business applications* which are interlinked via interfaces used for data exchange. In the course of time, the amount of data stored in these business applications continuously increases, among others as a result of the broadening business support. While the provision of business support inevitably means that employees get access on data that 'they need to know', access to further data has to be restricted. Role-based access mechanisms allow distinctive read and write permissions, while grouping access rights to logical units, so called *roles*. Against the background of an increased interconnectivity of business applications, the mechanisms of role-based access control deserve in-depth analyses. Especially the question, whether a role may transitively access confidential information is of high interest.

The need for a matrix describing the access permissions of a role in a comprehensive manner is confirmed by one of our industry partners acting in the automotive industry. Recent economic development has encouraged this partner to construct a car production plant in the United States to assemble and sell products right on site. From an IT perspective, this means that 170 business applications hosted in Germany will also provide business support for the new plant. Due to *Federal Rules of Civil Procedure*[1] as well as the amendments made

[1] see http://www.uscourts.gov/rules/newrules4.html for details.

on federal state level[2], an attorney filing a lawsuit possibly gets access to electronically stored information retrievable by making use of business applications. In case of the car manufacturer, claims considering specific automotive parts (e.g. front brakes, or gear drive) may allow the plaintiffs counsel in some circumstances to conduct detailed on-site investigations to find additional evidence. These investigations, which are subsumed by the term *eDiscovery* [Gei08], may also encompass data about parts which are not in the focus of the lawsuit. From the perspective of the industry partner, these parts are denoted *critical products*. Translated into the terminology of *business objects*, a critical product is mapped to a business object which in turn is stored within an business application. Hence, if a product is deemed critical since it may be subject to a litigation, the according business object in addition to the storing business applications are marked as critical, too.

The problem motivating the development and usage of a business object access matrix for the industry partner can be stated as follows: In the event of *substantiated suspicion*, a lawyer investigating on a critical product may legally have the right to undertake the role of those employees, who were dealing with the according business object. This in turn may allow the legal representative to gain access on additional business objects and related business applications which originally were not in scope of the litigation. Once access is granted to a business application, the lawyer could retrieve data in neighboring systems possibly leading to further claims. As a consequence, concrete links between roles, business objects, and business applications should be identified and later on adjusted, mitigating the risk which would be emanated by a lawyer who has comprehensive access rights.

In this respect, an enterprise-wide business object access matrix should address the following question: *Which role has permissions to access business objects stored in business applications?* The question is taken further by differentiating between direct and transitive access on business objects besides the distinction of the actual and desired access matrix. Whereas former refers to the possibility that a role manipulates business objects whose according data are not directly stored in a business application but which may be accessed transitively through the connection to a second business application, latter picks up the thought of distinguishing between current and target access state. The presented method addresses the above question is complemented with a set of relevant viewpoints visualizing the relations between role, business object, and business application, and allow to identify:

- Which business objects are stored in specific business applications?
- Which roles have access to which specific business applications?
- Which business applications are connected with other business applications?
- Which roles have access to which specific business objects?

2 Related work

In [JE07], Johnson and Ekstedt discuss models and analyses to support information system decision making on an enterprise level. Thereby, security is identified as one major aspect

[2] see http://www.uscourts.gov/rules/EDiscovery_w_Notes.pdf for details.

which is understood as consisting of the topics confidentiality, integrity, and availability. In respect to our example, especially the aspect of confidentiality, which considers the degree to which a business application protects data from being accessed by unauthorized users, is of interest. While Johnson and Ekstedt mentioned main causalities for security, like e.g. "quality of security policies" or "quality of organizational aspects of information security", no detailed method how to address confidentiality issues, is presented.

In [Win08], enterprise-wide information logistics (IL) is defined as a process of planning, implementing, as well as controlling cross-unit data flows in addition to the storage and provisioning of the respective data. While pointing out its general scope and characteristics, [Win08] also presents managerial challenges for an enterprise implied by IL and compares the concept to approaches like data warehousing, business intelligence, and management support systems. By centering around information which is spawned, processed, and used in an enterprise-wide context, the article motivates the fundamental necessity to manage information on an enterprise level in a holistic manner. In contrast to [Win08] who remains on a rather general point of view, we address a real-life case explicitly in proposing a solution.

The work of [DMR+05] thoroughly examines enterprise-wide data management, acting on the assumption of six distinct stages of development, i.e. isolated applications, usage of DBMSs, standardization of data, distinct responsibilities for data, information as enterprise resource, combination of information to knowledge. Covering the role and importance of data for a modern enterprise in detail, [DMR+05] also presents a matrix depicting which business function accesses and manipulates which specific entity. While the viewpoint is only briefly sketched by the authors, the approach presented in this article goes further in proposing a concrete method for designing enterprise-wide access views. Our matrix-based viewpoints consider three main concepts (role, business object, business application) including also transitive relations due to the reference to a concrete use case.

In his work, [Jun06] copes with the integration of data by elaborating different integration architectures which are divided in a typology, replication, transaction type, and synchronization control dimension. The work analyzes the management of information in an enterprise context by differentiating between business and technical perspectives. Thereby, the concept of a *information object type* as specified in the book can be directly mapped on the business object used in our work. However, while focusing on data integration, [Jun06] does not tackle the issue of access management through roles on business objects which are stored in different information systems, i.e. business applications.

Control Objectives for Information and related Technology (COBIT) represents a prominent framework for IT management created by the Information Systems Audit and Control Association in cooperation with the IT Governance Institute [IT 09]. The framework is focused on what is required to achieve adequate management and control of IT, and is positioned at a high level. Besides structuring IT in four core domains which are subdivided into 34 IT processes, COBIT suggests illustrative examples of those processes. Thereby, each example contains generic inputs and outputs, key activity goals, metrics, as well as activities and guidance on roles and responsibilities reified in a *Responsible, Accountable, Consulted and Informed* (RACI) chart. RACI charts, which are presented via matrices within the framework, identify who, i.e. which function, is responsible, accountable,

consulted, and/or informed when a specific IT process activity is performed. While the viewpoint structure in the form of a matrix is identical to the one presented in this article, its content and field of application is different.

3 Enterprise-wide access control on relevant business objects

The initial step in constructing an enterprise-wide access model is the creation of an *information model*, i.e. a conceptual model containing the relevant concepts of the problem domain. Figure 1 presents the model developed in cooperation with the industry partner – it contains the three main concepts of our problem description: role, business application, and business object. As described in the model, roles have access to business applications storing different business objects. Furthermore, business applications are connected via interfaces to other business applications, whereas interfaces may constitute of database dumps of another business applications, i.e. business objects are transfered without any access control leading to transitive access for roles as claimed in the Section 1.

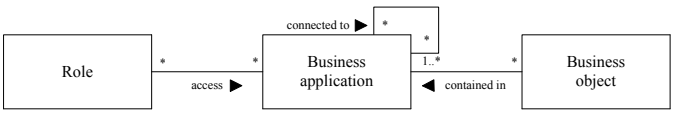


Figure 1: Information model for an enterprise-wide access management

A method for constructing enterprise-wide access views on business objects is shown in Figure 2 using the Business Process Modeling Notation (see [OMG08]). In the following, the method steps are explained in detail.

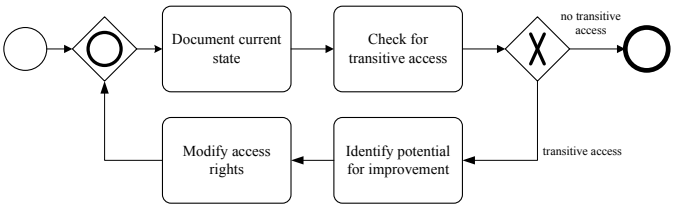


Figure 2: A method for creating enterprise-wide access views on business objects

Document current state gathers information about current access rights, hence documenting which business objects are stored in a business application which in turn may be accessed by one to many roles. Thereby, viewpoints making use of a table form can be used. The resulting views (cf. Figure 3) have to be filled with information gathered from different business units within the enterprise which have information on the access permissions. This documentation task can be performed in various ways, e.g. interviewing the respective stakeholders in the business units or observing actors during their IT-supported work. While the former methods are top-down, hence from a user's point of view, bottom-up approaches from a business application perspective can also be applied.

First, the car manufacturer documents which roles have direct access to which business applications, see Figure 3(a), and the stored business objects, see Figure 3(b). Afterwards,

the car manufacturer documents which business application accesses other business applications (cf. Figure 3(c)). For instance, system administrators reported that they send a nightly database dump of the CRM *to the* ERP *business application.*

	Business application		
Role	*CRM*	*ERP*	*Portal*
HR	●	○	●
Sales	●	○	●
R&D	○	●	○

(a) Role ⇒ Business application

	Business application		
Business object	*CRM*	*ERP*	*Portal*
Customer	●	○	●
Construction plan	○	●	○
Contract	○	○	●

(b) Business object ⇒ Business application

	Business application		
Business application	*CRM*	*ERP*	*Portal*
CRM	●	○	○
ERP	●	●	○
Portal	○	○	●

(c) Business application ⇒ Business application

Legend	
●	access
○	no access

Figure 3: Enterprise-wide access matrices

Check for transitive access performs an expert-based analysis for transitive access on business objects. In a first step, the access rights of roles via intermediary business applications on corresponding business objects are computed and visualized using a tabular viewpoint. The according view is in a second step enriched with information on transitive access, i.e. with information on business objects that may be reached from a role via exploiting the interrelations between the business applications.

Figure 4 depicts the results of the step checking for transitive access. This view is taken into account by experts from the car manufacturer, who assess the criticality of the corresponding direct and transitive access permissions.

	Business object		
Role	*Customer*	*Construction plan*	*Contract*
HR	●	○	●
Sales	●	○	●
R&D	○ [●]	●	○ [●]

Legend	
●	access
○	no access
[●]	transitive access

Figure 4: Role ⇒ Business object

Identify potential for improvement is conducted by experts, who analyze transitive access relations, and identify the subset violating enterprise policies or legal rights. In using the views and the therein contained knowledge, concrete action items can be derived which aim at restricting access on business objects. *Access violations are identified by the legal department at the car manufacturer, i.e. the role* R&D *accessing the* ERP *business application is also allowed to read the business objects* Customer *and* Contract *violating enterprise policies. According to Figure 4 the documented transitive properties represent access violations since the role* R&D *should not be allowed to read the* Customer *and* Contract *business object. Consequently, a root cause analysis is performed at the car manufacturer, revealing that the connection between* CRM *and* ERP *causes this access violations. Since business processes were improved, the* ERP *system required information on the business object* Customer*, and for this reason, a nightly database dump of the* CRM *is transfered to the* ERP.

Modify access permissions calls on well-known methods ranging from location-based restrictions of network access to role-based application layer access control. Thus, access permissions are modified in order to incorporate the improvements identified by experts. Afterwards, a new current state of the access permissions is documented (cf. Figure 2).

While the business object Customer *is not categorized as* critical product*, the business object* Contract *is. Thus, the car manufacturer immediately restricts access to the* CRM *business application and contacts the owners of the* ERP *and* CRM *business applications in order to conjointly work out a solution which prevents the* ERP *business application to access the data of the* CRM *while bewaring business functionality.*

4 Outlook

This article presented a method to construct enterprise-wide access views in order to obtain a holistic knowledge which roles have permissions to access business objects stored in specific business applications. Motivated by a real-world problem stated by one of our industry partners from the German car industry, this article suggests viewpoints in terms of matrices helping to identify role-based access rights on an enterprise level.

We are currently implementing the approach at the industry partner, where first results demonstrate its suitability. Nevertheless, we are aware of further challenges tying in with the presented method. To complete the proposed matrices with real data and therefore allow a reliable estimation about the business objects and business application instances a certain role may have access to, a huge amount of data has to be collected. Since the concrete information is spread out in either the business but also in the different IT units, a common terminology understandable by both sides has to be defined in advance.

References

[DMR+05] Rolf Dippold, Andreas Meier, André Ringgenberg, Walter Schnider, and Klaus Schwinn. *Unternehmensweites Datenmanagement - Von der Datenbankadministration bis zum modernen Informationsmanagement*. Vieweg, 4rth edition, 2005.

[Gei08] Ivo Geis. eDiscovery und Datenschutz. Website, 2008. http://www.ivo-geis.de/veroeffentlichungen/eDiscovery_und_Datenschutz.pdf, visited on April 20th, 2010.

[IT 09] IT Governance Institute. Framework Control Objectives Management Guidelines Maturity Models. Website, 2009. http://www.isaca.org/Knowledge-Center/cobit; visited on June 27th, 2010.

[JE07] Pontus Johnson and Mathias Ekstedt. *Enterprise Architecture – Models and Analyses for Information Systems Decision Making*. Studentlitteratur, Pozkal, Poland, 2007.

[Jun06] Reinhard Jung. *Architekturen zur Datenintegration: Gestaltungsempfehlungen auf der Basis fachkonzeptueller Anforderungen*. Vieweg+Teubner, Wiesbaden, Germany, 1st edition, 2006.

[OMG08] OMG. Software & Systems Process Engineering Meta-Model Specification (formal/2008-04-01), 2008.

[Win08] Robert Winter. Enterprise-wide information logistics: Conceptual foundations, technology enablers, and management challenges. In *30th International Conference on Information Technology Interfaces (ITI 2008)*, pages 41–50. IEEE Computer Society Press,U.S., Juni 2008.

Multi-Modell-Umgebung IT-Governance:

Einsatz mehrerer Best-Practice-Referenzmodelle

Stefanie Looso

Frankfurt School of Finance and Management
Sonnemannstraße 9-11
60314 Frankfurt
s.looso@frankfurt-school.de

Abstract: Unternehmen setzen zunehmend mehrere Best-Practice-Referenzmodelle (BPRM) der IT-Governance ein. Ein unsystematischer Einsatz von mehreren BPRM kann erhebliche Ineffizienzen bergen, daher ist ein systematischer kombinierter Einsatz erstrebenswert. Um den kombinierten Einsatz mehrerer Modelle gezielt zu unterstützen, sind genaue Kenntnisse der Kombinationsmöglichkeiten von BPRM in einer Multi-Modell-Umgebung notwendig. Diese werden in diesem Beitrag zunächst aufgearbeitet, bevor zwei Möglichkeiten präsentiert werden, die die Kombination von Modellen gezielt unterstützen.

1 Einführung und Problemstellung

Die für den weitgefassten Bereich IT-Governance verfügbaren Best-Practice-Referenzmodelle (BPRM) entsprechen unterschiedlichen Perspektiven auf die Unternehmens-IT. So wird beispielsweise COBIT (IT Governance Institute 2007) von Wirtschaftsprüfern und Auditoren für die IT-Prüfung bevorzugt, während im Bereich Systementwicklung verstärkt CMMI (Software Engineering Institute 2007) Verwendung findet. Der IT-Betrieb wiederum hat seine Prozesse vielfach an ITIL oder ISO 20000 ausgerichtet (KPMG 2004). Die Gründe für den Einsatz von mehreren BPRM (oder Teilbereichen mehrerer BPRM) sind vielfältig. Mögliche Gründe sind etwa: gesetzliche Vorschriften oder Empfehlungen, Marketingüberlegungen, Branchenvergleiche, persönliche Vorlieben von Entscheidern, Mode, etc. (Alter et al. 2009). Unternehmen stehen daher gegenwärtig vor der Herausforderung mehrere BPRM (bzw. deren Teilbereiche) einzusetzen oder selbiges zu planen (Cater-Steel et al. 2006). Verwendet ein Unternehmen verschiedene Modelle, wird dies zumeist als *Multi-Modell-Umgebung (MMU)* bezeichnet (Siviy et al. 2008; Looso et al. 2010).

Für den Einsatz von mehreren BPRM gibt es grundsätzlich zwei Möglichkeiten. Die Modelle können erstens unabhängig von einander (etwa nacheinander oder parallel) oder zweitens kombiniert eingesetzt werden. Insbesondere bei vorliegenden Überschneidungen ist der kombinierte Einsatz nach Ansicht der Verfasserin erstrebenswert (siehe Alter et al. 2009). Solche Überschneidungen liegen im Bereich der IT-Governance vor. Ursache der zum Teil erheblichen Überschneidungen ist bspw., dass der Gegenstandsbereich

der BPRM wächst und die jeweils neuen Versionen im Vergleich zu ihren Vorgängern größere fachliche Bereiche abdecken. Durch die Ausdehnung der Modelle ergeben sich so fortlaufend neue Schnittmengen und Redundanzen bei vormals tendenziell spezialisierten und nahezu überschneidungsfreien Modellen. Diese Überschneidungen können unter anderem zu Ineffizienzen beim Einsatz von BPRM führen. So führt die Verwendung von unterschiedlichen BPRM bspw. dazu, dass getrennte Sprachwelten entstehen und so bereits das Erstellen eines unternehmensweiten Glossars erhebliche Herausforderungen birgt (Alter et al. 2009). Weiterhin können die Anforderungen der verschiedenen BPRM an denselben Prozess kollidieren, indem unterschiedliche Schwerpunkte gesetzt oder Philosophien vertreten werden. Überschneidungen zwischen BPRM führen in Unternehmen daher etwa zu Doppeltarbeit, Mehrfachdokumentation oder im Extremfall zu redundanter Prozessmodellierung. Somit ist festzuhalten, dass der parallele Einsatz verschiedener BPRM die an sich schon hohe Komplexität von BPRM-Anwendungen zusätzlich erhöht. Bei fehlender Abstimmung können die Modelle so außerdem keine gemeinsame Steuerungswirkung entfalten. Der kombinierte Einsatz von BPRM ist demnach erstrebenswert. Um dies zu unterstützen wird in diesem Beitrag die Kombination von BPRM thematisiert.

2 Kombination von mehreren BPRM

2.1 Vorüberlegungen

Fraglich ist zunächst, wo die Kombination von BPRM ansetzen kann. Wie Alter et al. 2009 zeigen kann eine Kombination auf Modell- oder auf Metamodellebene erfolgen. Diese Kombinationsaktivitäten lassen sich weiterhin durch eine Unterscheidung von Referenzebene und Unternehmensebene klassifizieren. Die Unternehmensebene enthält die an die spezifische Situation eines Unternehmens angepassten BPRM. Beispielsweise sind unternehmensspezifische Modelle häufig inhaltlich oder strukturell verkürzt (siehe Looso 2010). Die Referenzebene enthält die Modelle in ihrer ursprünglichen Form. Die Aktivitäten *Kombination auf Metamodellebene* und *Kombination auf Modellebene* sind also weiter unterteilbar in:

- Kombination der Metamodelle zweier BPRM (Referenzebene)
- Kombination zweier unternehmensspezifischer Metamodelle (Unternehmensebene)
- Kombination zweier BPRM (Referenzebene)
- Kombination zweier angepasster BPRM (Unternehmensebene)

Zu dieser Unterteilung ist anzumerken, dass sich die beiden Kombinationsformen auf Referenz- und Unternehmensebene jeweils technisch nicht signifikant unterscheiden. Sie unterscheiden sich lediglich durch die zu kombinierenden Modelle und den Treiber der Kombination. Eine Modellkombination auf Referenzebene müsste im Gegensatz zu einer Modellkombination auf Unternehmensebene nämlich von den herausgebenden Institutionen der beiden BPRM durchgeführt bzw. unterstützt werden. Wäre dies nicht der Fall, würde das resultierende kombinierte Modell seinen Referenzcharakter verlieren. Eine

solche offizielle Modellkombination ist jedoch praktisch sehr unwahrscheinlich und eher von akademischem Interesse. Im Folgenden werden daher zwei Aktivitäten beschrieben, die von Unternehmen durchgeführt werden können um den gemeinsamen Einsatz zweier BPRM unterstützen. Die Kombination durch Schemaintegration kombiniert BPRM auf Metaebene, während die am Zusatznutzen orientierte Kombination die Kombination auf Modellebene unterstützt. Die beiden Formen ließen sich jedoch auch auf Referenzebene einsetzen.

2.2 Metamodell-Kombination durch Schemaintegration

Ein mögliches Verfahren zur Vereinigung zweier Metamodelle innerhalb eines Unternehmens ist die Schemaintegration, die bereits im Bereich der Datenbankforschung verwendet wird. Für das Zusammenführen von Komponenten sind dort im Allgemeinen mehrere Schritte vorgesehen (Spaccapietra et al. 1994; Spaccapietra et al. 1992; Rizopoulos et al. 2005). Nach der Vorintegration, die festgelegt, welche Metamodell-Komponenten in welcher Reihenfolge konsolidiert werden sollen, kommt der für diesen Beitrag maßgebliche zweite Schritt „Vergleich und Konfliktidentifikation“. In diesem Schritt erfolgt der Schemavergleich, bei dem Beziehungen zwischen den Metamodell-Komponenten aufgrund ihrer Modell-Komponenten identifiziert, analysiert und dokumentiert werden. Insbesondere sollen Konflikte und Inkonsistenzen aufgedeckt werden. Unterschieden werden in Anlehnung an Conrad 2002 vier semantische Beziehungen zwischen zwei Metamodell-Komponenten (Abbildung 1). Der Vergleich und die Konfliktidentifikation der BPRM COBIT und CMMI haben beispielsweise ergeben, dass die COBIT-Metamodell-Komponente Process und die CMMI-Metamodell-Komponente Process Area aufgrund ähnlicher Modell-Elemente (Instanzen) eine Überlappung aufweisen. Demgegenüber gibt es in CMMI keine Entsprechung für die COBIT-Metamodell-Komponente Metric. Auf der Grundlage der erwähnten Vergleichergebnisse wird in einem weiteren Schritt ein gemeinsames Metamodell erstellt. Ein solches ge-

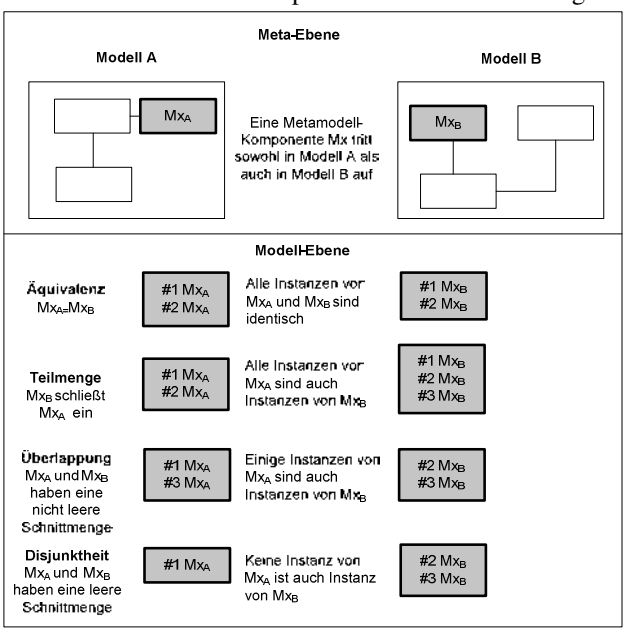


Abbildung 1 Semantische Beziehungen

meinsames Metamodell führt zu einem höheren Maß an Transparenz der unterschiedlichen Sprachwelten. Hierbei kommen im Zuge der Schemaintegration Integrationsregeln zum Einsatz, die festlegen, wie Metamodell-Komponenten, zwischen denen eine Beziehung besteht, in einem gemeinsamen Modell abgebildet werden. Bspw. kann bei Äquivalenz eine gemeinsame Metamodell-Komponente ohne Änderungen in das globale Modell aufgenommen werden; bei Überlappung bietet sich die Bildung eines integrierenden, gemeinsamen Metamodellelements (ggf. mit neuem Bezeichner) an, welches die Modellelemente beider BPRM semantisch voll umfasst. Das so entstandene gemeinsame Metamodell mehrerer BPRM ist eine sinnvolle Möglichkeit, den gemeinsamen Einsatz der BPRM zu unterstützen.

2.3 Modellkombination durch am Zusatznutzen orientierte Kombination

Die vorhandenen Ansätze zur Kombination von BPRM auf Modellebene beruhen zumeist auf Vergleichen. Deren Ziel ist jedoch nach Ansicht der Verfasserin nicht die Modellvereinigung, sondern das Aufzeigen von Schnittstellen (siehe bspw. Hochstein et al. 2003; Böhmann et al. 2004; Johannsen et al. 2007). Die bereits erwähnten vorhandenen Mappings vergleichen bspw. zwei BPRM direkt miteinander. Dabei zeigen sie in der Regel in einem 1:1-Vergleich Schnittstellen auf, d.h. sie suchen die Entsprechung einer Modell-Komponente im anderen Modell (siehe COBIT-Mappings oder itSMF et al. 2008). Diese Fülle an Ähnlichkeitsaussagen ist einerseits in ihren Ausmaßen sehr umfangreich und andererseits nicht auf eine gemeinsame Anwendung der Modelle hin ausgerichtet. Daher ist die Entwicklung eines Kombinations-Verfahrens notwendig. Mit der am Zusatznutzen orientierte Kombination von BPRM (ZNOK), wird in diesem Beitrag ein auf den gemeinsamen Einsatz von zwei BPRM ausgerichteter Ansatz vorgestellt. Im Fokus dieses Verfahrens steht der zu erwartende Zusatznutzen einer Kombination von BPRM gegenüber dem Nutzen eines einzelnen BPRM. Die Form der Kombination kann bspw. eingesetzt werden, wenn ein Unternehmen bereits ein BPRM anwendet und sich durch den kombinierten Einsatz des vorhandenen BPRM und des neuen BPRM einen zusätzlichen Nutzen verspricht. Der zusätzliche Nutzen ist jedoch ein subjektives, schwer zu operationalisierendes Konstrukt. Subjektive Einschätzungen des zusätzlichen Nutzens sind in diesem Fall jedoch unumgänglich, da der Zusatznutzen erst auf Unternehmensebene bestimmt werden kann. Die Zusammensetzung des Projektteams oder das Einbeziehen eines erfahrenen Beraters kann jedoch helfen die Bestimmung zu objektivieren.

Ansatzpunkt des ZNOK sind die Metamodelle des BPRM. Sie unterstützen die Auswahl, der für eine Kombination sinnvollen Metamodell-Komponenten. Im hier gewählten Beispiel wurden für COBIT die Metamodell-Komponente *Control Objective* und für CMMI die Metamodell-Komponente *Specific Practice* ausgewählt. Die Stellung der Elemente in den zugehörigen Metamodellen ist ein erster Hinweis für eine strukturelle Ähnlichkeit zwischen Control Objectives und Specific Practices. Die ZNOK ordnet nun der COBIT-Komponente Control Objective jede Modell-Komponente von CMMI zu, die einen Zusatznutzen für diesen Prozess liefert. Ziel ist es also nicht, wie in einem konventionellen Mapping, eine 1:1-Beziehung herzustellen, denn diese basiert auf Ähnlichkeit und hat daher einen geringen Zusatznutzen. Ziel der ZNOK ist es vielmehr, eine nutzen-

stiftende 1:n-Beziehung zu erstellen. Das bedeutet, dass in CMMI diejenigen Modell-Komponenten gesucht werden, die den COBIT-Prozess verbessern oder erweitern. Ein Auszug des Ergebnisses der ZNOK von COBIT und CMMI zeigt Tabelle 1.

Tabelle 1: Ausschnitt des Ergebnisses der COBIT-Domäne Plan & Organize

PO3.4 Technology Standards (Control Objective)
Hier fordert CMMI den verstärkten Einsatz von Reviews. Durch den Einsatz der CMMI Process Area (PA) *Verification*, speziell die Spezifische Praktik 2, werden für ausgewählte Prozessoutputs Partner Reviews durchgeführt. Zusätzlich wird empfohlen die PA *Process and Product Quality Assurance* einzusetzen um die Einhaltung der von COBIT geforderten Standards zu gewährleisten.
PO4.7 Responsibility for IT Quality Assurance (Control Objective)
Es empfiehlt sich einen *PPQA-Prozess* und die Reviews von *VER* gemäß CMMI in die vorhandene Qualitätssicherung einzugliedern, dadurch wird die Produkt- und Prozessqualität abgesichert.

Die ZNOK adressiert also die Kombination zweier BPRM auf der Unternehmensebene. Das kombinierte Modell hat jedoch ein BPRM als Basis, die Modelle sind also nicht gleichgewichtig kombiniert. Vorteilhaft an der ZNOK ist die Möglichkeit sie stufenweise anzuwenden. So kann etwa zunächst nur für einen bestimmten Teil des Ausgangsmodells ein Zusatznutzen im zweiten Modell gesucht werden.

3 Fazit und Ausblick

Überschneidungen zwischen BPRM können in Kombination mit einem unsystematischen Einsatz von BPRM Ineffizienzen auslösen, deren Beseitigung ökonomisches Potential birgt. In diesem Beitrag wurden daher zwei Ansätze für den systematischen Einsatz von mehreren BPRM in einer MMU vorgestellt. Beide Kombinationsansätze können sowohl auf Ebene der Referenzmodelle, d.h. durch die herausgebenden Institutionen, oder auch auf Ebene der unternehmensspezifischen Versionen der BPRM angewendet werden. Es gibt jedoch gegenwärtig keine Anzeichen für eine durch die jeweiligen Institutionen angestrebte Kombination von BPRM, die darauf basiert die Modelle wirklich zu einem gemeinsamen Modell zu kombinieren. Der Vergleich von BPRM auf Referenzebene ist jedoch immer wieder Ziel von Fachgruppen und Expertengremien. Die dort gewonnenen Erkenntnisse können wiederum von Unternehmen angewendet werden. Dadurch unterstützen die gewonnenen Erkenntnisse auf der Referenzebene die beiden hier beschriebenen Verfahren auf Unternehmensebene.

Ziel der Forschung muss es hingegen sein, den Einsatz mehrerer BPRM allgemeingültig zu beschreiben und objektive Kombinationsverfahren zu etablieren. Mit der Unterscheidung von Modell-Kombination und Metamodell-Kombination sind in diesem Beitrag zwei grundsätzliche Ansatzpunkte für eine Kombination beschrieben. Für beide Ebenen wurde jeweils ein Verfahren beschrieben, welches in weiterer Forschung konkretisiert werden muss. Weiterer Forschungsbedarf besteht insbesondere bezüglich der Subjektivität beider Kombinations-Verfahren. Ziel muss es sein, beide Verfahren allgemeingültig zu beschreiben und somit zu objektivieren. Zusammenfassend ist dieser Beitrag ein

erster Schritt hin zu einer solchen allgemeingültigen Beschreibung der Kombination von BPRM in einer Multi-Modell-Umgebung, er versteht sich außerdem als Grundlage einer zukünftigen wissenschaftlichen Diskussion.

Literaturverzeichnis

Alter, S.; Goeken, M.: Konzeptionelle Metamodelle von IT-Governance-Referenzmodellen als Basis der Kombination und Integration in einer Multi-Model-Umgebung. In (Hansen, H.R., Karagiannis, D., Fill, H.G. Hrsg.): Business Services: Konzepte, Technologien, Anwendungen. 9. Internationale Tagung Wirtschaftsinformatik, Band 1, Wien, 2009.

Böhmann, T.; Krcmar, H.: Grundlagen und Entwicklungstrends im IT-Servicemanagement. In HMD - Praxis Wirtschaftsinformatik, 237, 2004.

Cater-Steel, A.; Tan, W. G.; Toleman, M.: Challenge of adopting multiple process improvement frameworks: Proceedings of the 14th European Conference on Information Systems, 2006.

Conrad, S.: Schemaintegration. Integrationskonflikte, Lösungsansätze, aktuelle Herausforderungen. In Informatik – Forschung und Entwicklung, 2002, 17; S. 101–111.

Hochstein, A.; Hunziker, A.: Serviceorientierte Referenzmodelle des IT-Managements. In HMD - Praxis Wirtschaftsinformatik, 2003.

IT Governance Institute: COBIT 4.1, 2007.

itSMF; ISACA Hrsg.: ITIL-COBIT- Mapping. Gemeinsamkeiten und Unterschiede der IT-Standards. Symposion Publishing, Düsseldorf, 2008.

Johannsen, W.; Goeken, M.: Referenzmodelle für IT-Governance. dpunkt.verlag, Heidelberg, 2007.

KPMG: Summary of KPMG IS Governance Survey, London, 2004.

Looso, S.; Goeken, M.: IT Governance: Application of Best Practice Reference Models: Proceedings of the 18th European Conference on Information Systems, 2010.

Rizopoulos, N.; McBrien, P.; Magnani, M.: Schema Integration based on Uncertain Semantic Mappings: 24th International Conference on Conceptual Modeling, 2005; S. 31–46.

Siviy, J. Kirwan, P.; Marino, L.; Morley, J.: The Value of Harmonizing Multiple Improvement Technologies: A Process Improvement Professional's View, SEI, 2008.

Software Engineering Institute: CMMI, 2007.

Spaccapietra, S.; Parent, C.; Dupont, Y.: Model Independent Assertions for Integration of Heterogeneous Schemas. In VLDB Journal, 1992, 1; S. 81–126.

Spaccapietra, S.; Parent, C.: View Integration: A Step Forward in Solving Structural Conflicts. In IEEE Transactions on Knowledge and Data Engineering, 1994, 6; S. 258–274. 4.

Skills-Management - Schattendasein der SOA-Governance

Martin Sommer[1]

evodion Information Technologies GmbH
Högerdamm 41
20097 Hamburg
(IT-Consultant)

martin.sommer@soa-kompetenzen.de

Abstract: Die „Komponente“ Mensch stellt einen wesentlichen und erfolgskritischen Faktor bei der Realisierung von Service-orientierten Architekturen (SOA) dar. SOA erfordert neue Rollen und bringt neue Aufgaben mit sich. Auf operativer Ebene agiert das Skills-Management, welches durch geeignete Methoden sicherstellt, dass Personal korrekt ausgewählt bzw. passend weitergebildet wird. Skills-Management ist Bestandteil der SOA-Governance, das bislang ein Schattendasein fristete. Der vorliegende Beitrag konzentriert sich auf diesen organisatorischen Aspekt von SOA-Governance-Bestrebungen. Untermauert wird der Sachverhalt durch empirische Daten einer kürzlich durchgeführten Experten-Befragung.

1 Einleitung

SOA-Governance ist das zentrale Kontrollgremium zur Überwachung der strategisch konformen Realisierung und Weiterentwicklung einer SOA. Diese Governance-Disziplin umfasst weit mehr Aktivitäten als nur jene um die Management-Aktivitäten entlang des Service-Lifecycles. Neben der Betrachtung von technischen Fragestellungen (v.a. Identifizierung, Design und Implementierung von Services) fallen auch bedeutsame organisatorische Aspekte in den Rahmen der SOA-Governance. Beispielsweise müssen die erforderlichen Aufgaben den beteiligten Rollen [SS07, S.4] zugewiesen werden.

[1] Institut für Wirtschaftsinformatik, Professur Informationsmanagement (Prof. Dr. Franczyk), Universität Leipzig, Grimmaische Straße 12, 04109 Leipzig, (Externer Doktorand)

Die Ressource Mitarbeiter ist ein wesentlicher Erfolgsfaktor von IT-Projekten, wie die Studien von [Sc01] und [Ka06] belegen. Untersuchungen, die sich speziell auf den Projektkontext von SOA beziehen, zeigen ein ähnliches Bild. In der durchgeführten Untersuchung von (vgl. [DD07, S. 24]) wurden die kritischen Erfolgsfaktoren für die SOA-Einführung untersucht. Die Probanden dieser Studie gaben zu 88,6%an, dass die „Abstellung und Qualifizierung von Mitarbeitern“ für den Erfolg der SOA-Einführung relevant bzw. sehr relevant sind. In einer anderen Untersuchung der Firma evodion Information Technologies GmbH (vgl. [Ev07]) wurde herausgearbeitet, dass Unternehmen das fehlende Know-how als wesentliches Problem sehen.

Die Einführung von neuen Systemen erzeugt einen zweifachen Kompetenzbedarf, der durch das Wissen über das neue System (bzw. Paradigma) und der durch die Systemeinführung bedingten Veränderungen auf Ablauf- und Aufbauorganisationsebene entsteht. Diesen gilt es möglichst effektiv zu decken (vgl. [So09; BN07]). Vor diesem Hintergrund und der in der Praxis weiterhin präsenten Verunsicherung gegenüber SOA[2] stellt sich die Frage, wie intensiv die Skills der Mitarbeiter praktisch bei der Personalauswahl für SOA-Projekte Beachtung finden. Besonders die fachlichen und methodischen Kompetenzen werden kritisch gesehen, denn sie verkörpern das elementare Verständnis über die Serviceorientierung. Dieser Untersuchungsgegenstand wurde in einer dafür konzipierten Experten-Befragung näher beleuchtet. Sie steht im Mittelpunkt des nachstehenden Kapitels.

2 Expertenbefragung

Über einen zweiwöchigen Zeitraum Anfang September 2009 wurden per Onlinebefragung 77 Fachleute befragt, ob und wie potentielle Projektkandidaten auf ihre Kompetenzausstattung für den Einsatz in einem SOA-Projekt hin untersucht wurden. Die allgemeinen Kennzahlen über die Beteiligung wurden in Tabelle 1 zusammengefasst.

Kennzahl	Beschreibung	Wert absolut	Wert relativ
Einladungen	-	322	-
Rücklaufquote	Anzahl der Personen, die auf die Einladung reagiert haben	97	30,12%
Non-Drop-Outs	Anzahl der Teilnahmen, welche die Umfrage bis zum Ende durchgeführt haben.	77	79,39%
Drop-Outs	Anzahl der Fälle, welche die Umfrage angefangen, aber nicht bis zum Ende bearbeitet haben.	20	20,61%

Tabelle 1: Allgemeine Umfragekennzahlen

[2] Diese Verunsicherung wird speziell durch Mehrdeutigkeiten in der Fachterminologie bedingt sowie durch die teils leeren Versprechungen aus Berater- und Herstellerkreisen (vgl.[Ma07, S.10])

Der Fragebogen besteht aus vier Fragekomplexen mit insgesamt 31 Fragen. Der erste thematisiert den Aufbau des SOA-Projektes und die bereits eingenommenen Rollen. Gefolgt wird dieser durch ein Fragekomplex, über den mehr Daten zum Management von Kompetenzen gewonnen wurden. Im dritten Abschnitt wurden die Probanden dazu aufgefordert, eine subjektive Bewertung von Kompetenzen in den einzelnen Phasen eines SOA-Projektes für eine selbst gewählte Rolle durchzuführen. Dazu wurde eine fünfstufige Likert-Skalierung vorgegeben. Der letzte Teil der Umfrage bestand aus soziodemographischen Fragen zum Teilnehmer selbst. Das Profil der Teilnehmer ist sehr unterschiedlich. Sie kamen aus unterschiedlichen Branchen und haben unterschiedlichste Rollen in SOA-Projekten eingenommen. Im weiteren Verlauf des Kapitels werden die wesentlichen Erkenntnisse der Befragung thematisiert.

2.1 Evaluierte Kompetenzklassen

Die Teilnehmer der Umfrage wurden u.a. gefragt, ob vor ihrem wichtigsten SOA-Projekt die Mitarbeiter in Hinblick auf die notwendige Kompetenzausstattung für die zukünftigen Aufgaben evaluiert worden sind[3]. 33 Probanden gaben an, dass in ihrem wichtigsten SOA-Projekt die Skills im Vorfeld evaluiert wurden. In einer weiteren Frage wurden sie dazu angehalten zu konkretisieren, welche Klassen dabei im Vordergrund standen. Als Ergebnis wurden die fachlichen Kompetenzen 26 mal, die methodischen Kompetenzen 25 mal, die sozialen Kompetenzen 15 mal und die persönlichen Kompetenzen 14 mal genannt (siehe Tabelle 2). Die Kompetenzbündel grenzen sich klar von einander ab. Nicht verwunderlich ist, dass primär das fachliche Können im Vordergrund stand, denn dieses beinhaltet die außerordentlich wichtigen serviceorientierten Prinzipien. Darüber hinaus wurde ebenfalls großen Wert auf methodische Aspekte gelegt. Als sekundär entscheidend wurden die weichen Kompetenzen befunden. Ihr Niveau erreicht nicht die Intensität der harten Kompetenzen, ist jedoch verhältnismäßig stark ausgeprägt. Ein Indiz dafür könnte die Kommunikationsstruktur in den vorliegenden SOA-Projekten sein, in denen stark kommunikative Fähigkeiten abverlangt wurden. Begünstigt werden kann dies durch eine agile Vorgehensweise und große Teams. Beide erfordern einen höheren Kommunikationsaufwand unter den Mitgliedern des Projektes.

[3] Gestellte Frage: „Wenn Sie die Frage B1 mit Ja beantwortet haben, welche Kompetenzen wurden betrachtet?“. Dazu wurden die vier als kanonisch aufgefassten Kompetenzklassen vorgegeben [SS04] und mit Beispielen untermauert. Fachkompetenz ist zur Bewältigung einer beruflichen Aufgabe erforderlich. Bezogen auf den IT-Bereich stellen beispielsweise das Grundlagenwissen und Fertigkeiten aus dem Bereich Systementwicklung eine fachliche Kompetenz dar [NR05]. Methodische Kompetenzen sind situationsübergreifend anwendbar und meistens flexibel einsetzbare kognitive Fähigkeiten [NR05]. Als Beispiel kann hier das Problemlösen von komplexen und neuartigen Aufgaben genannt werden. Kommunikative und kooperative Fähigkeiten werden als soziale Kompentenzen zusammengefasst. Als Beispiel für diese Klasse kann die Kooperationsbereitschaft und die partnerzentrierte Interaktion und Verständnisbereitschaft genannt werden [NR05], [Sc09]. Schließlich bilden die personenbezogenen Dispositionen die letzte Kategorie. Beispiele für persönliche Kompetenzen sind Einstellungen, Wertevorstellungen, Zuverlässigkeit, Durchsetzungsvermögen und andere Dispositionen zu nennen. Es konnten mehrere Kompetenzklassen gewählt werden.

Kompetenzklasse	Absolute Häufigkeit	Relative Häufigkeit
Fachliche Kompetenzen	26	32,50%
Methodische Kompetenzen	25	31,30%
Soziale Kompetenzen	15	18,80%
Persönliche Kompetenzen	14	17,50%
Summe	80	100%

Tabelle 2: Gewichtung der evaluierten Kompentenzklassen

2.2 Methoden der Skills-Evaluierung

Ein weiterer Untersuchungsgegenstand in der Experten-Befragung war die Methodik, mit der die Skills evaluiert wurden[4]. Tabelle 3 fasst die normalisierten Ergebnisse der entsprechenden Fragestellung zusammen.

Methode	**Modus / Partizipant**				**Anzahl Nennungen**
	VG**	PK**	EXT**	Unb.	
Unbekannte Evaluierungsmethode	x				9
	x	x			2
			x		1
Interview	x	x			5
		x	x		1
Interview + Profilevaluierung	x	x			3
QSE-Verfahren				x	1
Collective Mind Methode	x	x			1
Nicht erforderlich (bekannt aus früheren Projekten)					1
Keine Antwort					9
∑					33
Legende: ** VG = ein oder mehrere Vorgesetzte, PK = Projektkandidat, EXT = Externe Berater bzw. Projekt-externe					

Tabelle 3: Verwendete Methodik zur Evaluierung von Skills (ex ante)

[4] Gestellte Frage: „Wenn Sie Frage B1 mit Ja beantwortet haben, nach welchen Verfahren und durch wen wurden die Kompetenzen der Mitarbeiter identifiziert und bewertet?“. Da die möglichen Antworten nicht abgeschätzt werden konnten, wurde ein Freitext-Feld zur Beantwortung der Frage zur Verfügung gestellt.

(Frage B1: „Wurden die Potentiale bzw. Kompetenzen (z.B. Fachwissen etc.) der Mitarbeiter im Vorfeld Ihres wichtigsten SOA-Projektes ermittelt?“.)

In der Tabelle werden die Dimension der Methode und die Dimension des Modus unterschieden. Festgestellt werden kann, dass lediglich in zwei Fällen standardisierte Verfahren eingesetzt worden sind. Zum einen wurde die QSE-Methode genannt, wobei das Akronym nicht weiter spezifiziert wurde und damit offen bleibt, ob es sich beispielsweise um das Quality Software Engineering oder die Qualitätssicherung und -entwicklung handelt. Zum anderen wurde die Collective Mind Methode [KO09] genannt. Sie ist ein Framework, auf dessen Basis weiche Kompetenzen operationalisiert werden können. Die Methode basisiert auf dem etablierten Myers-Briggs Type Indicator (MBTI®) [BM95]. Die übrigen benannten Formen sind überwiegend Interviews in den unterschiedlichen Modi. Der am stärksten ausgeprägte Modus war die Evaluierung durch den Vorgesetzten und den Projektkandidaten zusammen. Am zweithäufigsten wurde eine Evaluierung ausschließlich durch den Vorgesetzen durchgeführt.

In weiteren Fällen traten neutrale Externe in den Evaluierungsprozess ein. Eine Selbstevaluierung der Projektkandidaten wurde in der Umfrage nicht beobachtet.

2.3 Bedeutung der individuellen Kompetenzen bei Auswahl des IT-Dienstleisters

SOA-Projekte werden in den seltensten Fällen ausschließlich in Eigenregie durchgeführt. Meist werden Unternehmen durch IT-Dienstleister unterstützt. Die Auswahl eines Dienstleisters erfolgt i.d.R. auf Grundlage harter Fakten, wobei Kosten-Aspekten immer eine große Bedeutung zukommt. Dies vorausgesetzt wurden nur zusätzliche Kriterien abgeklärt. Oft ist es dem Kunden wichtig, dass die externen Mitarbeiter sich sowohl aus fachlich / methodischer Sicht, als auch von ihrem persönlichen Wesen gut in das Team integrieren können. Der zentrale Untersuchungsgegenstand war die Frage, wie stark ein solches Kriterium bei den untersuchten SOA-Projekten zum Tragen kam. Mit 24,60% aller Antworten wurde deutlich, dass positive Erfahrungen aus früheren Projekten das wichtigste Entscheidungskriterium zur Auswahl des Dienstleisters für das SOA-Projekt darstellte. Neben organisatorischen Kompetenzen, die beispielsweise durch eine Spezialisierung auf Technologien oder Produkten gebildet werden, wurden die personenbezogenen Kompetenzen in der gleichen Häufigkeit genannt (jeweils 18,65%). Dies zeigt sehr deutlich, dass individuelle Kompetenzen durchaus für SOA-Projekte ein nicht unbedeutendes Entscheidungskriterium darstellen, um den IT-Dienstleister auszuwählen.[5]

[5] Gestellte Frage: „Nach welchen Kriterien wurde die IT-Dienstleisterauswahl getroffen?“. Es waren Mehrfachantworten möglich. Als Antworten wurden die Kriterien aus Tabelle 4 vorgegeben. Die Kritieren „Positive Erfahrungen aus früheren Projekten“, „Vertrauen“ und „Zuverlässigkeit“ wurden in der Tabelle aggregiert.

Gründe für die Auswahl des Dienstleisters	Absolute Häufigkeit	Relative Häufigkeit
Positive Erfahrungen aus früheren Projektem, Vertrauen, Zuverlässigkeit	74	55,20%
Unternehmensbezogenes Know-How (z.B. Spezialisierung auf Technologien oder Produkte)	25	18,65%
Personengebundene (individuelle) Kompetenzen (z.B. Skillprofil erfüllt)	**25**	**18,65%**
Planungssicherheit und finanzielle Stabilität	8	6,00%
Organisatorische Befähigung (z.B. CMMI)	2	1,50%
Summe	134	100%

Tabelle 4: Bedeutung von individuellen Kompetenzen für die Dienstleister-Auswahl

2.4 Gründe der Nicht-Evaluierung

Damit die praktische Relevanz von Skills-Management für die SOA-Governance stärker untermauert wird, wurden die Teilnehmer nach den Gründen gefragt, wieso in deren SOA-Projekten keine Skills-Evaluierung (ex ante) stattgefunden hat. Vermutet wurde, dass vor allem Datenschutzgründe bzw. die Betriebsverfassung dies nicht zuließen. Folgende Rangliste zeigt das Antwortergebnis (inkl. absolute Häufigkeit der Nennungen):

- Ausgleich der Kompetenzdefizite durch ‚Learning by Doing' (n = 15)
- Instrument zur Identifizierung, Messung und Bewertung für Tätigkeiten aus dem Service-orientierten Umfeld fehlt (n = 8)
- Betrachtung der Mitarbeiterkompetenzen wurde bisher nicht in Erwägung gezogen (n = 8)
- Kompetenzen waren vorhanden und mussten nicht evaluiert werden (n = 5)
- Betrachtung von Kompetenzen ist überflüssig (n = 1)
- Datenschutz- und Betriebsverfassung (n = 1)
- Sonstige Gründe (n = 10)

Der Hauptgrund, weshalb Mitarbeiter auf Kompetenzebene nicht untersucht worden sind, war die Antizipation, dass sie erforderliches Wissen und erforderliche Fertigkeiten während des Projektes erlernen. Als zweitwichtigster Grund wurde genannt, dass ein passendes Instrument fehlt, um Skills-Management effektiv zu betreiben. Mit der gleichen Häufigkeit wurde es nicht in Erwägung gezogen, Mitarbeiter einer Evaluierung zu unterziehen. Mit fünf Nennungen wurde angegeben, dass eine Evaluierung nicht erforderlich sei, weil die Kompetenzen der Mitarbeiter bekannt sind. Dass die Betriebsverfassung oder der Datenschutz eine Barriere bilden, um effektives Skills-Management zur Personalauswahl für SOA-Projekte zu betreiben, konnte demnach nicht belegt werden. Dieser Grund wurde lediglich einmal genannt.

3 Fazit

Wie in der empirischen Studie belegt werden konnte, werden Ansätze des Skills-Managements im Feld praktiziert. Dabei werden in der Regel meist keine standardisierten Evaluierungs-Methoden verwendet, um Mitarbeiter auf ihre Kompetenzausstattung im Vorfeld eines SOA-Projektes zu untersuchen. Der Autor verfolgt dabei die These, dass ein offenes Service-orientiertes Kompetenzmodell (für neue Rollen / Tätigkeiten) fehlt oder für Unternehmen zu zeitaufwendig bzw. unrentabel zu implementieren ist. Wie die Daten belegen, gibt es das Anwendungsszeanario die Kompetenzen über die Unternehmensgrenzen hinaus zu messen. Ein Modell was diese Anforderungen berücksichtigt, ermöglicht vor Einbindung der Dienstleister deren Human Ressourcen in Kombination mit den internen Ressourcen zu evaluieren und Schlüsse über die Teameignung zu ziehen. Eine standardisierte Evaluierungs-Methodik kann hier helfen die Qualität des Projektergebnisses und damit auch die Güte der SOA zu steigern, was letztendlich auch zu Wettbewerbsvorteilen führt.

Literaturverzeichnis

[BM95] Briggs Myers, I.; Myers, P.B.:Gifts differing: Understanding Personality Type. Davis-Black Publishing, 1995

[BN07] Blutner, D.; Neuhaus, H. M.: Mensch-Maschine-Schnittstellen unter Bewährungsdruck: Kompetenzerwerb bei Einführung neuer IT-Systeme. In: Kompetenzentwicklung in realen und virtuellen Arbeitswelten. Gfa-Press, Dortmund, 2007; S. 123–126.

[DD07] Daum, M.; Durch, M.: Erfolgsfaktoren serviceorientierter Architekturen. In: HMD – Praxis der Wirtschaftsinformatik. dpunkt-Verlag, Heidelberg, 2007.

[Ev07] Bark, M.: Evodion SOA Umfrage 2007: Pragmatische Herangehensweise.ap Verlag, 2007.Verfügbar unter: http://www.ap-verlag.de/OnlineArtikel/SOA/20080708zb%20Evodion%20SOA%20Umfrage%202007.htm, abgerufen am: 25.04.201

[Ka06] Kappelman et al.: Early Warning Signs of IT Project Failure: The Dominant Dozen. In: Information Systems Management. Taylor & Francis, Vol. 23, 2006.; S.31-36.

[KO09] Köhler, J; Oswald, A.: Die Collective Mind Methode. Projekterfolg durch Softskills. Springer-Verlag, Berlin, Heidelberg, 2009

[Ma07] Masak, D.: SOA? Serviceorientierung in Business und Software. Springer-Verlag, Berlin, 2007.

[NR05] Reinhardt, K.; North, K.: Kompetenzmanagement in der Praxis. Mitarbeiterkompetenzen systematisch identifizieren, nutzen und entwickeln. Gabler, Wiesbaden, 2005

[Sc01] Schmidt et al:. Identifying Software Project Risks: An International Delphi Study. In: Journal of Management Information Systems, M E Sharpe Inc, Vol. 17, 2001; S 5-36

[Sc09] Schleidt, B.: Kompetenzen für Ingenieure in der unternehmensübergreifenden virtuellen Produktentwicklung. Technische Universität Kaiserslautern, Kaiserlautern, 2009

[So09] Sommer, M.: Transformation bei Mensch und Maschine. Parallele Umstrukturierungsprozesse in Unternehmen. Am Beispiel der Einführung einer Service-orientierten Architektur (SOA). Meine-Verlag, Magdeburg, 2009.

[SS04] Kompetenzmodelle – Erfolgsfaktoren im HR-Management? Ein strategie- und evidenzbasierter Ansatz der Kompetenzmodellierung. In: Personalführung. DGfP e.V., Düsseldorf, Heft 10, 2004; S. 18-26

[SS07] Schelp, J.; Stutz, M.: SOA-Governance. In: HMD – Praxis der Wirtschaftsinformatik. dpunkt-Verlag, Heidelberg, 2007.

Semantische Analyse zur Unterstützung von SOA-Governance

Michael Niemann, Julian Eckert, Ralf Steinmetz
vorname.nachname@kom.tu-darmstadt.de
Multimedia Communications Lab, Technische Universität Darmstadt
Rundeturmstrasse 10, 64283 Darmstadt

Abstract: Im Hinblick auf sich verändernde Marktbedingungen und die damit verbundenen Flexibilitätsanforderungen ist die effiziente Steuerung und Kontrolle von IT-Systemen auf Basis Service-orientierter Architekturen (SOA) primäres Ziel von SOA-Governance-Frameworks. Die Einführung und Anwendung, d.h., die Bewertung von Prozessen hinsichtlich best-practice-Frameworks, ist in der Regel aufwändig und ressourcenintensiv. Eine Nutzung des Wissens in Governance-Frameworks mit Hilfe von Ontologien birgt das Potenzial, Anwendung und Umsetzung von Governance-Strategien maßgeblich zu vereinfachen und zu beschleunigen. In diesem Beitrag werden Anwendungsmöglichkeiten von Ontologien vorgestellt, welche die Abbildung von Wissen aus SOA-Governance-Frameworks und so die automatische Generierung entscheidungsunterstützender Informationen sowie die Analyse von Governance-Frameworks im Allgemeinen ermöglichen.

1 Einführung

In den letzten Jahren haben sich die Anforderungen und Erwartungen an IT in Unternehmen drastisch verändert. Das Entstehen neuer Märkte und neuer Geschäftsmodelle erfordert erhöhte Flexibilität und Adaptivität der IT-Organisation in Unternehmen [BKR03]. Das Paradigma Service-orientierter Architekturen (SOA) wird als ein möglicher Ansatz für die Bewältigung dieser Herausforderung angesehen [PTD+06]. Ein wesentlicher Erfolgsfaktor von großen IT-Systemen, insbesondere SOA, ist jedoch die Einhaltung regulativer, normativer und rechtlicher Anforderungen [JG07] [NERS08]. Governance-Frameworks wurden entworfen, um Systeme hinsichtlich der Realisierung dieser Vorschriften effizient zu steuern – Vertreter sind CObIT [IT 07], ITIL [Off07], und SOA-spezifische wie das SOA-Governance-Framework der Open Group [The09]. Die Umsetzung dieser Best-Practice-Frameworks in IT-Abteilungen, d.h., die Realisierung und Bewertung der gegebenen Prozesse ist eine ressourcenintensive Aufgabe. In den meisten Fällen kann eine interne Untersuchung von bestimmten Prozessen im Vorfeld wertvolle Einsichten und erste Hinweise darauf liefern, wie eine Verbesserung interner Prozesse erreicht werden kann.

Obgleich sie oft ausreichend Struktur bieten, gibt es nur wenige wissenschaftliche Ansätze, Governance-Frameworks als Datenmodelle darzustellen. Die Modellierung und Nutzung des hierin gesammelten Wissens in Form von Ontologien eröffnet ein weites Feld von automatisierten Anwendungen, insbesondere der deutlichen Verbesserung der Anpas-

sung und Umsetzung eines Frameworks. Verwendet als automatisierte Referenz und als Basis für Modellierungs- und verschiedene Analyse-Methoden kann eine Ontologie dazu genutzt werden, effizient relevante Informationen zu identifizieren – insbesondere solche, welche nicht explizit durch die Framework-Beschreibung gegeben sind. Kombiniert mit semantischen Anwendungen wie semantischer Suche oder text-basierten Vergleichstechniken können darüber hinaus Entscheidungsprozesse unterstützt werden. Die Bereitstellung solcher Techniken in IT-Abteilungen hat das Potenzial, die Bewertung interner Prozesse deutlich zu vereinfachen und Prozessoptimierungen hinsichtlich Kosten, Effizienz, Flexibilität und Zeit weniger resourcen-intensiv zu realisieren.

In diesem Beitrag werden Anwendungsmöglichkeiten von Ontologien im Rahmen von SOA-Governance vorgestellt, welche Framework-Analysen, Framework-Vergleiche sowie die Umsetzung von Governance-Strategien unterstützen.

1.1 SOA-Governance

Das SOA-Paradigma wird als ein ganzheitlicher Ansatz zur Ausführung von Geschäftsprozessen innerhalb von IT-Architekturen angesehen. Als architektonisches Paradigma definiert SOA Mechanismen, Grundsätze und Bedingungen, wobei *services* einen zentralen Aspekt darstellen. Generell realisieren services Wiederverwendbarkeit in verschiedenen Szenarien durch Rekonfiguration. SOA-Systeme erleichtern darüber hinaus die Integration verschiedener IT-Technologien sowie die Abbildung von Wertschöpfungsketten als *value networks* auf IT-Ebene [PTD+06][Jos07][Erl05][HS05]. In den letzten Jahren wurde SOA-Governance als Erfolgsfaktor für die erfolgreiche Anpassung und Betrieb einer SOA erkannt, insbesondere für große Systeme [MB06][NJRS09]. SOA-Governance-Ansätze bieten Mittel, um die Möglichkeiten von SOA effektiv zu nutzen. Sie konzentrieren sich auf die reibungslose Einführung und den erfolgreichen Betrieb einer SOA als IT-Architektur. Durch Befolgung von Leitlinien und best practices, SOA Maturity Modellen und anderen Governance-Mechanismen wird die Integrität des Systems und die Anpassungsfähigkeit an Geschäfts- und Verwaltungsprozesse gewährleistet.

Eines der größten SOA-Governance-Frameworks, die derzeit öffentlich zugänglich sind ist das *SOA Governance Framework der Open Group (OSGF)*[The09]. Die Autoren konkretisieren eine Reihe von wichtigen Elementen wie Prozesse, Rollen, Verantwortlichkeiten, Artefakten, etc. Das OSGF besteht aus zwei großen Teilen: dem *SOA Governance Reference Model* (SGRM), und einem Vorgehensmodell, der *SOA Governance Vitality Method* (SGVM). Die wichtigsten Teile des SGRM sind Leitprinzipien, Prozesse, Artefakte, Rollen und Verantwortlichkeiten sowie Governance-Technologiekompetenzen. Die Prozess-Definition unterscheidet *governing* und *governed* Prozesse. Das Rollenmodell ist umfangreich und beschreibt die Mehrheit der potenziellen Rollen im Rahmen einer SOA und definiert insbesondere Rollegruppen, so genannte *structures*. Das Rückgrat der SGRM wird durch die SGRM Artefakte gebildet. Artefakte dienen als Input-und Output-Dokumente für alle Arten von Aktivitäten, Phasen und Prozesse, die im OSGF definiert sind. Der zweite Teil ist ein Regelkreis ähnlich dem Deming-Zyklus, die SGVM. Sie definiert vier Phasen: *Plan*, *Define*, *Implement* und *Monitor*, welche auf die kontinuierliche Verbesse-

rung von Governance-Strukturen und Methoden durch Wiederholung des Zyklus abzielen. In *Plan* wird der Ist-Zustand der SOA-und SOA-Governance im Unternehmen analysiert und mit Referenz-Architekturen verglichen. In der Phase *Define* werden Prozesse, Rollen-, Transformations-Pläne und Roadmaps definiert. In *Implement* werden diese schrittweise umgesetzt. In der *Monitor* Phase werden die *governed* und *governing* Prozesse überwacht und mittels Metriken bewertet, um zu entscheiden, ob eine Anpassung und Reiteration des Zyklus notwendig wird.

Das OSGF definiert eine umfangreiche innere Struktur. Eine vielversprechende Möglichkeit für die effiziente Nutzung dieses Wissens ist die Modellierung als Ontologie [Gru93].

2 Anwendungsbereiche für semantisch unterstützte Analyse im Bereich SOA-Governance

Governance Frameworks können als semi-strukturierte Ressourcen von Expertenwissen interpretiert werden. Sie beschreiben abstrakte Referenz-Prozesse und Strukturen und werden meist als umfangreiche Textdokumente (vgl. ITIL, COBIT oder OSGF) zur Verfügung gestellt. Einige von ihnen sind gut strukturiert, andere stellen in erster Linie Beschreibungen auf Text-Basis dar. Aufgrund ihrer Komplexität und des ganzheitlichen Anspruchs ist es einem Laien kaum möglich, alle expliziten und impliziten Beziehungen zwischen den zahlreichen Elementen eines solchen best-practice-Frameworks zu erkennen und zu verstehen. Nach Goeken und Alter (2009) wird aus wissenschaftlichem Umfeld wenig Anleitung und methodische Unterstützung für Governance im Allgemeinen zur Verfügung gestellt [GA09].

Die Idee ist nun, dieses strukturierte Fachwissen durch explizite Formalisierung maschinell nutzbar zu machen – ein typischer Anwendungsbereich von Ontologien [MS01]. Die Modelierung von Governance-Frameworks ermöglicht deren Wiederverwendung in verschiedenen Kontexten, z.B. für die Einführung und Nutzung von Frameworks, Prozessbewertungen, den Vergleich von Frameworks, und vielem mehr. Governance Frameworks aggregieren und formalisieren Domänenwissen zu einem hohen Anteil, was für die Erstellung von Ontologien ausgenutzt werden kann. Zudem gibt es häufig wiederkehrende Elemente in Governance Frameworks. Als Hauptanwendungsgebiete einer solchen Governance-Ontologie unterscheiden wir die automatisierte Unterstützung von *Framework-Analysen*, den *Vergleich von Frameworks*, sowie die automatisierte Unterstützung von Aufgaben der *Framework-Anwendung und -umsetzung*.

Die *Framework-Analyse* befasst sich mit der gezielten Erkennung von logischen Schwächen bezüglich der Definitionen der Elemente des Frameworks sowie der Ableitung von impliziten Beziehungen, die in der Beschreibung unerwähnt bleiben. Definitionen aus Governance Frameworks sind explizit und implizit im Text und durch wiederkehrende Strukturen und Beziehungen gegeben – in manchen Fällen werden diese nicht durchgängig in der Spezifikation einhalten. Diese werden in Form von Inferenzregeln modelliert. Als Beispiel sei die Definition von *checkpoint* aus dem OSGF erwähnt. Ein *process* wird hier als zusammengesetzte Aktivität, während eine *activity* als atomare Aktivität definiert wird.

Ein *checkpoint* ist als "eine der zentralen Aktivitäten [...] im Rahmen eines Prozesses" definiert. Allerdings besteht jede explizit in der Spezifikation erwähnte Instanz von *checkpoint* aus weiteren *activities*. Der Definition zufolge sind *checkpoints* nun atomare Aktivitäten, entsprechend der Instanzendefinitionen sind es *processes*. Eine Klassifizierung kann in diesem Fall automatisch logische Inkonsistenz feststellen.

In den Bereichen IT- und SOA-Governance existieren eine Vielzahl von Frameworks. Oft addressieren diese ähnliche Herausforderungen aus einer anderen Perspektive oder auf einer anderen Abstraktionsebene – es existieren eine Reihe von offiziellen "Mappings" (z.B. [IT 08]). Der Einsatz von Ontologien erleichtert die Nutzungs des Wissens vorhandener Zuordnungen sowie deren Ausbau. Eine Domänen-Dachontologie bildet Einheiten wie z.B. *process, activity, role, responsibility, artifact, input, output, goal, metric*, etc. sowie Beziehungen wie *hasInput, hasOutput, realizesGoal, isResponsibleFor*, etc. ab. Diese werden von der Mehrheit der Governance-Frameworks definiert und können daher als allgemeingebräuchlich angesehen werden. Nach dem Schema von Guarino stellt diese eine *domain ontology* dar [Gua97]. So wird es möglich, explizite Elemente *eines* Frameworks (wie z.B. einen Prozess mit Input- und Output-Beziehungen) durch das Modell eines *zweiten* klassifizieren zu lassen. Mit Textverarbeitungstechniken und Metriken für semantische Ähnlichkeit werden die Ergebnisse verfeinert.

Aufgaben der *Framework-Anwendung und -umsetzung* zielen auf die Vereinfachung der Einführung eines Frameworks ab, sowie auf die Erleichterung des Zugangs zu Framework-Wissen. Grundsätzlich wird *ex-ante* und *ex-post*-Analyse unterschieden. Die *ex-ante*- Untersuchung unterstützt die Einführung und erleichtert das Verständnis und den Zugang zu Frameworks. Hauptziele sind ein Verständnis zu vermitteln, z.B. durch das "Surfen" im visualisierten Framework, die aktive Untersuchung von Elementen, Beziehungen und deren Kombinationen sowie die Überprüfung von modellierten Situationen mit Anforderungen (*scenario and impact analysis*), um relevante Organisationsbereiche zur Anwendung sowie sonstige Ansatzpunkte für Teile des Frameworks zu identifizieren.

Ein Beispiel für einen solchen "schnellen Zugriff" wäre die Beantwortung der Frage "Was sind Best Practices, um services interface-compliant zu entwickeln?" Die Antwort, basierend auf einer kombiniert semantischen und Text-basierten Suche, wäre: einen Prozess "service implementation" mit einem Checkpoint "approve service implementation" zu definieren, welche u.a. die Richtlinie "service interface guidelines" verwenden. Nach dem OSGF ist dieser Prozess vom Typ "governedProcess" und Teil der "service lifecycle governance"-Aktivitäten. Eine weitergehende Nutzung umfasst die Auflistung aller tangierten Artefakte sowie die automatisierte Berechnung des Prozess-Artefakt-Workflows, dessen Teil der Prozess ist. Strukturelle Untersuchungen, wie z. B. die Untersuchung von Anfragen, ob es ein wiederkehrendes (implizit) modelliertes Verhältnis zwischen Zielen und Leitlinien gibt, die Identifizierung von aggregierten Artefakten oder Tätigkeiten oder die Berechnung eines Verknüpfungsgrades für einen bestimmten Prozess, und so weiter, sind darauf aufbauend möglich. Alle diese Informationen werden durch die Frameworks definiert, jedoch nicht ausdrücklich in der Dokumentation beschrieben.

Die *ex-post*-Analyse zielt auf die Untersuchung von Situationen und die Simulation von Szenarien und Auswirkungensanalysen ab, um Informationen zur allgemeinenen Entscheidungsunterstützung zu generieren. Diese Analyse basiert auf mit Hilfe der Ontologie mo-

dellierten Situationen. Mittels automatischer Klassifizierung kann die Art und Lage gegebener Elemente bezüglich eines bestimmten Frameworks bestimmt werden. Zum Beispiel kann ein bestimmtes modelliertes Artefakt A anhand der Inferenzregeln, die durch das OSGF definiert wurden, daraufhin untersucht werden, ob es als ein bestimmtes Element intepretiert würde, z.B. als *guideline* oder *aggregated artifact*. Insbesondere kann so automatische zwischen *governing* und *governed* Artifakten unterschieden werden. Die Einstufung eines Elements liefert wertvolle Informationen über dessen Interaktion mit anderen Artefakten oder Prozessen sowie Ähnlichkeiten zu weiteren Elementen. Dies stützt die Bewertung einer Ist-Situation durch das Explizit-Machen von Informationen, welche zusätzlich als Grundlage für weitere Entscheidungen wertvoll sein kann.

3 Zusammenfassung

IT- und SOA-Governance-Frameworks sind bewährte Mittel zur Regulierung von IT im Allgemeinen. Die Dokumentationen bündeln dabei Praktiker- und Expertenwissen. Die Modellierung und automatisierte Anwendung von Wissen aus diesen Governance-Frameworks birgt enormes Potential zur Wissensnutzung. Aufgrund ihrer Fähigkeiten bezüglich *reasoning* und Klassifizierung stellen sich Ontologien als die hierzu vielversprechendste Technologie dar. Wir definieren drei Anwendungsszenarien: die automatisierte Unterstützung von *Framework-Analysen*, den *Vergleich von Frameworks*, sowie die automatisierte Unterstützung von Aufgaben der *Framework-Anwendung und -umsetzung*. Abgesehen von der Identifikation von Ansatzpunkten für die Framework-Anwendung stellen die Bewertung interner Prozesse in Bezug auf Teile des IT-Systems und best practices sowie die Unterstützung von Auditingprozessen die wichtigsten Ziele dar. Ontologien verbessern die Anwendungsmöglichkeiten von Meta-Modell-Strukturen auf diesem Gebiet. Der skizzierte Ansatz ist in der Lage, durch Klassifizierung neue Prozesse (oder andere Elemente) automatisch in die durch die Ontologie definierte Taxonomie einzuordnen – wobei "neu" sich hier auf vom Benutzer angegebene Elemente beziehen kann (*Framework-Analyse und -umsetzung*), oder auf Prozesse, welche durch andere Governance-Frameworks definiert wurden (*Vergleich von Frameworks*). Abgesehen von Zeit- und Kostenersparnissen wird zusätzlich der direkte Zugang zu Expertenwissen vereinfacht. Die weitere Planung umfasst die Modellierung von Governance Frameworks - das OSGF wurde bereits teilweise in OWL umgesetzt, CObIT 4.1 ist in Planung. Insbesondere wurde die Entwicklung einer übergeordneten Taxonomie begonnen. Weitere Schritte umfassen die Umsetzung eines entsprechenden Matching-Algorithmus sowie die Evaluation.

Literatur

[BKR03] J. Becker, M. Kugeler, and M. Rosemann, editors. *Process Management. A Guide for the Design of Business Processes*. Springer-Verlag, 2003.

[Erl05] Thomas Erl. *Service-Oriented Architecture - Concepts, Technology, and Design*. Prentice Hall Professional Technical Reference, Boston, 2005.

[GA09] Matthias Goeken and Stefanie Alter. Towards Conceptual Metamodelling of IT Governance Frameworks. Approach - Use – Benefits. In *Proceedings of the 42nd Annual Hawaii International Conference on System Sciences.*, Waikoloa, Big Island, Hawaii, January 2009.

[Gru93] Tom Gruber. A Translation Approach to Portable Ontology Specifications. *Knowledge Acquisition*, 5(2):199–220, 1993.

[Gua97] Nicola Guarino. Semantic Matching: Formal Ontological Distinctions for Information Organization, Extraction, and Integration. In M. T. Pazienza, editor, *Information Extraction: A Multidisciplinary Approach to an Emerging Information Technology*, pages 139–170. Springer Verlag, 1997.

[HS05] Michael N. Huhns and Munindar P. Singh. Service-Oriented Computing: Key Concepts and Principles. *IEEE Internet Computing*, 9(1):75–81, 2005.

[IT 07] IT Governance Institute (ITGI). *CobiT 4.1: Control Objectives for Information and Related Technology*. IT Governance Institute (ITGI), Rolling Meadows, IL, 2007. `http://www.itgi.org/cobit`.

[IT 08] IT Governance Institute (ITGI). *COBIT Mapping: Mapping ITIL V3 With COBIT 4.1*. IT Governance Institute (ITGI), 2008.

[JG07] Wolfgang Johannsen and Matthias Goeken. *Referenzmodelle für IT-Governance - Strategische Effektivität und Effizienz mit COBIT, ITIL & Co.* dpunkt.verlag Heidelberg, 1. edition, 2007.

[Jos07] Nicolai M. Josuttis. *SOA in Practice: The Art of Distributed System Design*. O'Reilly Media, Inc., 8 2007. http://www.soa-in-practice.com.

[MB06] Eric Marks and Michael Bell. *SOA: A Planning and Implementation Guide for Business and Technology*. John Wiley & Sons, Inc., New Jersey, USA, 2006.

[MS01] Alexander Mädche and Steffen Staab. Ontology Learning for the Semantic Web. *IEEE Intelligent Systems*, 16(2):72–79, 2001.

[NERS08] Michael Niemann, Julian Eckert, Nicolas Repp, and Ralf Steinmetz. Towards a Generic Governance Model for Service-oriented Architectures. In *Proceedings of the Fourteenth Americas Conference on Information Systems (AMCIS 2008), Toronto, ON, Canada*, 2008.

[NJRS09] Michael Niemann, Christian Janiesch, Nicolas Repp, and Ralf Steinmetz. Challenges of Governance Approaches for Service-Oriented Architectures. In *Proceedings of the Third IEEE International Conference on Digital Ecosystems and Technologies (IEEE DEST 2009)*, pages 634–639. IEEE, IEEE Computer Society, Jun 2009.

[Off07] Office of Governance Commerce (OGC). *ITIL v3: Information Technology Infrastructure Library Version 3*, volume 1-5. London: The Stationary Office, 2007. `http://www.itil.org`.

[PTD+06] Michael P. Papazoglou, Paolo Traverso, Schahram Dustdar, Frank Leymann, and Bernd J. Krämer. Service-Oriented Computing: A Research Roadmap. In Francisco Cubera, Bernd J. Krämer, and Michael P. Papazoglou, editors, *Service Oriented Computing (SOC)*, Dagstuhl Seminar Proceedings. Internationales Begegnungs- und Forschungszentrum fuer Informatik (IBFI), Schloss Dagstuhl, Germany, 2006.

[The09] The Open Group. *SOA Governance Framework - Technical Standard*. Number C093. The Open Group, Thames Tower, 37-45 Station Road, Reading, Berkshire, RG1 1LX, United Kingdom, August 2009.

Konzepte der Informationssicherheit in Standards am Beispiel ISO 27001

Danijel Milicevic, Matthias Goeken

IT-Governance-Practice-Network, Frankfurt School of Finance & Management
Sonnemannstr. 9-11, 60314 Frankfurt am Main
{ d.milicevic, m.goeken } @ frankfurt-school.de

Die Dynamik der Informationstechnologie und steigende Komplexität von Informationssystemen schlägt sich auch im Informationssicherheitsmanagement nieder. Informationssicherheitsstandards haben sich als generische Lösungen für eine Vielzahl von Aufgaben etabliert. Inwieweit derartige Standards die benötigte ganzheitliche Betrachtung vorsehen ist bislang nicht geklärt. Im vorliegenden Beitrag wird ein Metamodell des ISO 27001 Standards auf Basis einer qualitativen Datenanalyse erstellt, um die betrachteten Konzepte zu identifizieren und einen Einblick in die Kernelemente des Standards zu liefern.

1 Einleitung

Medienwirksame Sicherheitsvorfälle, Bemühungen von Datenschützern und eine verschärfte Gesetzgebung sorgen dafür, dass die Informationssicherheit eins der wichtigsten Themen für CIOs und IT-Manager darstellt. Während in der Vergangenheit der Mangel derartiger Aufmerksamkeit und Unterstützung durch das Management häufig als eins der größten Probleme benannt wurde, ist mittlerweile die steigende Komplexität ein wichtiger Faktor. Hieraus ergibt sich ein Bedarf nach einem umfassenden Sicherheitsmanagement und der Bedarf nach unternehmens- und branchenübergreifenden Richtlinien und Referenzen führte zur Bildung von Informationssicherheitsstandards. Als ein Bündel von sogenannten Best Practices bieten sie eine Übersicht der facettenreichen Informationssicherheitsdomäne. Einige Forscher stehen derartigen Standards kritisch gegenüber. So stellen bspw. [SW09] ihre Validität und inhaltliche Detailtiefe in Frage. Der vorliegende Beitrag beschäftigt sich damit, wie umfassend Informationssicherheitsstandards die Problemdomäne erfassen.

Der Aufbau von Informationssicherheitsmanagementsystemen sollte einen umfassenden oder gar „ganzheitlichen“ Ansatz verfolgen, falls man Effektivität gewährleisten möchte [BW07]. Zum Zweck der Analyse mit Blick auf „Ganzheitlichkeit“ leiten wir Konzepte der Informationssicherheit aus einem prominenten Standard ab und modellieren Sie auf einer Ebene, welche einen Vergleich mit in der Wissenschaft entstandenen Ontologie zulässt (auf Grund von Platzgründen nicht Teil dieses Beitrages). Zuvor werden die Forschungsmethodik und das Vorgehen bei der Metamodellierung erläutert (Teil 2). Das Metamodell des ISO 27001 wird in Teil 4 beschrieben.

2 Forschungsmethodik

In der vorliegenden Arbeit wird eine theoretische Fundierung über semi-formale Modelle der Informationssicherheit angestrebt. Falls der Untersuchungsgegenstand der Modellierung Modelle sind und nicht etwa Realobjekte, so spricht man von Metamodellen. [St96] betrachtet, wie durch Metaisierung Hierarchien von Modellen konstruiert werden. Das Metaisierungsprinzip beschreibt „denjenigen Aspekt eines Modells, der in der übergeordneten Modellierungsstufe abgebildet wird." Sie betont, dass verschiedene Möglichkeiten bestehen, um von den Instanzen der Wirklichkeit zu Modellen und schließlich zu Metamodellen kommt [siehe auch Le05]. Das Metaisierungsprinzip definiert demnach in einer Modellhierarchie den Abstraktionsmechanismus zur Strukturierung der Objekte der jeweils darunterliegenden Ebene. Beispiele sind die linguistische, ontologische und physische Metaisierung. Für eine detaillierte Darstellung sei an dieser Stelle auf [Le05 und GH05] verwiesen. Im Rahmen der vorliegenden Fragestellung ist insbesondere die ontologische Metamodellierung von Bedeutung, bei der die Abstraktion von Modellelementen gemäß ihrem Inhalt stattfindet.

Metamodelle stellen die zugrundeliegende, häufig implizite, Struktur von Modellen dar. Diese Repräsentation kann je nach Motivation verschiedenartig genutzt werden. Zum einen bieten Metamodelle methodische Unterstützung bei der Erweiterung und Veränderung von Modellen. Berücksichtigt man die Struktur eines Modells bei einer derartigen Adaption, ist eine Konformität mit dem ursprünglichen Modell bzw. Standard wahrscheinlicher. Ebenfalls können Metamodelle zur Integration mehrerer Modelle dienen, als auch Grundlage für die Entwicklung einer entsprechenden Werkzeugunterstützung sein. Im vorliegenden Beitrag ist die Motivation analytischer Natur, so dass die Metamodellierung des Standards als Grundlage für eine Analyse und Evaluation der Vollständigkeit durch Vergleich mit einer Sicherheitsontologie genutzt wird. Bei der Erstellung von Modellen, somit auch Metamodellen, bedarf es einer Sprache und einer Methode im Sinne einer Vorgehensweise, die die Identifikation wie auch Repräsentation der relevanten Objekte im Modell unterstützt. Die Sprache bzw. Sprachauswahl wird auch plastisch „way of modeling" bezeichnet, die Vorgehensweise als „way of working" [VHW91]. Hier werden UML-Klassendiagramme verwendet. Im Folgenden soll der „way of working" detaillierter betrachtet werden.

Als methodische Fundierung des Konstruktionsprozesses von Metamodellen findet im vorliegenden Beitrag eine Anlehnung an die Grounded Theory und die qualitative Datenanalyse (QDA) statt. Aufgrund der Seitenbeschränkung soll hier von einer detaillierten Darstellung abgesehen werden und stattdessen auf die relevante Literatur verwiesen werden, wie z.B. [CS90, Gr95]. Die grundlegende Idee der Grounded Theory ist das Arbeiten mit empirischen Daten wie die Transkription von Interviews, Protokolle und Dokumente, mit welchen der Forscher im Feld konfrontiert ist. Der Fokus liegt dabei auf dem induktiven Generieren von Theorien, welche in den jeweiligen empirischen Daten „verankert" (grounded) sind. Eine zentrale Aktivität in der qualitativen Datenanalyse stellt dabei die Kodierung dar. Kodierung bedeutet hierbei die Konzeptualisierung von Daten und Herleitung von Kategorien, sowie Beziehungen zwischen ihnen. Die in den Daten gefundenen Ereignisse und Instanzen werden

analysiert als potentielle „indicators of phenomena … which are thereby given conceptual labels“ [CS90, p. 7]. Diese Konzeptualisierung ist der vorab beschriebenen ontologischen Metaisierung sehr ähnlich. Im Folgenden wird ein Informationssicherheitsstandard als Ursprungsdokument verwendet und mittels induktiver Kategorisierung in ein Metamodell überführt.

3 Informationssicherheitsstandards

Eine der größten Herausforderungen im Informationssicherheitsmanagement sind unvollständige Informationen über die Risiken, denen Informationssysteme ausgesetzt sind sowie Gegenmaßnahmen. Dies begründet unter anderem die Popularität von Planungsmodellen, Checklisten und Richtlinien als Referenzen im Informationssicherheitsmanagement. Während einzelne Unternehmen Bedrohungen gegenüber ihren Informationssystemen identifizieren und Gegenmaßnahmen abwägen und implementieren, kristallisieren sich erfolgreiche Aktivitäten und Prozesse heraus (Best Practices). Im Rahmen von Standardisierungsbemühungen durch sog. Standard-Setter fließen derartige Best Practices in erste Entwürfe für neue Standards ein. Diese haben selbst im Rahmen der Informationssicherheit unterschiedliche Foki und Ausrichtungen. Bei der Auswahl eines Informationssicherheitsstandards wurden zwei Anforderungen formuliert: 1) der Standard muss eine umfassende und breite Betrachtung der Informationssicherheit als Ziel haben und 2) er sollte - wenn auch beschränkt - eine gewisse Repräsentativität für den tatsächlichen Umgang mit Informationssicherheit in der Praxis darstellen.

Nach einer Evaluation mehrerer Standards fiel die Wahl auf ISO 27001 [In05]. Der ISO-27002-Standard, welcher die eigentliche Best-Practices-Richtlinie für den Aufbau eines Informationssicherheitsmanagementsystems (ISMS) ist, wird in der Regel nicht vollständig adaptiert [LG10]. Der ISO-27001-Standard hingegen ist ein Zertifizierungsstandard. Dadurch ist gewährleistet, dass zertifizierte Unternehmen ihn in Gänze implementiert haben (ggf. mit Erweiterungen). Dies bedeutet im Umkehrschluss, dass der Zertifizierungsstandard, zumindest für zertifizierte Unternehmen, die tatsächliche Praxis im Informationssicherheitsmanagement darstellt.

4 Metamodellierung von ISO 27001

Durch die Wahl des ISO-27001-Zertifizierungsstandards als Untersuchungsgegenstand ist das Primärdokument bereits bestimmt. Als Primärdokument wird im verwendeten Atlas.ti-Werkzeug das zu kodierende Dokument bezeichnet. Eine weitere Eingrenzung findet dadurch statt, dass lediglich der Anhang A „Control objectives and controls“ aus dem Standard kodiert werden soll. Um potentielle linguistische Verzerrung zu reduzieren (Präferenzen des Kodierers gegenüber bestimmten Wörtern) wurde die Entscheidung getroffen im ersten Schritt eine In-vivo-Kodierung zu nutzen, bei welcher der zitierte Begriff selbst als Code dient. Nach der vollständigen Kodierung wurden so 153 Codes generiert, welche in 275 Zitaten verankert (grounded) sind. Diese Menge an

Codes erhielt die Bezeichnung „base set“. Im folgenden Schritt wurden Codes zusammengeführt, welche Wordvariationen oder Synonyme darstellten. Die bereinigte Menge („consolidated set“) enthielt 124 Codes. Bei genauer Betrachtung fiel auf, dass viele Codes lediglich einfach durch ein Zitat verankert waren und gegebenenfalls eher als Attribute unter Modellierungsgesichtspunkten verstanden werden würden. So verkörpern die beiden Codes „information in transit“ und „stored information“ lediglich einen Zustand des Konzeptes „information“, welches wiederrum dem Konzept „asset“ zugeordnet werden konnte. Um die Menge der Codes weiter zu reduzieren und auf die Kernkonzepte abzustellen, wurde die Bedingung eingeführt, dass ein Code durch mindestens zwei Zitate verankert werden muss. Der Prozess und das Resultat („core set of codes“) sind in Abbildung 1 dargestellt.

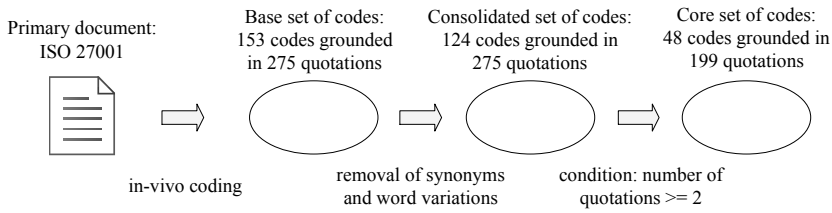


Abbildung 1: Prozess der Kodierung und induktiven Kategorisierung

Die so hergeleiteten Konzepte befinden sich auf der Metamodellebene. Wir definieren Kernkonzepte als Konzepte, welche weder Typen noch Subkonzepte anderer Konzepte sind.

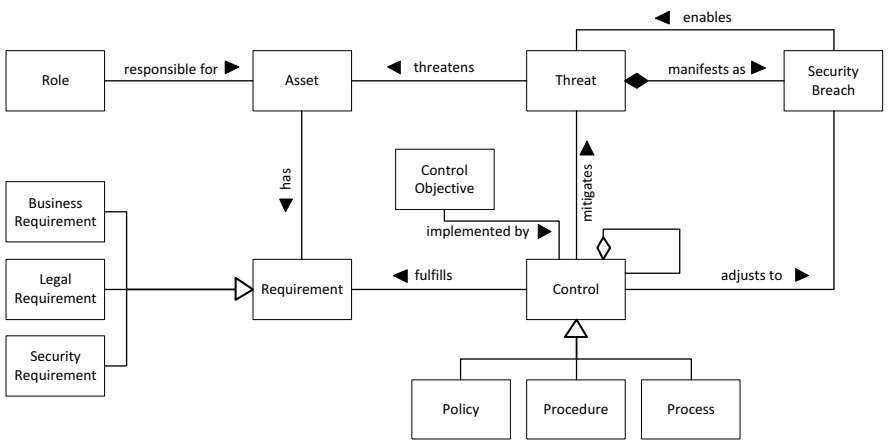


Abbildung 2: ISO 27001 Metamodell

Abbildung 2 zeigt das konstruierte Metamodell des ISO-27001-Standards. Unter den 48 finalen Codes wurden die folgenden Codes als derartige Kernkonzepte identifiziert: „asset“, „threat“, „control“, „requirement“ und „role“. „Assets“, also Vermögensgegenstände, repräsentieren einen Wert, der für die Organisation

schützenswert ist, während „threat“ das Konzept ist, welches diesen Wert gefährdet und „controls“ entsprechend Mittel zur Erreichung des erwähntes Schutzes darstellen.

Das Konzept „requirement“ ist in Form von drei Subkonzepten bzw. Typen vertreten: 1) „security requirement“, 2) „business requirement“ und 3) „legal requirement“. Diese Unterscheidung impliziert eine potentielle Ebenenbetrachtung für die Domäne Informationssicherheit, wie bereits von einigen Forschern suggeriert (siehe [DB01]). Im Vergleich zu den anderen Kernkonzepten finden sich relativ wenige Zitate, welche das Konzept „role“ im Standard verankern. Allerdings ist mit 10 Zitaten das Konzept „responsibility“ stark ausgeprägt, welches die Beziehung zwischen „role“ und „assets“ qualifiziert. Um diesen Indikator für ein Ownership-Paradigma (wie z.B. Process Ownership) nicht zu ignorieren, wurde die Entscheidung getroffen das Konzept „role“ trotz verhältnismäßig schwacher Fundierung mit in das Metamodell aufzunehmen.

Nachdem die fünf Kernkonzepte eingepflegt und „requirement“ und „control“ in ihre Subkonzepte (mit starkem „grounding“ durch Zitate) aufgeteilt wurden, wurden vorher ausgeschlossene Codes mit einfachen Zitaten als Anker re-evaluiert. Hierdurch wurden drei Codes mit einer semantischen Ähnlichkeit identifiziert: „security event“, „security incident“ und „security breach“, welche im Metamodell unter „security breach“ subsummiert wurden. Zwar mag eine Differenzierung über den Grad der Schwere derartiger Ereignisse möglich sein, doch entschied man sich die Codes zusammenzuführen und in das Metamodell aufzunehmen, da eine Analyse der „control“ und „threat“ Konzepte ergab, dass diese Codes eine wichtige Rolle in den Controls A.8.2 (Umgang mit Mitarbeitern, wie z.B. Awareness-Schulungen), A.10.10 (Monitoring) und A.13.2 (Security Incident Management) spielen [In05]. Zuzüglich wurde das Element „control objective“ hinzugefügt, welches eine hohe Bedeutung als strukturelles Element im Standard besitzt. Es ist anzumerken, dass in dem analysierten Teil des ISO 27001 Standards das erwartete Konzept Kennzahlen kaum Beachtung findet, obwohl es eine wichtige Rolle im Rest der ISO 27000-Standardfamilie innehat. Ebenso bestehen kaum Hinweise für eine Zuordnung zwischen den Konzepten „role“ und „controls“ oder „control objectives“. Aus einer Governance-Perspektive wäre die Zuweisung von Verantwortlichkeiten und Entscheidungsrechten bei Implementierung des Standards von hoher Bedeutung.

5 Fazit

Im vorliegenden Beitrag wurden Konzepte der Informationssicherheit hergeleitet. Hierzu fand eine ontologische Metaisierung der Informationssicherheitsmanagementstandards ISO 27001 mithilfe einer induktiven Kategorisierung statt. Hierdurch wurde Einsicht in die Struktur des Standards erreicht. Als methodologisches Werkzeug wurde die Metamodellierung herangezogen und seine grundsätzliche Eignung für Modellvergleiche verargumentiert.

Eine Klärung und Abgrenzung von (ontologischen) Metamodellen und Ontologien fand nicht statt. Aufgrund von Platzgründen wird ein extensiver Vergleich mit mehreren Ontologien in einer Nachfolgepublikation behandelt werden. Die Positionierung von konzeptionellen Modellen, Metamodellen, Referenzmodellen und Ontologien zueinander ist nicht stark erforscht und stellt ein zukünftiges Betätigungsfeld für die Forschung dar. In einer Erweiterung der vorgestellten Arbeit sollen ebenfalls die Instanzen verstärkt betrachtet werden.

Literaturverzeichnis

[BW07] Baker, W. H.; Wallace, L.: Is Information Security Under Control?: Investigating Quality in Information Security Management, IEEE Security and Privacy, 5, 1, 2007, 36-44.

[CS90] Corbin, J. M.; Strauss, A.: Grounded theory research: Procedures, canons, and evaluative criteria, Qualitative Sociology, 13, 1, 1990, 3-21.

[DB01] Dhillon, G.; Backhouse, J.: Current directions in IS security research: towards socio-organizational perspectives, Information Systems Journal, 11, 2, 2001, 127-153.

[GH05] Gitzel, R.; Hildenbrand, T.: A Taxonomy of Metamodel Hierachies. Working paper, Lehrstuhl für ABWL und Wirtschaftsinformatik, 2005.

[Gr95] Grubner, T. R.: Towards principles for the design of ontologies used for knowledge sharing, International Journal of Human-Computer-Studies, 43, 5, 1995, 907-928.

[In05] International Organization for Standardization and International Electrotechnical Commission. Iso/iec 27001:2005, information technology - security techniques - information security management systems- requirements, 2005.

[Le05] Leist-Galanos, S.: Methoden zur Unternehmensmodellierung – Vergleich, Anwendungen und Integrationspotenziale, Logos, Berlin, 2006.

[LG10] Looso, S.; Goeken, M.: Application of Best-Practice Reference Models of IT Governance, Proceedings of European Conference on Information Systems, 2010.

[St96] Strahringer, S.: Metamodellierung als Instrument des Methodenvergleichs. Shaker, Aachen, 1996.

[SW09] Siponen, M.; Willison, R.: Information security management standards: Problems and solutions, Information & Management, 46, 2009, 267-270.

[VHW91] Verhoef, T. F.; Hofstede, A. H. M. T.; Wijers, G. M.: Structuring Modelling Knowledge for CASE Shells. In (Andersen, R. et al. Hrsg.): Proceedings of the third international Conference CAiSE'91 on Advanced Information Systems Engineering, 502-524.

Information-Governance

- Herausforderungen in verteilten Umgebungen -

Wolfgang Johannsen

IT Governance Practice Network
c/o Frankfurt School of Finance & Management
Sonnemannstraße 9-11
60314 Frankfurt am Main
johannsen@its-okay.com

Abstract: Information-Governance - das Management des Vermögensgegenstandes Information im Spannungsfeld von „Performance“ und „Conformance“ - wird zwischen Corporate Governance und IT-Governance positioniert.

Der Beitrag liefert eine Einordnung des Themas insbesondere in Bezug auf die etablierte IT-Governance und die Betrachtung von Aspekten, die sich aus der Verteilung der Informationen über Unternehmens- und Rechtsgrenzen hinaus für eine Information-Governance ergeben.

1. Information-Governance

Informationen werden zunehmend als Vermögenswert eingestuft. Nie zuvor konnten sie in solch großen Mengen zugänglich gemacht, gespeichert, ausgewertet und übertragen werden. Und nie zuvor waren Wirtschaft, öffentliche Verwaltung, Polizei und Militär sowie private Haushalte so stark wie heute von der Verfügbarkeit großer, heterogen strukturierter, verteilter und über vielfältige Medien zugreifbarer Informationsmengen abhängig.

Das Kernproblem unter den Gesichtspunkten des strategischen Alignments, d.h. dem Imperativ der Angleichung der Geschäftsstrategie und IT-Strategie („Performance“), mit dem Ziel der Optimierung des Leistungsbeitrages der IT zum Geschäftserfolg, sowie mit dem Ziel der Compliance („Conformance“), also der Sicherstellung von Konformität mit gesetzlichen und sonstigen regulatorischen Anforderungen, besteht hier in der unkontrollierten Verteilung von Informationen. Dies Problem verschärft sich sprungartig, sobald Verteilung und Diffusion über Unternehmensgrenzen hinausgehen.

In der Bewältigung dieses Problemkreises zeichnen sich zwei alternative Ansätze ab: der Ausbau bisheriger Managementtechniken oder die Etablierung einer gänzlich neuen Information-Governance. In beiden Fällen würden Informationen dann nicht mehr als „Seiteneffekte der Technologie", sondern als der zentrale Gegenstand der Governance selbst behandelt werden. Während der Autor eine Ergänzung und Erweiterung bestehender Referenzmodelle bevorzugt, um einen hohen Integrationsgrad zwischen beiden Governance-Ansätzen zu erzielen, wird gelegentlich auch - wie z.B. von Logan (Gartner Group) [Lo10] - eher der zweite Ansatz favorisiert.

Information-Governance befasst sich mit den strategisch orientierten Fragestellungen im Zusammenhang mit dem Informationsmanagement. Dies soll ganzheitlich unter Einbezug von Prozessen, Menschen, Technologien und Strategien erfolgen.

Selbstverständlich werden die bekannten Methoden der IT-Governance und des IT-Managements sowie die dabei verwendeten Methoden, Prozesse und Referenzmodelle ebenso wenig überflüssig wie Datenbanken, Data Warehouses und ECM-Systeme. Die erweiterten Anforderungen einer Information-Governance zwingen jedoch zu deren Überprüfung und ggfs. zur Erweiterung und Ergänzung.

Abgrenzung Information-Governance und IT-Governance

Wie häufig bei der Entwicklung neuer Themen zu beobachten ist, steht eine begriffsscharfe Definition der Inhalte nicht am Anfang, sondern ergibt sich evolutionär und schrittweise im fachlichen Diskurs. Die bisher vorliegenden Definitionen zum Begriff Information-Governance lassen eine gemeinsame Richtung erkennen, sind jedoch nicht deckungsgleich. Exemplarisch sei hier die recht umfassende Definition der Gartner-Group [Lo10] angeführt:

> Information-Governance umfasst die Festlegung von Entscheidungsbefugnissen und eines „Frameworks" zur Förderung eines gezielten Umganges mit der Bewertung, Erzeugung, Speicherung, Nutzung, Archivierung und Vernichtung von Informationen. Es umfasst Prozesse, Rollen, Metriken und Standards zur Sicherstellung der Effektivität und Effizienz bei der Nutzung von Informationen, um die gesetzten Ziele einer Organisation bestmöglich zu erreichen.

Während also eine Information-Governance den Prozess des unternehmensweiten Informationsmanagements im Fokus hätte [Tu10], steuert die IT-Governance den Prozess des unternehmensweiten IT-Alignments und der unternehmensweiten IT-Compliance [JG07] und ist wesentlich auf die die Beherrschung der Technologien und ihres Einsatzes in Unternehmen ausgerichtet.

Abbildung 1 illustriert die erweiterte Aufgabenstellung der Information-Governance gegenüber der IT-Governance. Dabei wird u. a. auf die Einordnung der ISACA im Cobit-Referenzmodell [Is07] und auf das von Kampffmeyer in [Ka09] eingeführte Schema zum „Electronic Record Management" Bezug genommen.



Abbildung 1 Beziehung zwischen Information-Governance und IT-Governance

Im Kern fordert Information-Governance die Auseinandersetzung mit der Semantik der Daten. Daraus ergibt sich nahezu zwingend die Notwendigkeit zur Beurteilung des geschäftlichen Wertes von Daten und Informationen. Diese liegt außerhalb der klassischen IT-Aufgaben, die vielmehr die Gestaltung und den Betrieb der Systeme, der damit zusammenhängenden Rollen, Verantwortungen und Prozesse und auch der Informationsarchitekturen verlangen.

2. Umsetzung mit Referenzmodellen und Frameworks

Die Umsetzung von Information-Governance wird bisher nicht durch anerkannte Frameworks und Referenzmodelle unterstützt bzw. die vorhandenen decken diesen Bereich nur sehr unvollständig ab. Im Folgenden werden drei Beispiele für Frameworks aufgeführt und diskutiert, die mittel- und langfristig eine Rolle bei der Umsetzung von Information-Governance spielen können.

2.1. IBM Data Governance Council Maturity Model

Eine interessante Herangehensweise an die Aufgabenstellungen innerhalb der Information-Governance zeigt das Reifegradmodell des „IBM Data Governance Council Maturity Model“ [IBM07]. Hier wird eine in fünf Reifestufen geordnete Charakterisierung von Fähigkeiten der Information-Governance gegeben. Damit folgt dieser Ansatz weitgehend dem vom Software Engineering Institute der Carnegie Mellon University (SEI) entwickelten und etablierten Reifegradmodell „Capability Maturity Model Integrated“ (CMMI) [SEI09].

Im IBM-Reifegradmodell wird eine Einordnung in Reifestufen für 11 Fähigkeits-Kategorien (Organisatorische Strukturen, Nachhaltigkeit, Regeln, Werteerzeugung, Risikomanagement, Informationssicherheit, Informationsarchitektur, Datenqualität, Klassifizierung und Metadaten, Lifecycle Management, Audit-Informationen und Berichtswesen) vorgenommen.

2.2. Cobit (Control Objectives for Information and related Technology)

Das Cobit-Referenzmodell in der aktuellen Version 4.1 [Is07] gilt als das als ausgereiftes und differenziertes Rahmenmodell für IT-Governance. Es beschreibt den gesamten IT-Lebenszyklus anhand von vier Aufgabendomänen und 34 sogenannten IT-Prozessen mit zugehörigen Einzelaufgaben und -zielen. Das Rahmenwerk selbst wird durch diverse Dokumentationen zu Begleitthemen wie dem IT-Wertbeitrag und dem Risikomanagement, aber auch zur Praxisunterstützung hinsichtlich Implementierung, Kontrollsystemen und dem Einsatz von Cobit im IT-Assurance ergänzt.

Information-Governance in dem hier vorgestellten Sinn wird von Cobit lediglich teilweise und indirekt unterstützt (vgl. Abbildung 1). Vielmehr steht bisher die ganzheitliche Beherrschung der IT unter den beiden Gesichtspunkten „Conformance“ und „Performance“ im Mittelpunkt. Ersten Veröffentlichungen zum Nachfolgemodell von Cobit 4.1, dem für 2011 zu erwartenden Cobit 5.0, ist zu entnehmen, dass dort der Information eine höhere Bedeutung eingeräumt werden wird.

2.3. ISO/IEC 38500

Der 2008 erschienene Standard ISO/IEC 38500:2008 Corporate Governance of IT [Is08] führte zu einer Positionierung der Internationalen Standardisierungsorganisation (ISO) im Bereich IT-Governance. Der Standard definiert ein High-Level-Referenzmodell, das Führungskräften das Verständnis für die Bedeutung und Umsetzungspflichten hinsichtlich rechtlicher, regulatorischer und ethischer Anforderungen im Bereich der Unternehmens-Governance der IT (Enterprise Governance of IT) verdeutlichen hilft.

Information-Governance als Aufgabenstellung wird vom ISO-Standard 38500 weder explizit noch implizit hinreichend adressiert. Der Fokus liegt nahezu ausschließlich auf den Management-Aufgaben, die sich aus der IT-Infrastruktur ergeben.

3. Reifegrade verteilter Information-Governance

Mit zunehmender Verteilung der IT-Infrastruktur wächst die Komplexität der Information-Governance. Aus dieser Einordnung ergibt sich eine Hierarchie der Management-/Governance-Disziplinen. Das klassische IT-Management bildet darin den mit Technologiefragen befassten Kern und wird in der IT-Governance um eine ganzheitliche und strategische ausgerichtete Sicht auf IT-Compliance und IT-Alignment ergänzt.

Entscheidend für Umfang und Komplexität der Governance-Anforderungen im Cloud-Computing ist der Grad der Öffentlichkeit [Is09], der eine Cloud „ausgesetzt“ ist. Dieser reicht von der, in Eigenregie betriebenen, „Private Cloud“ über die „Group-Cloud“ deren Nutzer-/ Anbieter-Netz klar abgegrenzt ist, bis hin zur „Public Cloud“ in der diese Abgrenzungen aufgehoben sind. Hybride Mischformen sind möglich [Sc09].

Im Cloud-Computing – wie in verteilten Systemen generell – geht typischerweise die Dezentralisierung der Infrastruktur mit der Dezentralisierung der Verantwortung einher. Handelt es sich dann bei den dezentral verarbeiteten Informationen um solche, die für ein Unternehmen einen hohen Wert besitzen (Personaldaten, Produkt- und Verfahrensdaten, Verträge etc.) sind auch die Anforderungen an die IT-Governance und insbesondere an die Information-Governance am höchsten.

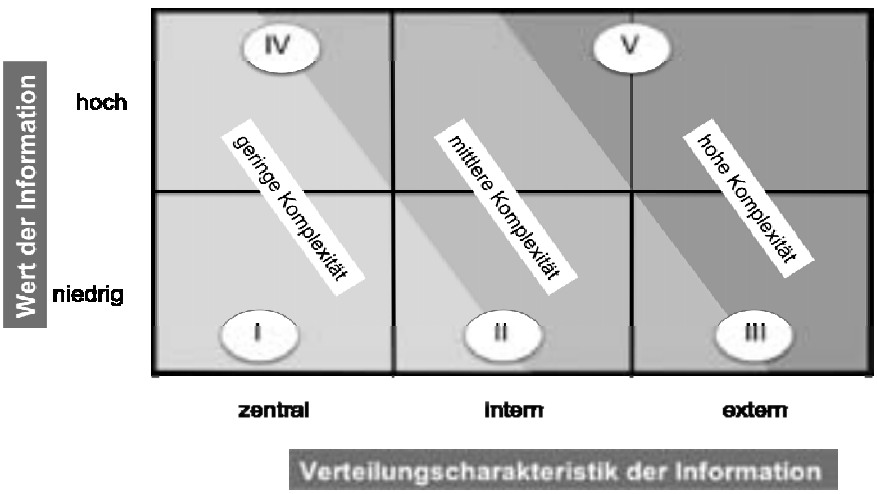


Abbildung 2 Ausprägung der Information-Governance nach Reifegradstufen und Umsetzungskomplexität

In Abbildung 2 werden die Anforderungen, die sich aus dem Wert der Information und ihrer Verteilungscharakteristik ergeben, drei Komplexitätsgruppen (gering, mittel, hoch) und fünf Reifegraden zugeordnet:

Reifegrad I: **Wert der Information gering, keine Verteilung.**

Reifegrad II: **Wert der Information gering, interne Verteilung.**

Reifegrad III: **Wert der Information gering, externe Verteilung.**

Reifegrad IV: **Wert der Information hoch, keine Verteilung.**

Reifegrad V: **Wert der Information hoch, interne und intern/externe Verteilung.**

Diese Klassifizierung in Reifegrade ergänzt die Fähigkeits-Kategorien im „IBM Data Governance Council Maturity Model“ um die Kategorie der „Verteilung“. (Eine nähere Darstellung ist hier aus Platzgründen nicht möglich.) Damit ist ein erster Schritt getan, die Governance-Anforderungen in verteilten Umgebungen (heute der Normalfall) in Abhängigkeit vom Wert der Informationen zu differenzieren.

4. Resümee

Die vorliegende Kurzdarstellung motiviert und diskutiert Information-Governance und mögliche Umsetzungsansätze. Es zeigt sich, dass bisher keine tauglichen Referenzmodelle zur Bewältigung des (verteilten) Information-Governance vorliegen. Durch eine Erweiterung des „IBM Data Governance Council Maturity Model“ wird ein Ansatz zur Einbeziehung der Verteilungsproblematik in dieses Modell vorgestellt.

Sinnvolle weitere Schritte wären nach Meinung des Autors die Erweiterung des Cobit-Referenzmodells und damit einhergehend des Standards ISO 38500 um die dynamische Prozessdarstellungen, um den Lebenszyklus der Informationen vs. der IT-Systeme aktiv steuern zu können.

5. Literaturverzeichnis

[IBM07] IBM: The IBM Data Governance Council Maturity Model: Building a roadmap for effective data governance, (www-935.ibm.com/services/uk), Somers, 2007.

[Is07] ISACA, IT Governance Institute: COBIT 4.1, Framework, Control Objectives, Management Guidelines, Maturity Models, (www.isaca.org.), Rolling Meadows, USA, 2007.

[Is08] ISO/IEC, Corporate governance of information technology, International Standard, 2008

[Is09] ISACA: Cloud Computing: Business Benefits with Security, Governance and assurance Perspectives. In: ISACA Emerging technology White Paper, ISACA, (www.isaca.org) Rolling Meadows, USA, 2009.

[JG07] Johannsen, W.; Goeken, M.: Referenzmodelle für die IT-Governance. dpunkt-Verlag, Heidelberg 2007 (2. Auflage, Sommer 2010).

[Ka09] Kampffmeyer, U.: GRC –Governance, Risk Management & Compliance, IBM Whitepaper, 2009.

[Lo10] Logan, D.: What is Information Governance? And Why is it So Hard? In: blogs.gartner.com, Gartner Group, 2010

[SEI09] CMMI for Services, The Software Engineering Institute, Carnegie Mellon University, 2009

[Tu10] Tuft, B.: Data Governance oder Information Governance? (www.silicon.de), 2010

Workshop

5. Modellbasiertes Testen und Testautomatisierung (MoTes 2010)

Fevzi Belli, Christof J. Budnik, Axel Hollmann, Francesca Saglietti, Mario Winter

MOTES2010 - Modellbasiertes Testen

5. Workshop im Rahmen der Jahrestagung der Gesellschaft für Informatik am 28. September 2010 in Leipzig

Einführung und Übersicht über die angenommenen Beiträge

Fevzi Belli
Universität Paderborn
belli@upb.de

Christof J. Budnik
Siemens Corporate Research
christof.budnik@siemens.com

Axel Hollmann
Universität Paderborn
hollmann@upb.de

Francesca Saglietti
Universität Erlangen-Nürnberg
saglietti@informatik.uni-erlangen.de

Mario Winter
Fachhochschule Köln
mario.winter@fh-koeln.de

1 Einführung

Folgekosten fehlerhafter Software demonstrieren eindrucksvoll, dass die systematische Entwicklung von Software in einer für den Einsatz ausreichenden Qualität auch nach Jahrzehnten Forschung nicht die Regel darstellt. Eine Ursache ist sicherlich in den Projekten selbst zu finden, da häufig bekannte Qualitätssicherungsmaßnahmen entweder gar nicht oder zumindest nicht in der nötigen Form eingesetzt werden, um Termin- bzw. Budgetanforderungen gerecht zu werden. Sicherlich ist aber auch eine Ursache, dass beispielsweise für Software eines speziellen Typs oder für Software, die nach speziellen Implementierungs- und Modellierungstechniken entwickelt wurde, nicht die geeigneten Qualitätssicherungsmaßnahmen zur Verfügung stehen.

Testen ist eine der wichtigen analytischen Maßnahmen der Praxis zur Sicherung der Qualität von Software. Beim Modellbasierten Testen wird die zu testende Software über ein Modell betrachtet, das sich auf besondere Eigenschaften des Prüflings oder seiner Einsatzumgebung konzentriert, meist auf das Verhalten der zu testenden Software. Häufig verwendete Modelltypen sind hierbei endliche Automaten und verschiedene Flussgraphen sowie -diagramme. Testmethoden auf der Basis dieser Modelle lehnen sich beispielsweise an der formalen Verifikation, Kontroll- und Datenflussanalysen wie aber auch dem Model Checking an.

2 Workshopüberblick

In dem fünften Workshop der MOTES-Reihe am 28. September 2010 haben zehn Teilnehmer und Teilnehmerinnen aus Forschung und Praxis im Rahmen der Jahrestagung der Gesellschaft für Informatik in Leipzig ihre Arbeiten in den folgenden Sitzungen ihren Fachkolleginnen und -kollegen vorgestellt und ihre Erfahrungen bei der Anwendung modellbasierten Testens ausgetauscht.

1. Sitzung:
- *System Models vs. Test Models - Distinguishing the Undistinguishable*? Stephan Weißleder, Hartmut Lackner
- *Event-Based Mutation Testing vs. State-Based Mutation Testing – Comparison Using a Web-based System*. Mutlu Beyazit, Timon Sebastian Deistler, Nida Gökçe
- *Modellbasierte Testgenerierung aus Spezifikationen mit parallelem Verhalten*. Jan Krause, Christian Diedrich

2. Sitzung:
- *Model-Based Testing of Automotive HMIs based on a Test-Oriented HMI Specification Model*. Linshu Duan, Heinrich Hussmann, Alexander Höfer
- *Dynamische Äquivalenzklassen im Klassifikationsbaum für Zustandsbehaftete Systeme*. Harald Cichos, Andy Schürr
- *Automatisierter funktionaler Steuergerätetest mit der EXtended Automation Method* (*EXAM*). Dirk Zitterell, Sebastian Thiel
- *Enabling statistical testing for component-based systems*. Thomas Bauer, Robert Eschbach

3. Sitzung:
- *Modelle als Mittel zur Verstärkung von strukturellen Tests sicherheitskritischer Software*. Dirk Wischermann, Wolfgang Schröder-Preikschat
- *Test Case Generation for Visual Contracts Using AI Planning*. Matthias Schnelte, Baris Güldali
- *Evaluierung des Fehlererkennungspotentials modellbasierter Komponenten- und Integrationstestfälle*. Florin Pinte, Francesca Saglietti

3 Programmkommittee

Unser Dank gebührt folgenden Mitgliedern des Programmkomitees, die mit ihren fachkundigen Gutachten der Beiträge wesentlich zum Erfolg des Workshops beigetragen haben:

Colin Atkinson, Klaus Bothe, Jens Braband, Mario Dal Cin, Dimitris Dranidis, Volker Gruhn, Uwe Hehn, Maritta Heisel, Oliver Jack, Bruno Legeard, Sebastian Lehnhoff, Michael Linschulte, Zoltan Adam Mann, Alexander Pretschner, Thomas Roßner, Ina Schieferdecker, Sebastian Senge, Andreas Spillner, Andreas Ulrich, Carsten Weise, Stephan Weißleder, Hong Zhu.

Sommer 2010,
Fevzi Belli, Christof J. Budnik, Axel Hollmann, Francesca Saglietti, Mario Winter

System Models vs. Test Models - Distinguishing the Undistinguishable?

Stephan Weißleder and Hartmut Lackner

{stephan.weissleder|hartmut.lackner}@first.fraunhofer.de

Abstract: Models are purposeful abstractions used in todays software engineering processes. Experience shows that the application of models reduces engineering costs, allows for easy reuse, and improves product quality. Models can be used for several purposes at various levels of abstraction. In this paper, we discuss the differences of system models and test models. For that, we consider several aspects of models and investigate if they can be used as distinctive features of system models and test models.

1 Introduction

Models are the essence of many engineering processes: People create models to understand, analyze, transform, and develop systems. There are several corresponding engineering processes, e.g., model-driven engineering [Ken02] or the Model-Driven Architecture (MDA) [Obj09] defined by the Object Management Group (OMG). The general idea is to create models at an abstract level, to refine them in a step-wise manner using model transformations, and to generate (a part of) the whole system.

Models are used, e.g., to describe systems and tests. Correspondingly, they are referred to as *system models* and *test models*. The best possible use of these models is the automatic generation of the system or the test suite. In this paper, we investigate several aspects of models with the aim to identify distinctive features of system models and test models.

This paper is structured as follows. We present the related work in Section 2. In Section 3, we describe several objective aspects and investigate if they can be used to distinguish system models and test models. We conclude and discuss our findings in Section 4.

2 Related Work

Models for software engineering have been investigated for decades [Har87, Obj07]. They can be used for system creation [Küs06] or testing [UPL06, UL06]. There are only a few comparisons of system models and test models. For instance, Malik et al. [MJV+10] state that test models can only be used for testing. In contrast, one of our findings is that test models can also be used for implementation. Furthermore, Utting and Legeard [UL06] state that test models are usually more abstract than system models. There are, however,

no urgent grounds for it (see Section 3.6). In [GNRS09], the authors state that system models and test models are two different views on the same system - both models can be related using a holistic model. In addition, we describe that both models can be used for the same actions. Utting et al. [UPL06] present a taxonomy of model-based testing, in which they consider several aspects of test models. In contrast, our approach is focused on comparing system models and test models. Furthermore, we also consider a broader range of model aspects and do not restrict our considerations to model-based testing.

3 Aspects of Models

In this section, we present aspects of models and investigate them with respect to their use as a distinctive feature for system models and test models. Figure 3 shows the investigated aspects: On the left side, known characteristics are shown as presented in [UPL06]. On the right side, we present additionally identified aspects.

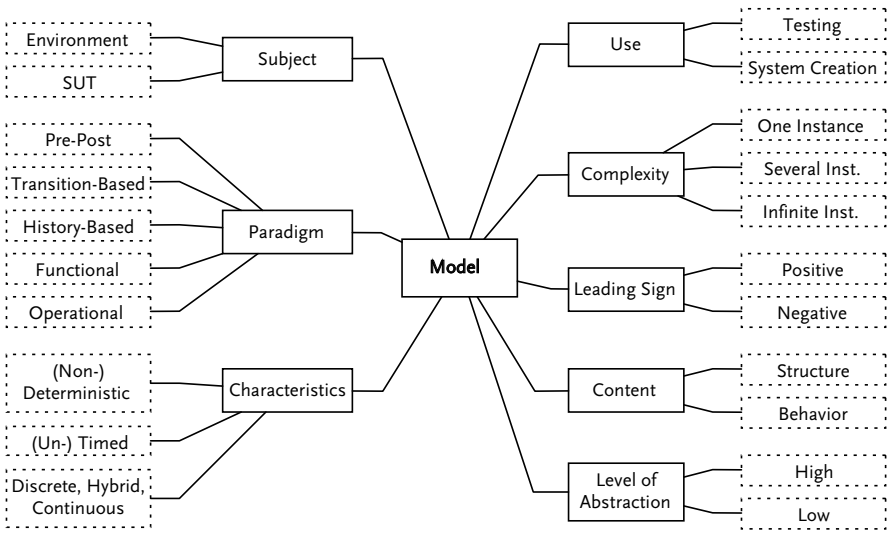


Figure 1: Model aspects.

In this paper, we focus on subject in Section 3.1, use in Section 3.2, complexity in Section 3.3, leading sign in Section 3.4, content in Section 3.5, and level of abstraction in Section 3.6. We leave the remaining aspects to future work.

3.1 Subject

Models are abstractions of an artifact - their subject. This subject is usually (a part of) a concrete system or its environment. Here, we investigate if system models or test models are fixed to one of these subjects.

Constructive system models are often used to create the system and, thus, describe the system. However, system models can also describe the interaction with other components by specifying the expected behavior as observed from outside. Thus, system models can describe the system as well as its environment.

Test models can also describe the system or the environment: The straight forward approach to model tests is to describe test cases, i.e. part of the system environment, at a higher level of abstraction, e.g. using activity diagrams of the Unified Modeling Language (UML). In contrast, tests can also be derived from descriptions of the expected system behavior. Several tools support test generation from system models [Sma, Con, Mic09].

As a result, both system models and test models can be (and actually are) used to describe both the system or its environment.

3.2 Use

Models are intentional abstractions. They are designed for a specific purpose. In this section, we investigate if test models are restricted to be used for testing and if system models are restricted to be used for system design.

System models describe the system and are intended to be used for system creation. They can also be used for testing. For instance, several model-based test tools use system models for automatic test generation [Sma, Con, Mic09].

Test models are designed for testing, i.e. describe tests of a system. A test contains input stimuli and expected system behavior to describe interaction sequences with the system, e.g. using sequence diagrams [Obj07]. Additionally, such sequence diagrams are often attached to use cases that are used for system design. Correspondingly, test descriptions like sequence diagrams can also be used for system design.

As a result, both system models and test models can be used for system design and system test. This is not surprising: Both kinds of models contain information. This information can be used for system implementation or testing.

3.3 Complexity

Models can describe structure or behavior. For instance, the UML [Obj07] defines several structural models, e.g., class diagrams, and behavioral models, e.g., state machines. Structural and behavioral models can allow exactly one possible instance, more than one but a

finite number of instances, and an infinite number of instances. For instance, multiplicities in class diagrams describe the number of possible related objects. The number of loops, the value type of parameters, etc. determine the number of behaviors in state machines.

Both system models and test models can be of arbitrary complexity. A large number of models with a low complexity can have a similar expressiveness as a small number of models with a high complexity. Examples are the following: State machines of high complexity often describe system behavior - they are used for system design and system test. While systems are often easy to generate from state machines, the generation of tests that cover all aspects is undecidable. Less complex behavioral models are often used to describe single tests or to describe the meaning of a use case in system design. Although there may be different challenges for system design and test design based on the complexity of the used model, system models and test models can be of arbitrary complexity in general.

3.4 Leading Sign

Models usually describe the expected (positive) structure or behavior of a system. Additionally, undesired structure or behavior can be described in negative models.

System specifications are based on positive models describing the intended functionality. The opposite, i.e. designing a system solely on negative models, is cumbersome and usually impossible. As a consequence, negative models are used in addition to positive models [SO05, vL04]. For security-relevant systems like, e.g., firewalls, the description of unwanted behavior is of utmost importance for system design.

Most test generators use positive test models to generate tests cases [Sma, Con, Mic09, Wei, GNRS09]. These test generators are constrained to test the specified functionalities of a system. Testing if unwanted system functionalities are prohibited is likewise important. The authors are aware of a comparatively small number of such approaches, e.g., deriving unspecified behavior from positive test models [KL09].

Both system models and test models can have positive or negative leading sign.

3.5 Content

Models can describe behavior or structure. There are various languages for behavioral and structural models. For instance, the UML 2.1 defines six structural diagrams and seven behavioral diagrams. There are other languages like, e.g., Petri Nets, B, Z, or AMSs.

The system model is often based on structure. For instance, structural elements like classes are defined before behavior can be assigned to them, e.g., using state machines. There are, however, also functionality-focused approaches that abstract from structural details using „the system“ as the only structure. Structural as well as behavioral information is crucial for system development. Thus, system models can and should describe both.

Testing can be focused on behavior [UL06] and on structure [SW08]. Thus, test models can be focused on describing behavior or structure, too. The border between structural and behavioral models, however, is not always sharp. For instance, behavioral tests can also be derived from class diagrams that contain operations with pre- and postconditions.

We conclude that system models as well as test models contain both behavior and structure.

3.6 Level of Abstraction

Models can reside at various levels of abstraction regarding detail and completeness. For instance, models can describe a wide range of artifacts from abstract ones like use cases via state machines to concrete models, e.g., containing source code or machine code.

System models can reside at almost every level of detail abstraction. Process models like the MDA [Obj09] prescribe that the process starts with abstract models that are refined in a step-wise manner until source code can be derived. Regarding completeness, system models can range from a complete description to partial descriptions that do not consider, e.g., already implemented library functions.

Like system models, test models can reside at all levels of abstraction regarding detail and completeness: Test models can also be refined in a step-wise manner. They can also be focused on a certain part of the system. Test models are often assumed to be more abstract than system models [UL06]. This, however, always depends on the intended usage: For testing general behavior, implementation details are rather unimportant. For security testing, these details are focused on, e.g., to prevent buffer overflows.

As a result, system models as well as test models can reside at any level of abstraction.

4 Conclusion and Discussion

The focus of our paper is the general distinguishability of system models and test models. We investigated several model aspects with respect to this issue. Our findings are that none of them could be used as a distinctive criterion for system models and test models, i.e. a model cannot be clearly identified as a system model or a test model.

There are several points to discuss. For instance, we investigated the impact of the considered aspects at a formal level and neglected common practices and experiences. Furthermore, the considered aspects were identified based on our experiences. There may be other aspects that provide a distinction for system models and test models. However, our investigations show at least that the presented aspects do not provide a distinction. There may be other applications like, e.g., test generation algorithms, for which it is important if the considered model describes the system or the test.

Acknowledgment. This work was supported by grants from the BMBF (Bundesministerium für Bildung und Forschung) via innovation alliance SPES2020.

References

[Con] Conformiq. Qtronic. http://www.conformiq.com/.

[GNRS09] Helmut Götz, Markus Nickolaus, Thomas Roßner, and Knut Salomon. iX-Studie: Modellbasiertes Testen. http://www.heise.de/kiosk/special/ixstudie/09/01/, 2009.

[Har87] David Harel. Statecharts: A Visual Formalism for Complex Systems. *Science of Computer Programming*, 8(3):231–274, 1987.

[Ken02] Stuart Kent. Model Driven Engineering. In *IFM '02: Proceedings of the Third International Conference on Integrated Formal Methods*, pages 286–298, London, UK, 2002. Springer-Verlag.

[KL09] Kathrin Kaschner and Niels Lohmann. Does my service have unspecified behavior? In Oliver Kopp and Niels Lohmann, editors, *Proceedings of the 1st Central-European Workshop on Services and their Composition, ZEUS 2009, Stuttgart, Germany, March 2-3, 2009*, volume 438 of *CEUR Workshop Proceedings*, pages 22–28. CEUR-WS.org, March 2009.

[Küs06] Jochen M. Küster. Definition and Validation of Model Transformations. *Software and Systems Modeling*, V5(3):233–259, 2006.

[Mic09] Microsoft Research. SpecExplorer. http://research.microsoft. com/en-us/projects/SpecExplorer/, 2009.

[MJV+10] Qaisar A. Malik, Antti Jaaskelainen, Heikki Virtanen, Mika Katara, Fredrik Abbors, Dragos Truscan, and Johan Lilius. Model-Based Testing Using System vs. Test Models - What Is the Difference? *Engineering of Computer-Based Systems, IEEE International Conference on the*, 0:291–299, 2010.

[Obj07] Object Management Group. Unified Modeling Language (UML), version 2.1. http://www.uml.org, 2007.

[Obj09] Object Management Group. Model Driven Architecture (MDA). http://www.omg.org/mda, 2009.

[Sma] Smartesting. Test Designer. http://www.smartesting.com.

[SO05] Guttorm Sindre and Andreas L. Opdahl. Eliciting security requirements with misuse cases. *Requir. Eng.*, 10(1):34–44, 2005.

[SW08] Daniel A. Sadilek and Stephan Weißleder. Testing Metamodels. In Alan Hartman Ina Schieferdecker, editor, *ECMDA'08: European Conference on Model Driven Architecture*. Springer, June 2008.

[UL06] Mark Utting and Bruno Legeard. *Practical Model-Based Testing: A Tools Approach*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2006.

[UPL06] Mark Utting, Alexander Pretschner, and Bruno Legeard. A Taxonomy of Model-Based Testing. Technical Report 04/2006, Department of Computer Science, The Universiy of Waikato (New Zealand), 2006.

[vL04] Axel van Lamsweerde. Elaborating Security Requirements by Construction of Intentional Anti-Models. In *ICSE '04: Proceedings of the 26th International Conference on Software Engineering*, pages 148–157, Washington, DC, USA, 2004. IEEE Computer Society.

[Wei] Stephan Weißleder. ParTeG (Partition Test Generator). http://parteg.sourceforge.net.

Event-Based Mutation Testing vs. State-Based Mutation Testing – Comparison Using a Web-based System

Mutlu Beyazıt[1], Timon Sebastian Deistler[1], Nida Gökçe[2]

[1]Faculty of Computer Science, Electrical Engineering and Mathematics,
University of Paderborn, Germany
beyazit@adt.upb.de, timond@teleos-web.de

[2]Faculty of Arts and Science, Muğla University, Turkey
on the leave at University of Paderborn, Germany
goekce@adt.upb.de

Abstract: The idea of using models in mutation analysis dates back to mid 1980s. Recently, it is shown that, with some proper adaptations, mutation testing can also be utilized as a black-box testing approach. For this purpose, a modeling structure, a set of mutation operators, coverage criteria and test case generation methods are required. This paper compares event-based and state-based realizations of this testing approach over a case study and achieves experimental and comparative results, while validating the applicability of the discussed notions.

1 Introduction

Mutation testing [DLS78] can be defined as a white-box testing technique based on generating faulty versions (mutants) of software by introducing small but typical faults using mutation operators. It aims at evaluating adequacy of test sets by their ability to reveal these mutants (or faults). On the other hand, using mutated models of the system [BBW06] can be deployed as a black-box testing technique, and test generation process can also be evaluated. For example, the original model can be mutated to generate test cases to be executed on the system. Doing so, a mutant is said to be *live* if all its test cases yield the expected output when executed on the system, and it is said to be *killed* if these outputs differ. Failures are also expected to be found during test execution. However, typically, many mutants can be generated, resulting in a massive testing effort, and, mutants, in terms of behavior, can be equivalent to each other.

The objective of this paper is to compare state-based and event-based mutation testing approaches. The state-based approach, discussed here, makes use of finite state automata [HMU01], more specifically Moore finite state machines (MFSMs) [Mo56]. They focus on states, events and outputs of these events in a modeled process. On the other hand, the event-based approach, modeled here by event sequence graphs (ESGs) [Be01], focus only on events of a process. When compared to MFSMs, ESGs are easier to learn and apply, therefore it is intuitive and common to think that MFSMs are more powerful in

terms of detecting the failures/faults. We will check the validity of this argument whether it does necessarily hold by making use of, and with respect to, the following aspects:

(1) *Modeling structure.* As already mentioned, the present paper makes use of ESGs [Be01] and MFSMs [Mo56].

(2) *Mutation operators.* Here, different, but structural, mutation operators are selected for ESGs and MFSMs. Only (a subset of) first order mutants are created, inserting small changes, i.e., not every mutant is generated or used.

(3) *Coverage criteria.* Coverage criteria are used to measure how well the system is exercised by a test suite [Bi00]. In present paper, two different coverage criteria, arc coverage and transition coverage, are used for ESGs and MFSMs, respectively.

(4) *Test case generation.* Minimum set of test cases, more specifically test sequences achieving the above mentioned coverage, are generated by solving slightly different variants of Chinese Postman Problem [EJ73] over the respective underlying graphs of ESG and MFSM models.

The rest of the paper is outlined as follows: Sections 2 and 3 explain the elements used in event (ESG) and state (MFSM) based approaches respectively. Then, in Section 4, the introduced concepts are applied over a case study and the results are given. Finally, in Section 5, the paper is concluded, also making some comparative remarks.

2 Event Sequence Graphs

Definition. An *event sequence graph (ESG)* can be used to represent a system's behavior interacting with user's actions. Nodes of an ESG are behavioral events of the system and arcs are sequences of these events. Thus, an ESG is a tuple $G = (N, A, [,])$ where

- N is a finite set of *nodes* or *events*,
- $A \subseteq N \times N$ is a finite set of *arcs* or *event sequences*,
- $[\in \mathrm{N}$ is a unique pseudo *start* node and, for $n \in N$, $(n, [) \notin A$, and
- $] \in \mathrm{N}$ is a unique pseudo *finish* node and, for $n \in N$, $(], n) \notin A$.
- each node is *useful*, i.e., for $n \in N$, n is reachable from start node and finish node is reachable from n.

Note that the start node has no incoming arcs and the finish node has no outgoing arcs. They are simply used to mark entry and exit events in a system. For that matter, the arcs connected to these nodes are often referred to as pseudo arcs, whose number is ≥ 2.

Mutation Operators. *3* mutation operators are defined for ESGs as follows: (1) *Insert arc (iA) operator* adds a new non-pseudo arc to the model, (2) *delete arc (dA) operator* removes an existing arc from the model, and (3) *reverse arc (rA) operator* reverses the direction of an existing non-pseudo arc in the model (See Figure 1). Application of dA and rA operators may invalidate the ESG model or the usefulness of some nodes. Thus, following formulae give only some upper bounds on the number of first order mutants:

- #iA_mutants = (#nodes - 2)^2 + 2 (#nodes - 2) - #arcs
- #dA_mutants = #arcs
- #rA_mutants = #arcs - #pseduoarcs - #loops - #symmetricarcs

where a symmetric arc is an arc whose reversion is also an arc.



(a) Original ESG (b) iA Mutant (c) dA Mutant (d) rA Mutant

Figure 1: Sample ESG Mutants

Test Generation. While generating test sequences from ESG models, the arc coverage criterion is quite a sensible choice. This criterion entails coverage of all event sequences of length *2*. In order to generate a test set achieving arc coverage from a given ESG, a variant of Chinese Postman Problem is solved. Thus, a minimum test set is constructed.

3 Moore Finite-State Machines

Definition. *Moore finite state machines (MFSMs)*, in contrast to ESGs, focus on states of the system and represent the events by including them in transitions. They also include outputs in the states. A MFSM is a tuple $M = (S, S0, E, O, T, G)$ where

- S is a finite set of *states*,
- $S0 \in S$ is a *start state* or *initial state*,
- E is a finite set of *input symbols* or *events*, called the *input alphabet*,
- O is a finite set of *output symbols*, called the *output alphabet*,
- $T: S \times E \rightarrow S$ is a *transition function*, and
- $G: S \rightarrow O$ is an *output function*.

Although, a MFSM does not have a set of final states, here, it is assumed that they have a non-empty set of final states F (to mark exit events or final states) and each state in the machine is useful. Also, note that, by definition, MFSMs are deterministic.

Mutation Operators. MFSMs are manipulated by the following *3* mutation operators: (1) *Insert transition (iT) operator* adds a new transition to the model, (2) *delete transition (dT) operator* deletes a transition from the model, and (3) *reverse transition (rT) operator* reverses an existing transition in the model (See Figure 2). Note that when compared to their ESG counterparts, MSFM mutation operators use state information to some level. Also, application of these operators may produce an invalid MFSM in the sense that the resulting FSM may be non-deterministic or some states may lose their usefulness. Thus, following formulae can be used to determine only some upper bounds on the number of first order mutants produced using the defined operators:

- #iA mutants = (#events) (#states)^2 - #transitions
- #dA mutants = #transitions

- #rA mutants = #transitions - #loops - #symmetrictransitions

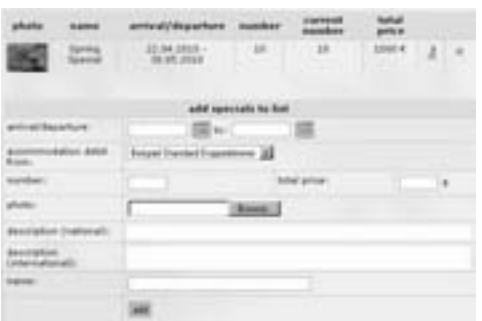
where a symmetric transition is a transition whose reversion is also a transition.



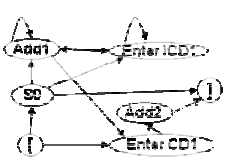
(a) Original FSM (b) iT Mutant (c) dT Mutant (d) rT Mutant

Figure 2: Sample MFSM Mutants

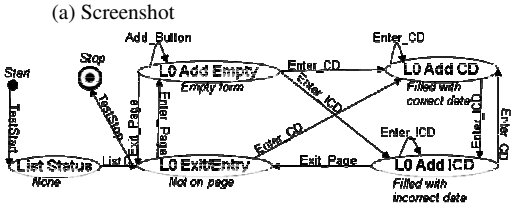
Test Generation. Transition coverage criterion is selected for generation of test sequences. Consequently, the problem of producing minimized test sets satisfying this criterion turns out to be another variant of Chinese Postman Problem. The additional feature of MFSMs is that, using the output function, it is possible to provide extra information such as suspected output in order to aid the test oracle during test execution.



(a) Screenshot



(b) Partial ESG Model



(c) Partial MFSM Model

Figure 3: ISELTA Specials Module – Main Interface and Simplified Models

4 Case Study

ISELTA is an online reservation system for hotel providers and agents. It is a cooperative product of the work between a mid-size travel agency (ISIK Touristik Ltd.) and University of Paderborn. For our case study, we will consider "Specials Module" of ISELTA (Figure 3). Through this module, in the given hotel, one is able to add special prices to the specified number of rooms of certain type for the determined period of time. Consequently, one can edit existing specials and also remove them. In our case study, testing of this module is carried out by two students separately.

Specials module is modeled using ESGs and MFSMs (See Figure 3 for partial models). The ESG model has *3* levels and it is a regular approximation to the system. In order to guarantee a valid approximation, some system-attributed events, e.g., S0 or "there are no existing specials" event in Figure 3(b), are not interpreted in test execution. They are used to differentiate between contexts, like "existing number of specials is *0*, *1*, or *≥2*" or "a special is being edited". On the other hand, MFSM model turns out to be another regular approximation to the system. It has *2* levels and, unlike ESGs, states are used to differentiate between contexts. In addition, thanks to the output function, corresponding test sequences also provide suspected outputs to guide the tester during test execution.

In order to test the system, aforementioned mutation operators are applied to generate mutants and these mutants together with original models are used to generate test sequences to be executed on the system to detect failures. Table 1 gives data on the number of mutants, which are selected differently based on the individual preferences (Numbers with respect to different mutants are not included in order to save space).

Table 1: Number of Mutants

Model	Upper Bound	Selected	Live	Killed	Equivalent
ESG	1507	120	44	76	0
MFSM	30040	126	24	102	8

Test execution results show that the majority of live mutants are dA mutants (for ESGs) and dT mutants (for MFSMs). Since greater number of dA mutants is used for ESGs, fewer mutants were killed. This can be explained by the fact that, in our case, dA or dT mutants are just correct sub models of the system. Thus, based on the used test generation methods, these mutants tend to yield test sequences which are relatively less effective in revealing failures or mutants. Furthermore, it is observed that the upper bound on the number of 1st order MFSM mutants is quite larger than that of ESG mutants. This stems from the fact that application of iT operator results in, as expected, a really great deal of mutants (approximately 96% of all mutants) because the operator includes states, increasing the number of possible operator applications and, in general, the testing effort.

Table 2: Detected Failures

No.	Brief Explanation
1	A special can be inserted with missing start date.
2	A previous date can be inserted for a special.
3	Departure date is not autocorrected.
4	Departure date is falsely autocorrected.
5	Today button does not work.

During test execution, killed mutants are able to detect the failures given in Table 2. Except for failure 5, both ESG-based and MFSM-based approaches manage to reveal the same failures (Failure 5 could not be detected using MFSMs, because today button was not included in the model). The results imply that, despite being simpler in structure, ESGs are able to achieve the same failure/fault-detection capability as MFSMs.

5 Conclusion

In this paper, mutation testing of a web-based system is carried out using an event-based and a state-based approach, i.e., using ESGs and MFSMs, respectively. To do this, regular approximations to the system are created, while taking particular measures to handle the execution of test sequences properly.

The results indicate that both approaches, contrary to the common belief, have similar failure-detection power when appropriately used. The advantage of ESGs is that they focus on only system events. Therefore, in an application with relatively large number of states, ESGs are easier to build and handle. However, direct use of state information is rendered impossible. On the other hand, FSMs include the states explicitly. Thus, they can easily be used to deliver auxiliary information like suspected outputs and provide assistance to the test oracle. Still, due to the inclusion of states, the length of the produced test sequences tend to be longer and mutation operators may easily lead to production of quite a large number of mutants.

In our case study, for both approaches and, naturally, with respect to the selected mutation operators, coverage criteria and test generation methods: (1) Repeated test sequences are observed in the produced test sets, and (2) the mutants produced by the applications of delete arc or delete transition operators are found to be less effective, or even ineffective, in failure/fault detection. Nevertheless, it still seems possible to obtain relatively less redundant test sequences by using higher order or different mutants, and, furthermore, the deletion operators can still prove to be more effective if they are used in combination with different coverage criteria and/or test sequence generation methods. In addition, it is naturally possible that the results may vary for other (type of) systems.

As a final remark, the testing approach used in this paper is a work in progress and being actively studied in our research group.

References

[BBW06] Belli, F.; Budnik, Ch., J.; Wong, E.: Basic Operations for Generating Behavioral Mutants, In: Proc. 2nd Workshop on Mutation Analysis in conjunction with ISSRE'06. IEEE CS, 2006; S. 10-18.

[Be01] Belli, F.: Finite-State Testing and Analysis of Graphical User Interfaces, Proc. of Int. Symp. on Softw. Reliability and Eng., IEEE Comp. Press, 2001; S. 34-43.

[Bi00] Binder, R.V: Testing Object-Oriented Systems. Addison-Wesley; 2000.

[DLS78] DeMillo R. A.; Lipton R. J.; Sayward F. G.: Hints on Test Data Selection: Help for the Practicing Programmer, Computer, v.11, no. 4, April 1978; S. 34-41.

[EJ73] Edmonds, J.; Johnson, e.L.: Matching: Euler Tours and the Chinese Postman, Math. Programming, 1973; S. 88-124.

[HMU01] Hopcroft, J. E.; Motwani. R.; Ullman, J. D.: Introduction to Automata Theory, Languages and Computation, 2nd Edition, Addison-Wesley, Massachusetts, 2001.

[Mo56] Moore, E. F.: Gedanken-experiments on Sequential Machines. Automata Studies, Annals of Mathematical Studies, Princeton University Press, 1956; 34, S. 129-153.

Modellbasierte Testgenerierung aus Spezifikationen mit parallelem Verhalten

Jan Krause, Christian Diedrich

ifak - Institut für Automation und Kommunikation e.V. Magdeburg
Werner-Heisenberg-Str. 1
39106 Magdeburg
jan.krause@ifak.eu, christian.diedrich@ifak.eu

Abstract: In diesem Beitrag wird eine Methode zur Generierung von Testfällen aus modellbasierten Verhaltensspezifikationen vorgestellt, wobei auch die Spezifikation von parallelem Verhalten unterstützt wird. Zur Realisierung werden etablierte Verfahren der Petri-Netz-Theorie und der constraintbasierten Programmierung benutzt. Auf der Grundlage eines entwickelten einfachen farbigen Petri-Netz-Modells (SPeNAt) können damit Testfälle mit parametrierbarem Abdeckungsgrad generiert werden, wobei auch sehr hohe, insbesondere für die Entwicklung sicherheitskritischer Systeme relevante, Abdeckungsgrade (wie z. B. „alle möglichen alternativen Pfade“, MC/DC an den Bedingungen, etc.) garantiert werden können.

1 Einleitung

In den letzten Jahren wurde eine Reihe von Methoden zur Testgenerierung entwickelt. Grundlage dieser Methoden ist immer eine formale Spezifikation bzw. ein formales Spezifikationsmodell. Meist arbeiten diese entwickelten Testgenerierungsmethoden nicht auf dem gesamten Zustandsraum des Spezifikationsmodells, wodurch auch die Testgenerierung für komplexe Modelle ermöglicht wird. Im Regelfall wird das Modell mit Eingaben gefüttert und die Modellreaktion wird protokolliert, wodurch sich das gewünschte Verhalten bzw. die Testfälle ergeben. Die Modelleingaben werden durch genetische Algorithmen (z. B. in [PSO08]), intelligente Heuristiken (z. B. in [SW07]) und/oder durch Methoden der constraintbasierten Programmierung (z. B. in [PL01]) ermittelt. Diese Prozedur wird solange fortgesetzt, bis die gewünschten, vorher spezifizierten Testziele (z. B. Abdeckungsgrad der Spezifikation) erreicht werden. Allerdings können mit diesen Methoden meist nur strukturelle Spezifikationsabdeckungen wie „alle Zustände“ bzw. „alle Transitionen“ erreicht werden. Höhere Abdeckungsgrade (z. B. „alle möglichen alternativen Pfade“) können nicht garantiert werden, da nicht auf dem gesamten Zustandsraum des Spezifikationsmodells operiert wird. Gerade in der Entwicklung sicherheitsrelevanter Systeme ist diese erreichte Spezifikationsabdeckung mitunter zu gering. Auch sind Testgenerierungsmethoden, die explizit die Spezifikation parallelen Verhaltens (auch auf Komponentenebene) unterstützen, den Autoren nicht bekannt. Durch die zukünftig erwartete Zunahme von mehrkernigen Systemarchitekturen wird allerdings die Spezifikation parallelen Verhaltens auch auf Komponentenebene vermehrt benötigt werden.

Die in dieser Arbeit vorgestellte Methode zur Testfallgenerierung erreicht höhere Spezifikationsabdeckungen (u. a. „alle möglichen alternativen Pfade“) der generierten Testfälle. Sie basiert und erweitert die Arbeiten aus [Ulr98] und [KHD08]. Erreicht wird die hohe Spezifikationsabdeckung durch die Analyse des Spezifikationsmodells mit bekannten und etablierten Methoden der constraintbasierten Programmierung und der Petri-Netz-Theorie, wodurch explizit die Beschreibung von parallelem Verhalten in der modellbasierten Spezifikation (auch auf Komponentenebene) unterstützt wird. Dafür wurde ein spezielles (Low-Level-)Petri-Netz-Modell (Sicheres Petri-Netz mit Attributen – SPeNAt) entwickelt, auf welches das Spezifikationsmodell abgebildet werden muss.

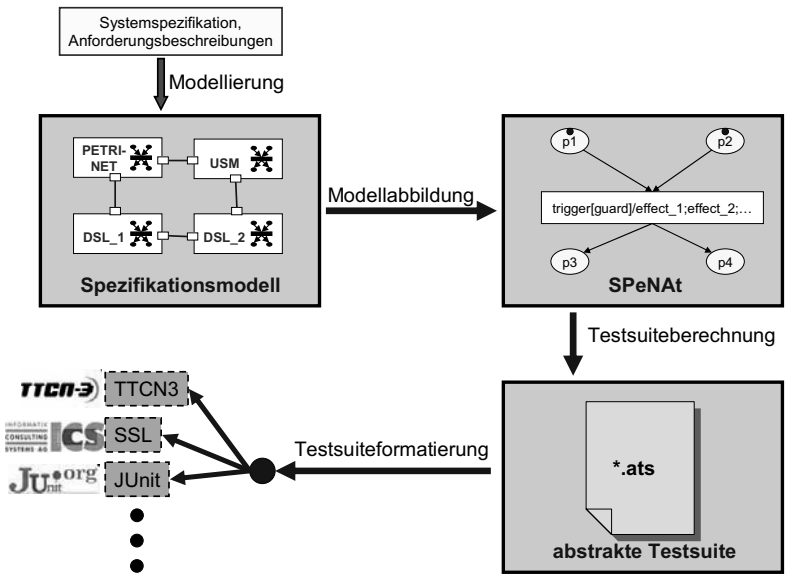


Abbildung 1: Schritte zur Testgenerierung und –formatierung

In Abbildung 1 werden die benötigten Schritte zur Testgenerierung und Testformatierung übersichtlich dargestellt. Alle Schritte werden durch entsprechende Software-Werkzeuge unterstützt, wobei etablierte Modellierungswerkzeuge und prototypische Implementierungen des ifak benutzt werden können. Auf der Grundlage der Systemspezifikation und/oder der beschriebenen Anforderungen wird zunächst das geforderte Systemverhalten modelliert und durch ein Spezifikationsmodell beschrieben. Dieses Spezifikationsmodell bildet die Grundlage für die Testgenerierung und spätere Testformatierung. Daher ist es von immenser Wichtigkeit, dass das Spezifikationsmodell die Systemspezifikation semantisch korrekt abbildet. Der Nachweis dieser Eigenschaften kann z. B. mit Methoden der formalen Verifikation (u. a. Model Checking [CGP99]) erfolgen, was in dieser Arbeit allerdings nicht weiter berücksichtigt wird. Vielmehr wird davon ausgegangen, dass das Spezifikationsmodell korrekt ist und zur Testgenerierung verwendet werden kann. Für die Modellierung des geforderten Systemverhaltens können dabei verschiedene Notationen wie z. B. Petri-Netze, UML-Zustandsmaschinen (USM), UML-Aktivitätsdiagramme und/oder domänenspezifische Sprachen (DSL) verwendet werden.

In der Praxis hat sich dabei die UML-Zustandsmaschine für die Modellierung reaktiven System- bzw. Komponentenverhaltens etabliert, wobei auch paralleles Verhalten mit einer UML-Zustandsmaschine modelliert werden kann.

2 Sicheres Petri Netz mit Attributen (SPeNAt)

Grundlage der entwickelten Methode zur modellbasierten Testfallgenerierung ist ein spezielles Petri-Netz-Modell (Sicheres Petri Netz mit Attributen – SPeNAt). Das Modell eines SPeNAts ist sehr einfach gehalten, wodurch Abbildungsvorschriften für gängige und etablierte Modellierungsnotationen wie z. B. UML-Zustandsmaschinen, UML-Aktivitätsdiagramme, DSLs möglich sind. Aktuell sind Abbildungsvorschriften für ein Spezifikationsmodell bestehend aus UML-Zustandsmaschinen definiert und in [KHD08] näher beschrieben.

Bei einem SPeNAt handelt es sich um ein sicheres Stellen/Transitionen Netz (siehe z. B. [ERV96]), welches um einige Aspekte erweitert wurde. So kann ein SPeNAt Attribute besitzen, die jeweils durch eine Datenstelle repräsentiert werden und durch einen (Daten-)Typ und einen eindeutigem Namen charakterisiert sind. Die Transitionen erhalten eine Beschriftung, die in Syntax und Semantik der Beschriftung einer Transition eines UML-Zustandsautomaten entspricht. Das bedeutet im Einzelnen, dass einer SPeNAt-Transition ein (externes, parametriertes) Ereignis, ein Guard sowie eine Menge an Aktionen zugeordnet sind. Eine SPeNAt-Transition schaltet, wenn ihre Vorgänger-Stellen markiert sind, ihr zugeordnetes (externes) Ereignis erscheint und ihr Guard zu *true* ausgewertet wird. Schaltet eine SPeNAt-Transition, werden ihre zugeordneten Aktionen ausgeführt und ihre Nachfolger mit entsprechenden Markierungen belegt.

In Abbildung 2 ist ein Beispiel für ein SPeNAt vor und nach dem Schalten einer Transition angegeben. Hier wird auch eine weitere Eigenschaft eines SPeNAts dargestellt. Datenstellen sind mit einer Transition immer durch eine Schlinge verbunden und dadurch immer markiert, da sie auch Teil der Anfangsmarkierung sind. Ob eine Datenstelle mit einer Transition verbunden ist, wird durch die Beschriftung der Transition (Guard und Aktionen) determiniert. Die Markierungen einer Datenstelle können durch verschiedene Werte des Typbereichs der Datenstelle erfolgen. Dieses Konzept für eine Petri-Netz-Markierung wird in der Literatur auch „farbige Token“ (siehe insb. [Jen09]) genannt, weshalb ein SPeNAt auch als einfaches farbiges Petri-Netz betrachtet werden kann.

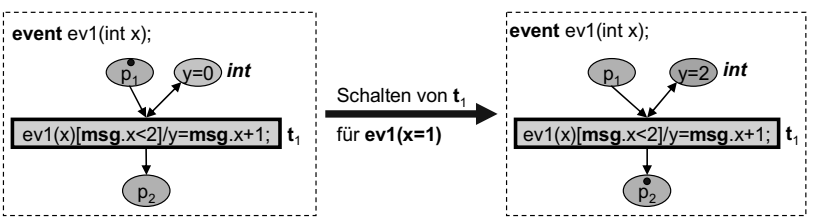


Abbildung 2: SPeNAt vor und nach dem Schalten einer Transition

3 Testgenerierung auf Grundlage der Spezifikationsmodellanalyse

Die Identifizierung der Testfälle wird auf Grundlage der Verhaltensanalyse des Spezifikationsmodells nach dessen Abbildung auf ein SPeNAt durchgeführt. Das Verhalten eines SPeNAt wird dabei durch die sukzessive Konstruktion des Entfaltungsnetzes berechnet. Dafür wird der Algorithmus aus [ERV96] verwendet und für die speziellen Eigenschaften eines SPeNAts angepasst [KHD08]. Die Arbeiten in [KHD08] wurden dahingehend erweitert, dass nun auch externe Ereignisse mit Parametern berücksichtigt werden können. In [KHD08] konnte das entsprechende Petri-Netz-Modell zwar auch schon Attribute verwenden (z. B. an den Guards der Transitionen), allerdings konnten nur externe Ereignisse ohne Parameter die Transitionen schalten. Erreicht wird diese Erweiterung durch die Verwendung von Methoden der constraintbasierten Programmierung, d. h. bei der Überprüfung der Schaltbarkeit einer Transition wird neben der benötigten Markierung der entsprechenden Vorgängerstellen auch ein Bedingungserfüllungsproblem (Constraint-Satisfaction-Problem – CSP) aufgestellt und ausgewertet. Dadurch können auch diese Informationen in der Präfixberechnung und damit in der darauf aufbauenden Testfallgenerierung mit berücksichtigt werden. Durch entsprechende umfassende Modellierung des geforderten Verhaltens im Spezifikationsmodell (z. B. Berücksichtigung des Fehlerverhaltens durch Modellierung mit *else*-Transitionen) werden somit Abdeckungsgrade nach dem v. a. bei der Entwicklung von sicherheitskritischen Systemen geforderten Kriterium MC/DC (siehe u. a. [TSWP10]) an den Bedingungen der SPeNAt-Transitionen erreicht.

Beim Entfaltungsnetz – präziser dem kompletten Anfangsteil des Entfaltungsnetzes (Präfix) – eines Stellen/Transitionen-Netzes handelt es sich um ein zyklenfreies, einfaches Bedingungen/Ereignisse-Netz [ERV96], das alle möglichen erreichbaren Markierungen des ursprünglichen Petri-Netzes codiert. Im Gegensatz zum Erreichbarkeitsgraphen eines Petri-Netzes werden aber im Entfaltungsnetz alle parallelen Ausführungen beibehalten, wodurch der Zustandsraum insbesondere bei parallelem Verhalten deutlich kompakter dargestellt wird [ERV96].

Die Eigenschaft des Entfaltungsnetzes, alle möglichen alternativen Prozesse zu enthalten [ERV96], wird bei der Testfallidentifizierung ausgenutzt. Wird jedem Prozess ein Testfall zugeordnet, wird ein Spezifikationsabdeckungsgrad „alle möglichen alternativen Pfade“ erreicht, welcher insbesondere bei der Entwicklung von sicherheitskritischen Anwendungen benötigt wird. Auf der Grundlage der sukzessiven Konstruktion des Entfaltungsnetzes können aber auch andere (einfachere) Testziele, wie z. B. Abdeckungsgrad „alle Transitionen“ o. ä., leicht identifiziert und somit auf den jeweiligen Anwendungsfall maßgeschneiderte Testfälle erstellt werden.

Ein sehr einfaches funktionales Anwendungsbeispiel wird in Abbildung 3 vorgestellt, um die Funktionsweise der Testfallgenerierung auf der Basis eines Spezifikationsmodells zu demonstrieren. Durch das Spezifikationsmodell werden drei Aktivitäten (Prozesse) durch entsprechende SPeNAt-Transitionen beispielhaft spezifiziert, wobei zwei Aktivitäten parallel ausgeführt werden können, aber vor der dritten Aktivität beendet sein müssen.

Insgesamt ist für dieses einfache Beispiel nur ein Testfall für eine vollständige Spezifikationsabdeckung notwendig (siehe rechten Teil aus Abbildung 3), wobei die konkreten benötigten Eingabeparameter aus der Auswertung des jeweiligen CSP berechnet werden. Durch die Verwendung der Entfaltungskonstruktion des SPeNAt-Spezifikationsmodells für die Testfallgenerierung kann das spezifizierte parallele Verhalten für die Testfallbeschreibung beibehalten werden. Somit sind bei gleicher Spezifikationsabdeckung deutlich weniger Testfälle notwendig als bei der Verwendung von Testfällen, die nur sequentielles Verhalten beschreiben können.

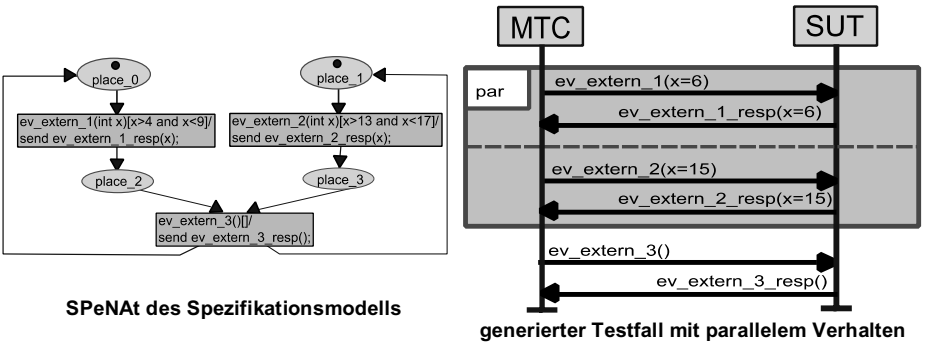


Abbildung 3: Funktionales Anwendungsbeispiel

An diesem Punkt besteht ein Testfall der generierten Testsuite aus einer Sequenz von SPeNAt-Transitionen mit entsprechenden Eingabeparametern, wobei Transitionen auch parallel in einem Testfall ausgeführt werden können. Für die konkrete Testrealisierung muss diese abstrakte Testsuitebeschreibung in ein Format gebracht werden, das im entsprechenden Testprozess verwendet werden kann. Dazu sind Informationen nötig, die sich aus dem jeweiligen Einsatzkontext ergeben. So müssen z. B. Datentypen und/oder Ereignisse bzw. Signale, die im Spezifikationsmodell auf einem eher abstrakten Niveau definiert wurden, auf verwendbare Strukturen der Testnotation des verwendeten Testwerkzeugs abgebildet werden.

Ein bekanntes und in der Praxis immer mehr etabliertes Format zur Testbeschreibung ist die durch die ETSI standardisierte *Testing and Test Control Notation* (TTCN-3), für die es mittlerweile auch mehrere Werkzeuge zur Testrealisierung gibt. Bei der Verwendung von TTCN-3 als Testnotation können die generierten Testsequenzen durch einfache Abbildungsregeln erzeugt werden. Allerdings müssen dann die Modellelemente auf die vorliegenden Gegebenheiten des SUT (Systems Under Test) bei der Implementierung des TTCN-3-SUT-Adapters angepasst werden, um eine automatische Realisierung des Tests zu ermöglichen. Die Spezifikation von parallelem Verhalten innerhalb eines Testfalls wird auch in TTCN-3 aktuell nicht vollständig unterstützt (partiell durch das *interleave*-Statement), ist aber in einer der nächsten Versionen (durch ein *par*-Statement) vorgesehen, so dass für die Testrealisierung aktuell noch auf nicht standardbasierte Werkzeuge zurückgegriffen werden muss.

4 Zusammenfassung

In diesem Papier wurde eine Methode zur Testfallgenerierung auf der Grundlage formaler Spezifikationen unter Berücksichtigung von parallelem Verhalten vorgestellt. In Abhängigkeit von den spezifizierten Testzielen können mit dieser Methode sehr hohe Spezifikationsabdeckungen (z. B. „alle möglichen alternativen Pfade“, MC/DC-Abdeckung an den Bedingungen) erzielt werden. Erreicht wird das durch die Verwendung eines einfachen, farbigen Petri-Netz-Modells (SPeNAt) und dessen Analysemöglichkeiten durch die Kombination von etablierten Methoden der Petri-Netz-Theorie und der constraintbasierten Programmierung. Falls erforderlich, wird auf dem gesamten Zustandsraum operiert, um die gewünschten Abdeckungsgrade zu erreichen. Dabei wird als Zustandsraumkodierung das komplette Präfix des Entfaltungsnetzes verwendet, der zum Einen eine kompakte Darstellung des Zustandsraumes erlaubt und zum Anderen eine einfache Identifizierung aller möglichen alternativen Prozesse – und damit der Testfälle – der Verhaltensspezifikation ermöglicht.

Literaturverzeichnis

[CGP99] E. M. Clarke, O. Grumberg, D. A. Peled: Model Checking, MIT-Press, 1999.

[ERV96] J. Esparza, S. Römer, W. Vogler: An improvement of McMillan's unfolding algorithm, In Proceedings of TACAS'96, pages 87-106, 1996.

[Jen09] Jensen, K.: Coloured Petri Nets: Modeling and Validation of Concurrent Systems, Springer Verlag, Berlin, 2009

[KHD08] J. Krause, A. Herrmann, Ch. Diedrich: Test case generation from formal system specifications based on UML State Machines. atp - International 01/2008, Oldenbourg-Verlag, 2008.

[PSO08] F. Pinte, F. Saglietti, N. Oster: Automatic Generation of Optimized Integration Test Data by Genetic Algorithms in Software Engineering, Workshop proceedings of MOTES'08, 2008.

[PL01] A. Pretschner, H. Lötzbeyer: Model Based Testing with Constraint Logic Programming: First Results and Challenges, Proceedings of WAPATV'01, Toronto, 2001.

[SW07] H. Schlingloff, S. Weißleder: Deriving Input Partitions from UML Models for Automatic Test Generation; In: Models in Software Engineering, Holger Giese (Ed.), Springer LNCS 5002, 2007.

[TSWP10] http://www.testingstandards.co.uk/glossary.htm, 2010.

[Ulr98] A. Ulrich: Testfallableitung und Testrealisierung in verteilten Systemen, Dissertation "Otto-von-Guericke"-Universität Magdeburg, 1998.

A Test-Oriented HMI Specification Model for Model-Based Testing of Automotive Human-Machine Interfaces

Linshu Duan
AUDI AG
85055 Ingolstadt
Germany
linshu.duan@audi.de

Heinrich Hussmann
Ludwig-Maximilians-
Universität München
80333 Munich Germany
heinrich.hussmann@
ifi.lmu.de

Alexander Höfer
AUDI AG
85055 Ingolstadt
Germany
alexander.hoefer@audi.de

Abstract: While model-based testing is widely-used for function tests, testing the human-machine interface (HMI) remains a manual, demanding and time consuming task. Numerous research efforts have been addressing model-based HMI testing in the last years. However the existing approaches cannot serve our test purposes or needs. In the research project at AUDI AG a model-based testing approach is proposed for testing both the logical behavior and the graphical interface of automotive infotainment HMIs. In this paper we focus on menu behavior tests and introduce some important details of the test-oriented specification model. A test-oriented HMI specification model describes the expected menu behavior and contains needed information for the test generation. Methods of test generation and execution will be introduced briefly.

1 Introduction

The complexity of infotainment HMIs is growing with the functionalities of infotainment systems [Sys05]. A user interface giving a faultless experience is one of the most important requirements of today's infotainment systems. During the development phases, 50 percent of the time and more than 50 percent of the costs are expended for testing [MS04]. While testers of non-HMI applications have been enjoying the convenience of tools to automate their tests, GUI testers still toil in their tedious works. Difficulties of HMI testing compared to function tests are fully discussed in [CYM10]. In infotainment testing area, HMI tests are usually distinguished from function tests. Accordingly we distinguish HMI errors from function errors. A function error is e.g. deleted entries are still existing, while HMI errors can be classified in three categories: menu behavior error, widget behavior error and display error. Accordant tests detecting these errors are menu behavior tests, widget behavior tests and display tests [DHH10]. In [Ben10] and [Ben07] task models are proposed for generation of integration tests which belong to functions tests. In our research a model-based approach for testing the infotainment HMI is proposed. Testing goals, specific coverage criteria and evaluation methods are introduced in previous paper [DHH10]. In this paper we focus on the menu behavior tests and introduce in Section 3 the test-oriented specification model for the menu behavior tests. Test generation and execution will be introduced in Section 4. Firstly related researches are discussed.

2 Related Work

In the previous paper [DHH10] some related works in GUI application testing were introduced. It was explained why these existing approaches cannot fulfill our testing demands. Automotive manufacturers and suppliers have also been engaged in model-based testing of infotainment HMIs for some years. In [Fle07] a model-based HMI prototyping tool is suggested from the prospect of a supplier to manufacturers for a model-based specification. One of the motivations is that a model-based HMI specification should allow model-based HMI testing. In [SBK05] discontinuities are regarded as the reason why model-based testing is still not accomplished for infotainment HMIs. Numerous stake-holders participate in the development process and each of them has his own perspectives, approaches and tools. To resolve this problem a HMI framework, which is quite similar to the one in [Fle07], is proposed for the whole HMI development process: HMI specification, implementation and testing. In [GB09] domain specific languages (DSLs) are defined for each layer of an infotainment HMI. That means the HMI specification is composed of models described with the defined DSLs. Such specification should allow automated test generation and execution.
To accomplish model-based testing for infotainment HMIs, the mentioned approaches are looking for solutions in the HMI framework. However our experience has shown that the absence of a suitable HMI framework is not the core problem. There exist varied problems, e.g. the process problem and designer vs. IT-engineer problem [PVH07]. In our opinion, another problem is the lack of a complete model-based HMI testing approach covering the whole testing process, which can serve as a pilot solution. A pilot solution, which is proved to be possible and useful, is the best motivation for the management to resolve the process problem or other no-technical problems.

3 Test-Oriented HMI Specification Model

A test-oriented HMI specification model is a model which describes the expected HMI behavior and contains sufficient information for testing. Depending on the HMI development process, a test-oriented HMI specification model could be created by HMI designers and test engineers corporately or only by test engineers in the test generation phase on the bases of a normal HMI specification. A normal HMI specification is a product of the HMI design phase, it could be informal or model-based. As shown in Figure 1, a test-oriented HMI specification model can be made with the information contained in the normal HMI specification and additional information needed for testing which is defined by the test engineers. In this case the test-oriented HMI specification model does not have to own a visual form, it could be a composition of information in the memory. We have created a normal HMI specification and a test-oriented specification model with a HMI prototyping tool in order to show them visually.

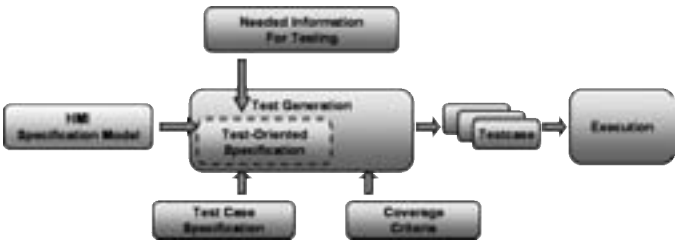


Figure 1: A Potential HMI Testing Process

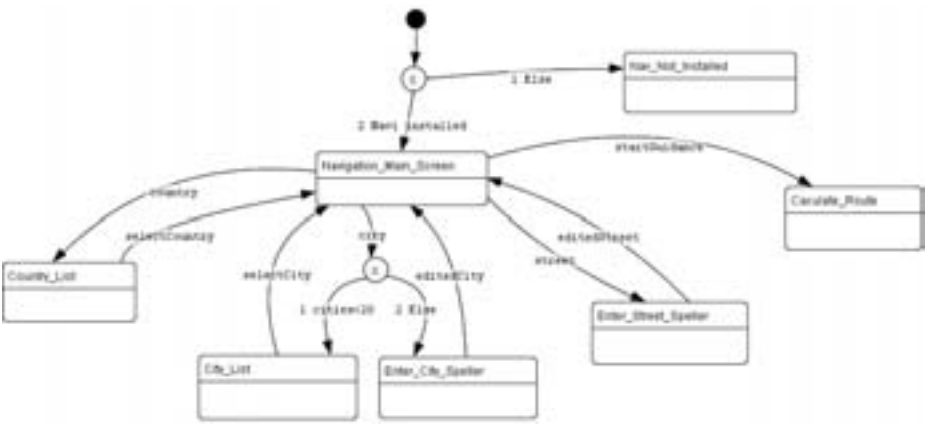


Figure 2: Simplified Cutout of a normal HMI Specification



Figure 3: Screen and Widgets

3.1 A normal HMI Specification

An HMI specification is usually composed of two parts according to the two layers of an HMI: i)Statecharts in Figure 2 defining the menu behavior. The menu behavior is how the HMI switches from one screen to another as reaction to incoming inputs. ii)Graphical elements i.e., screens and widgets shown in Figure 3 defining the HMI looks.

Figure 2 demonstrates a strongly simplified cutout from the statecharts of a normal HMI specification. The statechart cutout describes the simplified menu behavior of the feature navigation. Each state in the diagram is a view state, i.e., associated with a defined screen. As soon as the HMI enters a view state the according screen will be displayed. The HMI

stands in *Navigation_Main_Screen* when the feature navigation is accessed. Event action-reaction details of widgets are contained in the graphical layer, but not in the statechart. The first left outgoing transition in the diagram is executed by the event *country* which is fired by the button widget as soon as the user has pressed the button. The HMI returns from the state *Country_List* to the main screen, when the user has selected a country. From the main screen it is possible to enter a city as destination. If the selected country has less then 20 cities, the screen *City_List*, from which the user can directly select the expected city, will be entered. Otherwise, the user will be led to *Enter_City_Speller* which offers a speller to the user for character inputs. So the next state after the event *city* depends on the selected country and the used navigation database. By a complete user input the route guidance can now be started. The HMI accesses the screen *Calculate_Route*.

3.2 A Test-Oriented HMI Specification Model

A normal HMI specification defines only the expected behavior of an HMI. Tests can not yet be generated from this specification. In this paper two mostly important information needed for testing will be introduced. Firstly, in a normal specification many important conditions are not accessible for the test generation. The statechart in Figure 2 defines the possibility to reach *City_List* from the main screen. However, the condition to do this is hidden in the widget. The widgets used in the specification phase are usually implemented by the supplier and are available as program libraries. The hidden logic is that as long as the widget city is active, it fires the event *city* as reaction to the keystroke. The widget city is only active if the underlying application, which is not a part of the model-based specification, has evaluated the earlier selected country as valid and then set the widget active. This condition is hidden in the widget library and is not accessible for test generation. Similarly start guidance can only be performed if the supplied destination was valid. Without these conditions many invalid tests will be generated, e.g., starting the route guidance immediately from the main screen without any destination inputs. Secondly, contents of user inputs e.g., the city name to be entered do not exist in the specification. Generated tests are not complete without these user input contents.

Figure 4 shows the result after adding the lacking information into the specification. Actions and conditions are inserted into some of the transitions. For example, to be able to enter a city the condition *countrySelected==true* must be satisfied. This condition can only be satisfied by the action of the transition from *Navigation_Main_Screen* to *Country_List*. This condition constraints that the country must be selected at first. The conditions will be considered by the test generation going to be introduced in the section 4. In this way, it is ensured to generate correct and usable tests and to avoid invalid tests. Some self-transitions indicating the needs for user input contents are inserted into the diagram. Section 4 will introduce how user inputs can be integrated into generated tests.

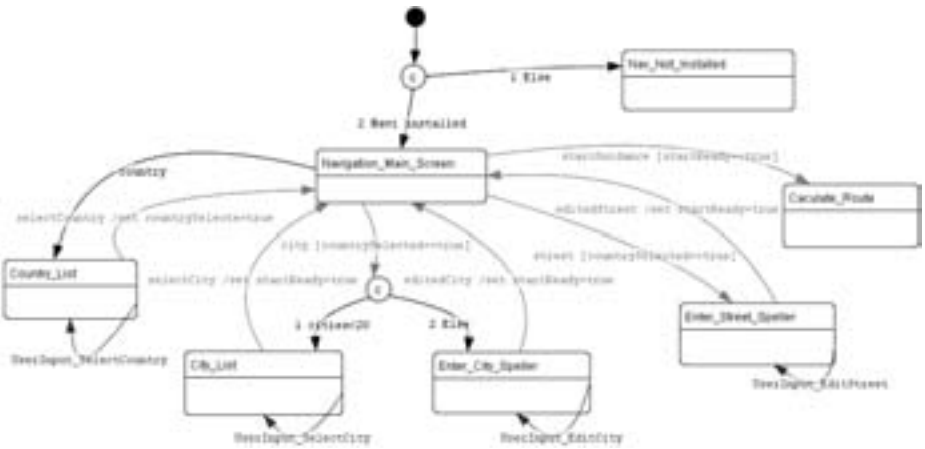


Figure 4: Sample of a Test-Oriented HMI Specification Model

4 Test Generation and Execution

In order to generate tests, test case specification has to be defined by testers and a test coverage criterion has to be selected for the forthcoming generation. Test case specification defines the contents and ranges of tests to be generated, e.g., the states and transitions a test should cover. Test generation algorithms need this information to select tests from a test-oriented HMI specification model. In our work, infotainment system specific coverage criteria are identified. Our previous paper [DHH10] has introduced them in detail.
Before a transition labeled with a condition is selected as the next test step, first of all it has to be verified whether actions of other transitions make the condition fulfilled. For example, to select the transition with event *city*, the transition with event *country* has to be selected as previous step. Transitions with *UserInput* events will be selected as steps before all other outgoing transitions from the same state. We explained that the event *country* is the reaction event fired by the widget country. So the test generation selects the action event *enter on widget country* as the test step instead of the reaction event *country*. Generated tests contain both test steps and expected screens or state names.
In the previous paper test tools for infotainment system function tests were introduced. Tests generated from the test-oriented HMI specification model can be executed automatically with the help of such test tools. Test actions performing needed user input contents have to be defined in the test tools before. They will be executed as sub-sequence before the rest of the generated tests is executed. Image processing is intensively used to locate the correct widget on the screen and verify whether the HMI stands in the correct menu as expected. Details of test execution, observation and verification will not be explained in this paper to avoid exceeding the paper range.

5 Conclusion

The purpose of our research is to find a model-based testing approach for infotainment HMIs. We focused on the testing of the logical behavior in this paper and introduced a test-oriented specification model which contains both the expected behavior of the HMI and sufficient information for testing purposes. Tests can be generated from the test-oriented specification model and automatically executed with the help of test tools for infotainment system function tests. In future papers, we will introduce details about testing the widget behavior, displaying effect as well as the test generation and execution. We believe that we have provided a practical and complete model-based testing approach for infotainment HMIs whose methods could also be used for other advanced HMIs.

References

[Ben07] Sebastian Benz. Combining test case generation for component and integration testing. In *A-MOST '07: Proceedings of the 3rd international workshop on Advances in model-based testing*, pages 23–33, New York, NY, USA, 2007. ACM.

[Ben10] Sebastian Benz. *Generating Tests for Feature Interaction*. Dissertation, Technische Universität München, München, 2010.

[CYM10] Tsung-Hsiang Chang, Tom Yeh, and Robert C. Miller. GUI testing using computer vision. In *CHI '10: Proceedings of the 28th international conference on Human factors in computing systems*, pages 1535–1544, New York, NY, USA, 2010. ACM.

[DHH10] Linshu Duan, Alexander Hoefer, and Heinrich Hussmann. Model-Based Testing of Automotive HMIs based on a Test-Oriented HMI Specification Model. In *Proceedings of the the 2nd International Conference on Advances in System Testing and Validation Lifecycle (VALID 2010), August 22-27 2010, Nice, France*. IEEE, August 2010.

[Fle07] Thomas Fleischmann. Model Based HMI Specification in an Automotive Context. In *Human Interface and the Management of Information. Methods, Techniques and Tools in Information Design*, volume 4557/2007, pages 31–39. Springer Berlin / Heidelberg, 2007.

[GB09] Holger Grandy and Sebastian Benz. Specification based testing of automotive human machine interfaces. In *GI Jahrestagung*, pages 2720–2727, 2009.

[MS04] Glenford J. Myers and Corey Sandler. *The Art of Software Testing*. John Wiley & Sons, 2004.

[PVH07] Andreas Pleuss, Arnd Vitzthum, and Heinrich Hussmann. Integrating Heterogeneous Tools into Model-Centric Development of Interactive Applications. In *MoDELS*, pages 241–255, 2007.

[SBK05] Reinhard Stolle, Thomas Benedek, and Christian Knuechel. Model-based Test Automation for Automotive Human Machine Interfaces. Technische Universität Berlin - Forschungsberichte der Fakultät IV, Bericht-Nr. 2005-1, ISSN: 1436-9915, 2005.

[Sys05] QNX Software Systems. Ready for the Future: Software Trends for In-Car Infotainment Systems. 2005.

Dynamische Äquivalenzklassen im Klassifikationsbaum für zustandsbehaftete Systeme

Harald Cichos und Andy Schürr
Fachgebiet Echtzeitsysteme
Technische Universität Darmstadt
Merckstraße 25
D-64283 Darmstadt
{harald.cichos,andy.schuerr}@es.tu-darmstadt.de

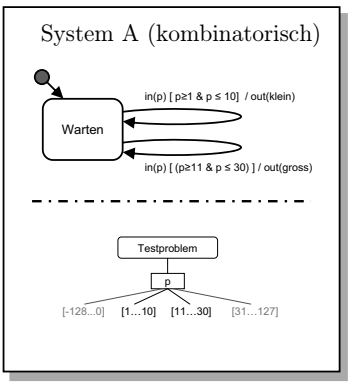
Abstract: Die Klassifikationsbaummethode ist eine weit verbreitete funktionsorientierte Methode zum Test von kombinatorischen Systemen. In der vorliegenden Arbeit wird eine Erweiterung der Klassifikationsbaummethode vorgestellt, die es ermöglicht, gültige und konkrete Testsequenzen auch für zustandsbehaftete Systeme abzuleiten. Zu diesem Zweck werden die im Klassifikationsbaum enthaltenen Äquivalenzklassen in Abhängigkeit vom Zustand des verhaltensbeschreibenden Testmodells aus diesem dynamisch erzeugt.

1 Einleitung

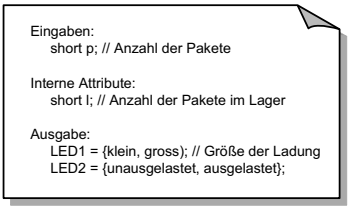
In der Praxis erfreuen sich funktionsorientierte Testmethoden, auch bekannt als Black-Box-Tests, großer Beliebtheit [Lig02]. Die Klassifikationsbaummethode (KBM) [GG06] ist ein solcher Ansatz, und erlaubt die systematische Herleitung von Eingabedaten für Testfälle, die anschließend auf das zu testende System angewendet werden. Bisher wird die KBM hauptsächlich zum Testen von kombinatorischen Systemen eingesetzt, da sie sich nur beschränkt zum Testen von zustandsbehafteten Systemen eignet. Begründet wird die Tatsache dadurch, dass sich bisher die durch einen Zustandswechsel erfolgten Änderungen im Systemverhalten nicht auf die Äquivalenzklassen im Klassifikationsbaum übertragen lassen. So unterstützt die KBM derzeit nur Äquivalenzklassen mit festen Wertebereichen, die sich nicht an den aktuellen Zustand anpassen können. Damit die KBM zukünftig auch auf zustandsbehaftete Systeme effizient anwendbar ist, müssen die internen Attribute und der Zustand eines Systems bei der Bildung von Äquivalenzklassen mit in Betracht gezogen werden. Da ein solches Vorgehen in der Regel sehr aufwändig und fehleranfällig ist, bietet sich eine Automatisierung an.
Wir schlagen daher vor, die zu einem Systemzustand passenden Äquivalenzklassen eines Klassifikationsbaums aus einem endlichen Automaten automatisch abzuleiten. Dadurch lässt sich die Komplexität und Fehleranfälligkeit beim Erzeugen des Klassifikationsbaums erheblich reduzieren. Anschließend können die für die Testfälle benötigten Testdaten aus den für den Zustand gültigen Eingabeäquivalenzklassen systematisch oder intuitiv ausgewählt und auf das zu testende System angewendet werden. In dieser Arbeit wird unser Ansatz näher erläutert und dessen Vorteile vorgestellt.

2 Einführung des Anwendungsbeispiels

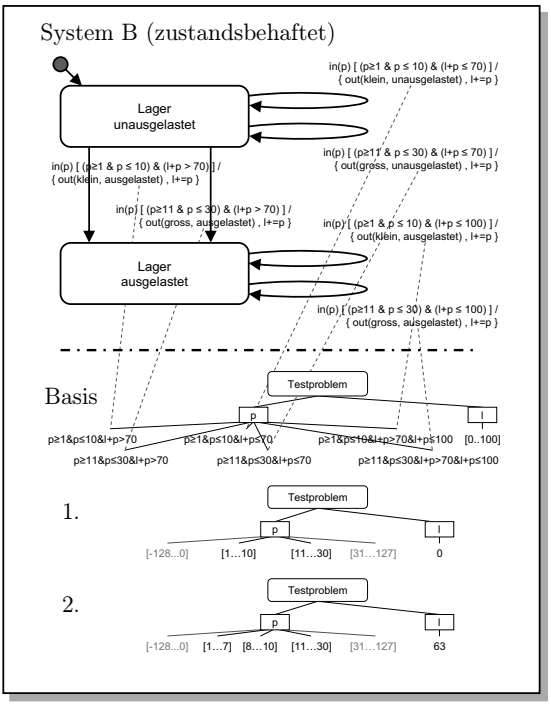
Zum besseren Verständnis des Ansatzes werden in diesem Kapitel zwei Systeme aus dem Transport-Logistik-Bereich vorgestellt (siehe Abbildung 1). Beide Systeme überwachen die Einlagerung von Paketen in ein Lagerhaus. Kleine Ladungen bestehen aus 1 bis 10 Paketen. Große Ladungen hingegen aus 11 bis 30 Paketen. Beide Systeme sollen als Ausgabe anzeigen, ob es sich um eine kleine oder große Ladung handelt. Zusätzlich dazu besitzt System `B` die Aufgabe die Auslastung des Lagers anzuzeigen. Dabei gilt das Lager erst bei mehr als 70 eingelagerten Paketen als ausgelastet. Eine formale Spezifikation der beiden Systeme ist in Abbildung 1 dargestellt. Bei System `A` handelt sich um ein kombinatorisches System, das auch als einfache Funktion interpretiert werden kann. System `B` hingegen ist ein zustandsbehaftetes System, da es sich entsprechend der Anzahl bereits eingelagerter Pakete unterschiedlich verhält. Für diese beiden Systeme soll im Folgenden der jeweils zugehörige Klassifikationsbaum erstellt werden. Im Anschluss könnten dann aus den Klassifikationsbäumen fehlersensitive Testfälle für die zu testenden Systeme erstellt werden.



a) Formale Spezifikation eines kombinatorischen Systems mit dem dazu passenden Klassifikationsbaum.



c) Ergänzende textuelle Spezifikation zu den zwei Systemen A und B.



b) Formale Spezifikation eines zustandsbehafteten Systems mit zwei aus dem Basis-Klassifikationsbaum abgeleiteten und dem internen Attribut entsprechenden Klassifikationsbäumen.

Abbildung 1: Formale Spezifikation zweier Systeme mit zugehörigen Klassifikationsbäumen.

3 Klassifikationsbaummethode

Beim Black-Box-Test darf das zu testende System nur über seine Schnittstellen getestet werden. Demzufolge liegt die einzige Möglichkeit eine Implementierung auf Konformität mit ihrer Spezifikation zu überprüfen in der Untersuchung des vollständigen Eingaberaumes. Jedoch ist das in realen Softwaresystemen aufgrund der damit verbunden Kosten selten praktikabel. Daher wird der Eingabewertebereich typischerweise in Äquivalenzklassen partitioniert. Dabei enthält eine Eingabeäquivalenzklasse nur solche Werte, die gleiches Ausgabeverhalten aufweisen, welches aber mehrere als äquivalent angesehene Ausgabewerte umfassen darf [Lig02]. Die Bildung von Äquivalenzklassen findet auch innerhalb der KBM Anwendung. Dabei wird zunächst die Spezifikation eines kombinatorischen Systems analysiert, um die testrelevanten Eingabeparameter zu identifizieren. Zugleich werden die für das Testproblem unwichtigen Informationen abstrahiert, um dessen Komplexität zu senken. Anschließend wird der Wertebereich aller testrelevanten Eingabeparamter anhand von Aspekten in Äquivalenzklassen unterteilt. Diese Unterteilung geschieht systematisch sowie schrittweise und führt letztendlich zu einer grafisch darstellbaren Baumstruktur, dem Klassifikationsbaum (siehe Abbildung 1a). Zum Erstellen eines Testfalls muss jedem Eingabeparameter aus einer seiner Äquivalenzklassen ein repräsentativer Wert zugewiesen werden. Zum Erstellen von fehlersensitiven Testfällen eignen sich erfahrungsgemäß die Grenzwerte der Äquivalenzklassen. Durch das Bilden von Äquivalenzklassen können redundante Testfälle vermieden werden, was zu einer Reduktion der Anzahl zu betrachtender Testfälle führt. Weitere Vorteile der KBM sind die Visualisierung des Testproblems und das mögliche Aufdecken von Unklarheiten in der funktionalen Spezifikation während der Erstellung des Baumes.
Die KBM setzt auf der Category-Partition-Method [OB88] auf und diente ursprünglich dem Erstellen abstrakter Testfälle. Sie wurde erst in einer späteren Arbeit [LB03] um die Eingaben von konkreten Testdaten erweitert. Die Kombination von Klassifikationsbäumen und zustandsbehafteten Systemen wurde erstmals in [OMS09] angedacht. Zum Themengebiet passende Arbeiten [WS08a, WS08b, Wei08] beschäftigten sich aber schon vorher mit der Bildung von Äquivalenzklassen und deren Einsatz beim Test von zustandsbehafteten Systemen. Der Classification-Tree-Editor (CTE) ist ein eigens für die KBM entwickeltes Werkzeug, das den Tester beim Erstellen eines Klassifikationsbaums unterstützt [GG06, LW00]. Der CTE erlaubt das Erstellen von Testsequenzen durch das Zusammenfassen von Testfällen, wobei ein Testfall jeweils die Voraussetzung für den nachfolgenden Testfall schafft. Allerdings passen sich die Äquivalenzklassen während einer Testsequenz nicht den internen Attributwerten eines zustandsbehafteten Systems an. Folglich kann der Repräsentant einer Äquivalenzklasse bei Vorliegen eines Zustandswechsels im Folgezustand ein anderes Verhalten verursachen, als das Verhalten der durch ihn charakterisierten Äquivalenzklasse. Aufgrund dessen sollten mit dem CTE erstellte Testsequenzen für zustandsbehaftete Systeme als potentiell fehlerhaft angesehen werden. Auch lassen sich mit dem CTE Bedingungen formulieren, die sich auf Äquivalenzklassen unterschiedlicher Eingabeparameter beziehen. So kann angegeben werden, dass eine Äquivalenzklasse nicht mehr ausgewählt werden darf, wenn zuvor einem anderen Parameter ein gewisser Wert zugewiesen wurde. Allerdings lassen sich diese Bedingungen nicht testfallübergreifend formulieren, weshalb sie für zustandsbehaftete Systeme ungeeignet sind.

4 Erweiterung auf zustandsbehaftete Systeme

Bei dem Versuch, für die beiden Systeme `A` und `B` (siehe Abbildung 1) jeweils einen Klassifikationsbaum zu erstellen, wird deutlich, dass die KBM in ihrer derzeitigen Form zwar für kombinatorische Systeme geeignet ist, nicht aber für zustandsbehaftete Systeme. Damit die KBM auch auf zustandsbehaftete Systeme angewendet werden kann, müssen sich die Äquivalenzklassen eines Klassifikationsbaums dynamisch an den Zustand und an die internen Attribute des Systems anpassen. Da die Abhängigkeiten zwischen dem Zustand bzw. den internen Attributen und den Äquivalenzklassen der Eingabewerte sehr komplex ausfallen können und daher für eine manuelle Herleitung zu fehleranfällig sind, sollten solche Äquivalenzklassen automatisch berechnet werden. Mit dynamisch hergeleiteten, dem Systemzustand entsprechenden Äquivalenzklassen sind Testsequenzen mit konkreten und gültigen Grenzwerten als Eingabedaten für zustandsbehaftete Systeme mittels der KBM möglich. Für das automatische Herleiten der Äquivalenzklassen wird jedoch eine formale Spezifikation des zu testenden Systems vorausgesetzt, beispielsweise ein Zustandsautomat (siehe Abbildung 1b).
Durch die Hinzunahme eines Zustandsautomaten gehört die KBM zukünftig zu den zustandsbasierten Testmethoden, die wiederum den modellbasierten Testverfahren zugeordnet werden. Zustandsbasierte Testmethoden werden eingesetzt, wenn sich neben den Eingabewerten auch der aktuelle Zustand des zu testenden Systems auf das Systemverhalten auswirkt. Bisher wurden zustandsbasierte Testmethoden vor allem zur Strukturüberdeckung eingesetzt. Durch die Kombination mit der Äquivalenzklassenbildung bietet sich jedoch das Testen von Grenzwerten an. Ein weiterer Vorteil dieser Kombination ist zudem, dass im hinzugenommenen Zustandsautomaten das erwartete Testverhalten enthalten ist, wodurch dieser auch als Testorakel eingesetzt werden kann. Bisher wird diese Funktion vom Tester übernommen. Schließlich eignet sich die Kombination, um die üblicherweise nicht spezifizierten, ungültigen Äquivalenzklassen automatisch zu identifizieren (siehe die äußeren Äquivalenzklassen des Klassifikationsbaums in Abbildung 1a) und damit, durch die Wahl eines entsprechenden Repräsentanten, das System auf unnötig implementierte Funktionen zu überprüfen.
Aus dem Zustandsautomaten kann für jeden Zustandsübergang die Äquivalenzklasse derjenigen Werte berechnet werden, die für den Zustandsübergang benötigt werden (siehe Basis-Klassifikationsbaum in Abbildung 1b). Für diese Aufgabe können insbesondere Verfahren der symbolischen Ausführung Anwendung finden [Kne92]. Diese sind in der Lage, selbst noch bei Zustandsübergängen mit komplexen Verzweigungsbedingungen (Kontrollflussartige Strukturen) die korrekten Äquivalenzklassen zu bestimmen. In einem zustandsbehafteten System wirken sich der Zustand und die internen Attribute in der Regel auf die Äquivalenzklassen der Eingabeparameter aus. Beispielsweise ist in Tabelle 1 gut zu erkennen, welche Auswirkung der Wert des internen Attributs `l` des Systems `B` auf die Äquivalenzklassen des Eingabeparameters `p` hat. In Abbildung 1b sind zum besseren Verständnis noch zwei Klassifikationsbäume dargestellt, deren Äquivalenzklassen jeweils an den Wert 0 bzw. 63 des internen Attributs `l` dynamisch angepasst wurden.
Da interne Attribute nicht direkt von außen beeinflussbar sind, kann deren Auswirkung auf die Äquivalenzklassen der Eingabeparameter leicht aufgelöst werden. So lässt sich die Äquivalenzklasse eines abhängigen Eingabeparameters sofort mit dem Wert des zugehöri-

ÄK des Eingabeparameters *p* in Abhängigkeit vom Wert des Attributs *l*	Zustand	ÄK für Attribut *l*
[1 … 10] , [11 … 30]	unausgelastet	[0 … 40]
[1 … 10] , [11 … (70 – *l*)] , [(71 – *l*) … 30]	unausgelastet	[41 … 59]
[1 … 10] , [11 … 30]	unausgelastet	60
[1 … (70 – *l*)] , [(71 – *l*) … 10] , [11 … 30]	unausgelastet	[61 … 69]
[1 … 10] , [11 … 30]	unausgelastet	70
[1 … 10] , [11 … (100 – *l*)]	ausgelastet	[71 … 89]
[1 … (100 – *l*)]	ausgelastet	[90 … 99]

Tabelle 1: Die Äquivalenzklassen des Eingabeparameters `p` des zustandsbehafteten Systems `B` (siehe Abbildung 1b) verändern sich in Abhängigkeit vom Wert des internen Attributs `l`.

gen internen Attributs verrechnen. Je mehr Abhängigkeiten durch das Festlegen von Werten aufgelöst wurden, desto übersichtlicher wird die Darstellung des dynamischen Klassifikationsbaums. Durch das Auflösen der Abhängigkeiten können auch mehrere Äquivalenzklassen eines Eingabeparameters zusammenfallen. In Abhängigkeit davon, für welche Äquivalenzklasse zuerst Repräsentanten gewählt und somit die Abhängigkeiten aufgelöst werden, kann der Klassifikationsbaum vorübergehend unterschiedliche Äquivalenzklassen ausbilden. Zusätzlich zu der Abhängigkeit zwischen den Eingabeparametern und den internen Attributen, können die Eingabeparameter aber auch untereinander voneinander abhängen. Diese Abhängigkeit lässt sich durch Vorbedingungen lösen. Diese werden den dynamischen Äquivalenzklassen eines abhängigen Eingabeparameter beigefügt und ermöglichen bei Auswahl eines Repräsentanten die Auflösung der Abhängigkeit.
Testsequenzen können mit dieser Erweiterung für zustandsbehaftete Systeme sowohl manuell als auch automatisch erstellt werden. Das manuelle Erstellen einer Testsequenz verläuft wie folgt:

Ausgehend vom Startzustand werden die Äquivalenzklassen berechnet, die die Bedingungen der ausgehenden Zustandsübergänge erfüllen. Anschließend werden, abhängig vom gewünschten Folgezustand, die Repräsentanten ausgewählt. Ausgehend vom erreichten Folgezustand wird der zuvor getätigte Schritt solange wiederholt, bis das Testziel der Testsequenz erreicht ist.

Ein Testziel kann beispielsweise das Erreichen eines Zustandes darstellen. Testsequenzen zu einem bestimmten Testziel lassen sich aber auch automatisch erstellen. Durch den Einsatz von Model-Checkern [GH99] lassen sich beispielsweise die kürzesten Testsequenzen zu einem Testziel automatisch berechnen. Unabhängig davon, auf welche Art die Testsequenzen erstellt wurden, können diese anschließend durch die KBM so modifiziert werden, dass die Testsequenzen gegenüber Fehlern sensitiver werden. Zu diesem Zweck müssen die Eingabewerte der sequentiell ablaufenden Testfälle durch die Grenzwerte der zugehörigen Äquivalenzklasse ausgetauscht werden. Allerdings ist dabei darauf zu achten, dass das eigentliche Verhalten der Testsequenz nicht verändern wird.

5 Zusammenfassung und Ausblick

In dieser Arbeit haben wir einen Ansatz beschrieben, wie sich die Klassifikationsbaummethode auf zustandsbehaftete Systeme effizient anwenden lässt. Dafür wird die Methode um eine Zustandsautomaten erweitert, der gerade der formalen Spezifikation des zu testenden Systems entspricht. Aus diesem Verhaltensmodell heraus kann der Eingabewertebereich in die für einen Systemzustand gültigen Äquivalenzklassen zerlegt werden. Durch die Auswahl und Festlegung von Grenzwerten als Repräsentanten der Äquivalenzklassen können dann insbesondere fehlersensitive und weniger redundante Testfälle für das zustandsbehaftete System abgeleitet werden. Aufgrund der mit der Berechnung der Äquivalenzklassen verbundenen hohen Komplexität und derer dynamischen Darstellung soll in weiteren Arbeiten ein Werkzeug entwickelt werden, das diese Aufgaben übernimmt.

Literatur

[GG06] M. Grochtmann und K. Grimm. Classification Trees for Partition Testing. *Software Testing, Verification and Reliability*, 3(2):63–82, 2006.

[GH99] A. Gargantini und C. Heitmeyer. Using Model Checking to Generate Tests from Requirements Specifications. *SIGSOFT Software Eng. Notes*, 24(6):146–162, 1999.

[Kne92] R. Kneuper. Validation und Verifikation von Software durch symbolische Ausführung. *Testen, Analysieren und Verifizieren von Software, Informatik Aktuell. Springer*, 1992.

[LB03] S. Lützkendorf und K. Bothe. Attributierte Klassifikationsbäume zur Testdatenbestimmung. *Fachgruppe TAV der Gesellschaft für Informatik*, 2003.

[Lig02] P. Liggesmeyer. *Software-Qualität*. Spektrum, Akad. Verlag, 2002.

[LW00] E. Lehmann und J. Wegener. Test Case Design by means of the CTE XL. *Proceedings of the 8th European International Conference on Software Testing, Analysis & Review (EuroSTAR 2000)*, 2000.

[OB88] T. J. Ostrand und M. J. Balcer. The Category-Partition Method for Specifying and Generating Functional Tests. *Commun. of the ACM*, 31(6):676–686, 1988.

[OMS09] S. Oster, F. Markert und A. Schürr. Integrated Modeling of Software Product Lines with Feature Models and Classification Trees. In *Proceedings of the 13th Int. SPL Conf.*, MAPLE 2009 Workshop. Springer, 2009.

[Wei08] S. Weißleder. Partition-Oriented Test Generation. *GI Jahrestagung*, 1:199–204, 2008.

[WS08a] S. Weißleder und H. Schlingloff. Deriving Input Partitions from UML Models for Automatic Test Generation. In *Models in Software Eng.: Workshops MODELS 2007*, Seiten 151–163. Springer, 2008.

[WS08b] S. Weißleder und H. Schlingloff. Quality of Automatically Generated Test Cases based on OCL Expressions. *ICST*, 2008.

Automatisierter funktionaler Steuergerätetest mit der EXtended Automation Method (EXAM)

Dirk Zitterell, Sebastian Thiel

AUDI AG
I/EE-65
Hardware-in-the-Loop Funktionserprobung
85045 Ingolstadt
dirk.zitterell@audi.de, sebastian1.thiel@audi.de

Abstract: Bei der Absicherung von Fahrzeugfunktionen spielen System- und Verhaltensmodelle der zu testenden Systeme sowie formale Methoden eine große Rolle. Jedoch ist ein Einsatz dieser Methoden auf Gesamtfahrzeugebene auf Grund der Komplexität des gesamten Systems nicht praktikabel. Hier werden Hardware-in-the-Loop-Prüfstände eingesetzt, welche System- und Verhaltensmodelle sowie reale Hardware verbinden. Im Volkswagen-Konzern wird die Testautomatisierungsmethode EXAM für die Prüfungsdurchführung und –auswertung eingesetzt. Neben der Vorstellung der Methode auf Basis eines modellgetriebenen Testentwicklungsansatzes wird auf die praktischen Erfahrungen mit dieser agilen Verfahrensweise eingegangen.

1 Einleitung

Getrieben durch höhere Anforderungen an Sicherheit, Komfort, Umweltverträglichkeit, Fahrspaß und Innovationen ist die Fahrzeugelektronik in den letzten Jahren immer komplexer geworden. Die hohe Komplexität ergibt sich durch eine stetig steigende Zahl an Fahrzeugfunktionen und eine zunehmende Vernetzung der Funktionen und die Verteilung der Funktionen auf mehrere Steuergeräte. Gleichzeitig stellen verkürzte Entwicklungszyklen, eine Derivatisierung der Fahrzeugprojekte, ein gestiegener Kostendruck und neue Normen (z.B. ISO 26262[1]) eine große Herausforderung bei der Entwicklung, Integration und insbesondere bei der Absicherung neuer Funktionen dar.

Die Aufgabe von Zulieferern und Automobilherstellern besteht darin, trotz erschwerter Rahmenbedingungen die fehlerfreie Funktion ihrer Produkte zu gewährleisten. Für eine effektive und effiziente Durchführung dieser Aufgabe sind modellbasierte Verfahren und der Einsatz effektiver Testautomatisierungssysteme hilfreich.

[1] Die ISO 26262 („*Road vehicles – Functional safety*“) ist eine entstehende ISO-Norm für sicherheitsrelevante elektrische/elektronische Systeme in Kraftfahrzeugen.

Zur Funktionsabsicherung der Steuergerätesoftware in den frühen Entwicklungsphasen werden MiL- (Model-in-the-Loop) und SiL- (Software-in-the-Loop) Testverfahren eingesetzt. Hierbei spielen Verhaltensmodelle der Systeme und formale Methoden eine große Rolle. Nach der Integration der Software ins Steuergerät liegt der Testfokus auf der Absicherung der gesamten Komponente (Steuergerät inkl. Software) bzw. der Funktion (Steuergeräteverbund). Der Einsatz von Hardware-in-the-Loop-Prüfständen ermöglicht eine Kombination von realen Komponenten und simulierten Komponenten. Die zu überprüfenden elektronischen Komponenten werden dabei untereinander mittels Bussystemen vernetzt und als Echtteile geprüft. Alle weiteren Fahrzeugkomponenten, das Fahrzeugverhalten und die Fahrzeugumgebung werden mithilfe von Verhaltensmodellen durch den HiL-Simulator in Echtzeit simuliert. Für eine effiziente Testdurchführung und hohe Auslastung des Prüfstands sind Testautomatisierungssysteme erforderlich. Einen detaillierten Überblick über den Absicherungsprozess in der Automobilindustrie geben u.a. [Sax08] und [SZ10]. Im Volkswagen-Konzern wird für die Testfallerstellung ein modellgetriebener Testentwicklungsansatz verwendet. Im Folgenden wird die hierfür eingesetzte Testautomatisierungslösung EXAM vorgestellt und auf praktische Erfahrungen mit dieser Methode eingegangen.

2 Testautomatisierungsmethode EXAM

Die EXAM-Methode basiert auf einem modellgetriebenen Testentwicklungsansatz (engl.: Model Driven Test Development, MDTD) und beschreibt die Modellierung, die Ausführung und die Auswertung von funktionalen Testfällen. Einen detaillierten Einblick über den wohldefinierten Testprozess mit EXAM im Volkswagen-Konzern gibt [TZ08]. Wir werden in diesem Kapitel neben dem EXAM-Testprozess vor allem die EXAM-Modellierung sowie die EXAM-Toolsuite vorstellen.

2.1 EXAM-Testentwicklungsprozess

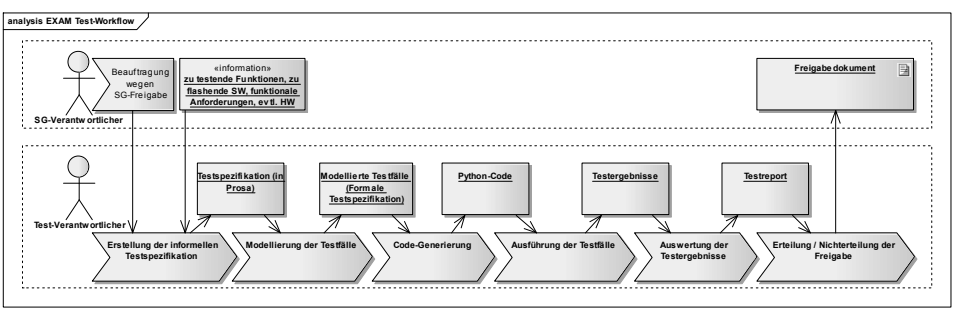


Abbildung 1: Prozessbild für die Entwicklung und Ausführung von funktionalen Testfällen mit EXAM

2.2 EXAM-Metamodell

Die Entwicklung von EXAM-Testfällen erfolgt basierend auf einem EXAM-Modell. Ein EXAM-Modell besteht im Allgemeinen aus den eigentlichen Testfällen, welche verschiedenen Testthemen wie z.B. Start-Stopp zugeordnet sind, sowie dem abstrakten Schnittstellenmodell *StandardAbstractCar*, welches das Testobjekt – d.h. das Fahrzeug – repräsentiert und zur Stimulation desselben bzw. zur Interaktion benutzt wird. Des Weiteren dient eine im EXAM-Modell enthaltene Funktionsbibliothek *(FunctionLibrary)* mit einem Satz von Funktionsbausteinen der Anbindung und Ansteuerung von Geräten (Diagnosetester, Datenlogger, Roboter etc.), zur Aufzeichnung von Datenlogs und Signalen sowie zur automatisierten Auswertung von Testergebnissen.

EXAM-Modelle werden in einer Teilmenge der UML2 [Uml05] formuliert, wobei neben domänenspezifischen Erweiterungen ein UML-Profil zum Einsatz kommt, welches die Anpassung an die Domäne des Testens von Fahrzeugelektronik vornimmt.

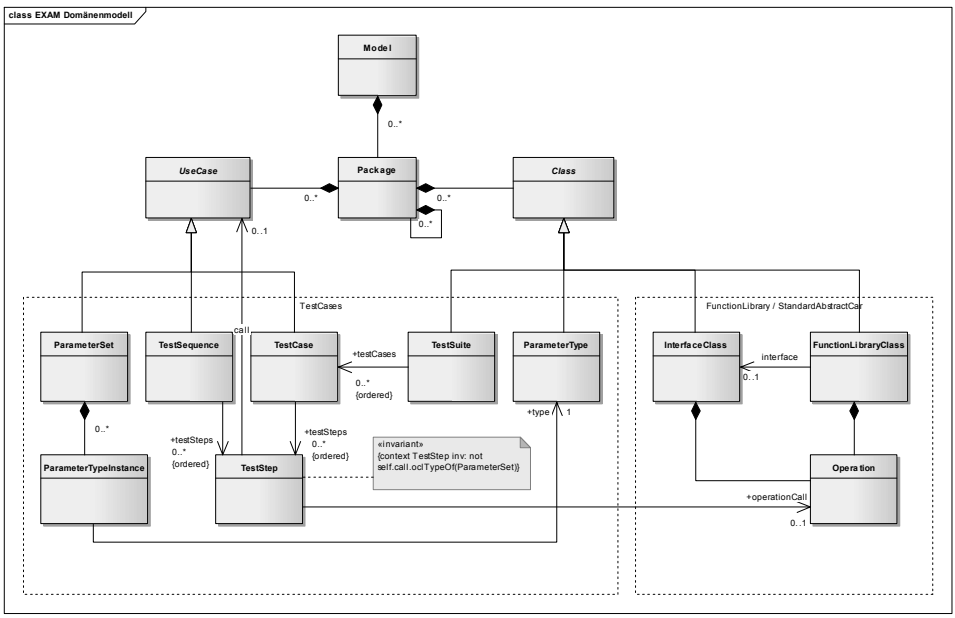


Abbildung 2: Vereinfachtes EXAM-Domänenmodell

Abbildung 2 zeigt in stark vereinfachter Form einen Ausschnitt des EXAM zugrunde liegenden Domänenmodells. Das Modell ist mit Hilfe von *Packages* als Hierarchie strukturiert. Die Funktionsbibliothek besteht hauptsächlich aus *InterfaceClasses* und implementierenden *FunctionLibraryClasses*. Testfälle werden mit Hilfe der Operationen von *InterfaceClasses* modelliert. Unterschiedliche Implementierungen erlauben hierbei die Plattformunabhängigkeit vom Prüfstand. So kann zum Einschalten der Zündung an einem Prüfstand bspw. eine Implementierung für ein Relais vorliegen, während an einem anderen Prüfstand etwa ein Roboter zum Einsatz kommt. Das EXAM-UML-Profil er-

laubt es, EXAM-Testfälle in UML zu modellieren. Beispielsweise wird ein EXAM-Testfall durch einen *UmlUseCase* mit Stereotyp *«testCase»* modelliert, eine EXAM-Testsuite durch eine *UmlClass* mit Stereotyp *«testSuite»* und ein EXAM-Parametersatz[2] durch einen *UmlUseCase* mit Stereotyp *«parameterSet»*. Das EXAM-Metamodell orientiert sich hierbei im Gegensatz zum *UML Testing Profile* [Utp05] an der Domäne der Hardware-in-the-Loop-Simulation.

2.3 EXAM-Toolsuite

Für eine effiziente Umsetzung der EXAM-Methode in der Praxis ist eine Unterstützung der Anwender durch Softwarewerkzeuge erforderlich. Aus diesem Grund wurden auf Basis der Eclipse-Rich-Client-Platform vier Softwaretools entwickelt: *EXAM modeller*, *EXAM generator*, *EXAM testrunner* und *EXAM reportmanager*. Die Tools decken den gesamten Testprozess von der Erstellung des formalen Testmodells bis zur automatisierten Testbewertung und der Erstellung der Testreporte ab.

Der *EXAM modeller* ermöglicht die grafische Modellierung der Testfälle auf Basis des EXAM-UML-Profil und die Verwaltung der EXAM Modelle. Für die Durchführung der Testfälle am Prüfstand müssen die grafisch modellierten Testfälle in ablauffähige Prüfskripte übersetzt werden. Diese Aufgabe übernimmt der *EXAM generator*. Im Volkswagen-Konzern wird ein zentraler *EXAM generator* eingesetzt der die Modelle inkrementell übersetzt. Der *EXAM testrunner* erlaubt die Konfiguration und Durchführung der Testabläufe und speichert die Testergebnisse in einer Reportdatenbank. Diese Daten können dann durch den *EXAM reportmanager* visualisiert, verwaltet und zu Testreporten zusammengefasst werden. Die EXAM-Tools sind Freeware und können gemeinsam mit einer Basis-Funktionsbibliothek von der EXAM-Webseite[3] heruntergeladen werden.

3 Erfahrungen aus der Praxis

EXAM hat sich in den vergangenen Jahren als Automatisierungsmethode in den einzelnen Marken des Volkswagen-Konzerns etabliert. In diesem Kapitel werden wir nun auf die mit ihr in der Praxis gemachten Erfahrungen eingehen.

3.1 Agile Vorgehensweise

Agile Werte und Prinzipien bilden nach [EJ04] das Fundament agiler Softwareentwicklung. Beispiele hierfür sind die effiziente Kommunikation zwischen den an der Entwicklung Beteiligten, die Kundennähe, die Offenheit gegenüber Änderungen, die Wieder-

[2] Parametersätze dienen der Parametrierung von Testfällen. Die Bedatung erfolgt in einem enthaltenen *UmlObjectDiagram*.
[3] EXAM-Webseite: http://www.exam-ta.de/

verwendung von Ressourcen, die Zweckmäßigkeit sowie das KISS-Prinzip[4]. Die meisten dieser Prinzipien werden von der EXAM-Methode adressiert und wirken sich positiv auf den Testprozess aus.

So erlaubt die Terminologie, die das EXAM-Metamodell definiert, eine effiziente und einheitliche Kommunikation mit allen am Testprozess beteiligten Stakeholdern und Rollen – sogar über Unternehmensgrenzen hinweg. Statt Quellcode bilden Modelle und Diagramme die Basis der Kommunikation zwischen Tester (Auftragnehmer) und Steuergeräteentwickler (Auftraggeber), welche zusammen ein Funktionstestteam bilden. Da zudem entwicklungsbegleitend getestet wird und die funktionalen Anforderungen an die Testobjekte stets Änderungen unterliegen, erlaubt EXAM ein ständiges *Model Refactoring*.

EXAM-Testmodelle werden in leistungsfähigen Datenbanken bei der Konzern-IT gehostet. Testentwickler und Tester, die verschiedenen Konzernmarken angehören, können auf diese Weise gleichzeitig auf einem gemeinsamen Modell (*Shared Model*) arbeiten. Neue Funktionalität, die von einem Entwicklungspartner in die Funktionsbibliothek eingebracht wird, steht unmittelbar allen auf dem Modell arbeitenden Entwicklern und Testern zur Verfügung, was nicht nur die schnelle Reaktion auf Änderungen ermöglicht, sondern auch die Wiederverwendung fördert und redundante Aktivitäten minimiert. Des Weiteren unterstützt EXAM markenübergreifende, auf Modulen basierende Steuergeräteentwicklungsprojekte (modulare Baukästen) durch kollaborative Testfallentwicklung.

3.2 Automatisierung

EXAM bildet mit seiner Architektur und seinen Werkzeugen das Fundament für eine durchgängige Prozessunterstützung und dadurch einen hohen Automatisierungsgrad an den einzelnen Schritten im Testprozess.

Zunächst wird vom zentralen Code-Generator inkrementell ausführbarer Code erzeugt, der auf die Prüfstandsrechner aufgespielt wird. Vor der Ausführung der Testfälle am Prüfstand wird diese Synchronisation ausgeschaltet, um die Aktualisierung der Prüfcodes während der Ausführung zu unterbinden. Des Weiteren werden für die automatisierte Ausführung von Testfällen am Prüfstand im Modell mit Hilfe von *«testCampaigns»* Testläufe für die Nacht oder das Wochenende modelliert, die aus einer Vielzahl von Testfällen bestehen und deren Testergebnisse dann am folgenden Werktag zur Verfügung stehen.

Während die Funktionsbibliothek bereits eine große Zahl an automatisierbaren Prüfungen ermöglicht, werden durch das Anbinden externer Funktionalität durch Services wie bspw. für die automatisierte Auswertung von Kamerabildern oder die automatisierte Auswertung von *Signal Traces* der Automatisierungsgrad immer weiter erhöht.

---

[4] Design-Prinzip in der Softwaretechnik: „Keep it simple and stupid.“ („Halte es einfach und leicht verständlich.“)

4. Zusammenfassung und Ausblick

Die Veränderungen im Entwicklungsprozess stellen gleichermaßen neue Anforderungen an die Funktionsabsicherung. Viele komplexe Funktionen von unterschiedlichen Fahrzeugprojekten müssen gleichzeitig und entwicklungsbegleitend abgesichert werden. Die Absicherung mit Hilfe von Verhaltensmodellen der Funktionen und formalen Methoden ist auf Funktionsebene und gerade auf Ebene des Gesamtfahrzeugs aufgrund der Komplexität nicht wirtschaftlich.

Im Volkswagen-Konzern werden mit Hilfe von Hardware-in-the-Loop-Prüfständen Verhaltens- und Systemsmodelle mit realer Hardware kombiniert. Für den effizienten Betrieb der Prüfstände wird seit etwa fünf Jahren die auf dem MDTD-Ansatz basierende Testautomatisierungsmethode EXAM erfolgreich eingesetzt. Neben der Vorstellung der Prozessschritte, der EXAM-Tools und der Modellierung mit Hilfe eines UML-Profils in einem zentralen *Repository* haben wir dargestellt wie EXAM eine agile Entwicklung der Testfälle ermöglicht. Gerade durch die Zusammenarbeit zwischen Testabteilung und Entwicklungsabteilung in Funktionstestteams und die Wiederverwendung von Testfragmenten in EXAM wird der Testprozess positiv beeinflusst.

Obwohl durch die Einführung der EXAM-Methode der Testprozess entscheidend verbessert wurde, stellen die aktuellen Rahmenbedingungen (z.B. kürzere Entwicklungszyklen, gestiegene Betriebskosten bei hohem Kostendruck etc.) weitere Anforderungen an die Methode. Schon heute enthalten die Testmodelle mehrere Millionen Elemente und es stehen mehr Testfälle zur Verfügung als operative Prüfstandszeit. Um die kostbare Ressource Prüfstandszeit effizienter zu nutzen, sind daher zusätzlich Verfahren notwendig, die bspw. semantische Modellierungsfehler vor der Ausführung am Prüfstand erkennen (*Model Checking*). Zusätzlich sind bei der modellgetriebenen Testentwicklung für die Zusammenarbeit in weitverteilten Teams effiziente Tools und Verfahren zum *Diffen* und *Mergen* sehr großer, persistierter Modelle gefragt, welche auf der Basis von Quellcode in der klassischen Softwareentwicklung bereits selbstverständlich sind.

Literaturverzeichnis

[EJ04] Jutta Eckstein, Nicolai Josuttis: Agile Softwareentwicklung im Großen: Ein Eintauchen in die Untiefen erfolgreicher Projekte, dpunkt.verlag, 2004, ISBN 978-3-89864-250-7.

[Sax08] Sax, E.: Automatisiertes Testen Eingebetteter Systeme in der Automobilindustrie, Hanser-Verlag, 2008.

[SZ10] Schäuffele, J.; Zurawka T.: Automotive Software Engineering: Grundlagen, Prozesse, Methoden und Werkzeuge effizient einsetzen, Vieweg+Teubner, 4.Auflage, 2010

[TZ08] Sebastian Thiel, Dirk Zitterell: EXtended Automation Method (EXAM) zur automatisierten Funktionserprobung von Steuergeräten in der Automobilindustrie. GI Jahrestagung (2) 2008: 625-630.

[Uml05] Object Management Group: Unified Modeling Language, 2005, http://www.uml.org/

[Utp05] Object Management Group: UML Testing Profile, 2005, http://utp.omg.org/

Enabling statistical testing for component-based systems

Thomas Bauer, Robert Eschbach

Fraunhofer-Institut für Experimentelles Software Engineering IESE
Fraunhofer Platz 1
67663 Kaiserslautern, Germany
{thomas.bauer | robert.eschbach}@iese.fraunhofer.de

Abstract: In this article, an automated statistical test approach for the systematic quality assurance of component-based systems is presented. From state-based component models a compact interaction test model is constructed, which describes component interactions and interaction sequences. The model is annotated with an interaction profile to facilitate the generation of representative test cases and the reliability estimation of the test object during the test evaluation. For the test case generation and reliability estimation model-based statistical testing was applied. A small example of a component-based system from the automotive domain is given.

1 Introduction

In component-based systems, faulty component interactions have a high impact on the dependability of the system. With the component-based development of software systems and the use of third-party components, the integration and interoperability testing of a distributed system has become an important quality assurance activity. Integration testing is performed in parallel to the integration stage where the system is composed from its components. The goal is to test system interactions and functionalities that are spread across several system components. Our work is restricted to interactions that are provoked by the event flow between components and the synchronization of components on certain events. Furthermore, we assume that state-based component models are available.

We propose an integration test approach that fully covers the component interactions and sequences of interactions by constructing a compact *interaction test model.* The model is derived from state-based component models that represent the functional component requirements. The interaction test model is enhanced by an *interaction profile* to enable statistical testing, generate representative sets of test cases, and estimate the product reliability based on the test results. The reliability estimates can be used as test stopping criterion, to assess the current status of the product quality, and to predict future problems after the product release.

The contribution of this article are the construction approach of interaction test models from state-based component models and the application of statistical testing to test component-based systems and assess the reliability w.r.t. component interactions. This paper is structured as follows: In chapter 2, related work in the field of model-based integration testing is presented. Our general approach is described in chapter 3. Chapter 4 contains a detailed description of the construction procedure of the interaction test models. The application of statistical testing and the usage profiles are covered by chapter 5. The article concludes with a summary and outlook in chapter 6

2 Related Work

The research in integration testing comprises a wide field of systematic test solutions for component-based systems. Interaction profiles and reliability estimations have not been covered. Most of the automated test approaches use different types of models, e.g. requirements models, component behavior models, system structure models, or code models. Several approaches have been developed which generate test cases from state-based component models [ABJ07, BRT03, DV07, HIM00, HP05, LR01]. These approaches focus on the efficient coverage of single component interactions. Longer sequences of component interactions are not considered. Some publications address the complexity in state machine composition during the integration testing. The approach in [RHP08] creates partial models from component state machines. Model checking is used to derive test cases based on several communication coverage criteria. The approach also works for extended finite state machine (EFSMs) with variables and conditions. [Pre03] proposes a solution for the generation of integration test cases from component test cases. He uses EFSMs for describing the component functionalities. The resulting integration test suite provides MC/DC (Modified Condition/Decision Coverage) of the underlying component models. [KCT02] aimed at reducing the number of states and transitions of composed state machines by incrementally creating reduced observationally equivalent state machines. The models were derived from code models (control and data flow models).

Some integration test approaches combine high level system models with low level component models. [SWK09] uses a so-called global choreography model for system modeling and local partner models for component modeling. Abstract test cases are generated from the global model and later refined into traces of the component models by applying model checking. [Be07] uses feature interaction models for the high level description and low level component models. Test cases are generated based on several feature interaction coverage criteria. Both solutions avoid the construction of the complete product model, but they do not support the systematic generation of longer interaction sequences. The generation of representative test cases by incorporating a usage profile is also not considered. .

3 Approach Overview

The main idea of our approach is the development of an interaction test model from component models. The interaction test model is a simplified system model which incorporates an interaction profile to describe importance and criticality of component interactions. Component test models (CTM) describe the possible input-output trajectories of components and theircontrol logic represented by a finite state machine. These models might be re-used from component testing or component design.

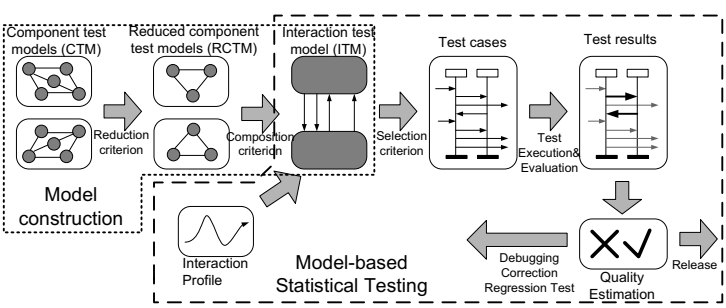


Figure 1: Steps of the test approach

Figure 1 shows the steps of our approach which is divided into two phases, model construction and model-based statistical testing. At first, CTMs are analyzed. All elements of the CTMs that are not needed for executing interactions and interaction sequences are removed. The remaining transitions are merged to simplify the models and to ease the model composition later. The resulting reduced component test model (RCTM) contains the condensed information that is required to generate valid sequences of component interactions from the component's perspective. The RCTM transitions are composite. Each of them contains a synchronized event, either as a stimulus or as a response. In the next step, the RCTMs are composed into the interaction test model (ITM) by applying product composition. Since all RCTM transitions contain synchronized events, the transitions are either merged or removed for the ITM. The ITM contains all reachable component interactions based on the underlying component test models. Due to the high number of component synchronizations, the resulting ITM becomes small compared to the complete product model which would be generated from the CTMs.

Additionally, the ITM is annotated with an interaction profile which describes the frequency and importance of different interactions. This allows the application of model-based statistical testing (MBST). The test case generation, execution, and evaluation is fully automated by the approach. Chapter 5 describes the application and adaptation of MBST in detail.

4 Model Construction

In order to systematically test component interactions and sequences of them, the possible connections of different component interactions are interesting. Looking at the CTMs we have identified three relevant aspects: the ways from the start state to the interacting transitions (*initial interaction paths*), the ways between two interacting transitions (*connecting interaction paths*), and the ways from an interacting transition to the exit state (*final interaction paths*). Each aspect can be described by a set of particular paths. These paths do not contain any synchronized transitions except for the last transition. This assures that other interactions are not triggered by accident during the test execution.

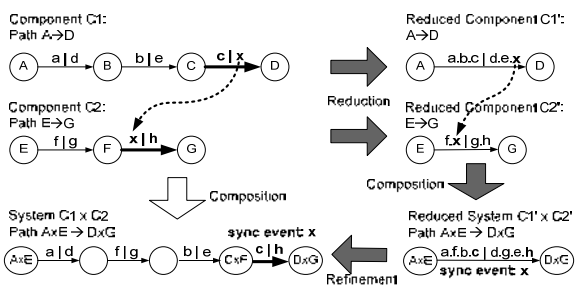


Figure 2: Model construction

In the first step, we extract the component-specific information that is needed to construct these three artifacts. Therefore, we have to determine all synchronized transitions of the CTMs, i.e., all transitions with synchronized events. Then, we determine the initial and final paths for each synchronized transition and the connecting paths for each pair of synchronized transitions. Based on that, the RCTMs can be generated containing the interaction-relevant information of the component. The RCTMs are composed into a so-called interaction test models (ITM). Figure 2 shows the mapping of CTM paths to transitions of the RCTM and their composition into transitions of the ITM. Assuming that A and E are the start states of the CTMs *C1* and *C2*, the paths in the figure represent the initial paths of *C1* and *C2*. Both models synchronize on

event *x* when *C1* is in state *C* and *C2* in state *F*. In order to simplify the model structure, we combine the transition paths of the CTMs into single composite transitions of the RCTMs. Each transition of a RCTM represents either an initial path α, a connecting path β, or a final path γ. The composite transitions of the α- and β-paths contain exactly one component interaction at the end, in our example the flow of event *x*. The resulting RCTM will only contain the source and target states of the composite transitions. The other states can be easily determined by deriving the sequence of atomic transitions from the composite transition.

When composing two RCTMs, the last atomic transitions of the composite transitions may synchronize, here: $c|x$ of *C1* and $x|h$ of *C2*. The result in the product is an atomic transition $c|h$ with the internal event *x*. The non-synchronized atomic transitions of the underlying RCTMs are combined in series with preserved order. Here, we chose an interleaving approach for the *C1'* and *C2'* transitions, resulting in the input sequence *a.f.b* with the output *d.g.e*. The transition of the composed ITM *C1'xC2'* is also composite. It can be refined into a sequence of atomic transitions, which corresponds to a valid path in the unreduced product model *C1xC2*.

5 Model-based Statistical Testing

Model-based statistical testing (MBST) is a black-box testing technique that enables the generation of representative test cases from the tester's or user's perspective [WPT95]. The internal model is a discrete time Markov chain which describes the stimulation of the test object by its environment. The operational profile (in MBST usage profile) of the test object, which is annotated in the test model, represents frequency, criticality, or cost of the test execution [Mu93]. Based on the test results and the underlying operational profile the system reliability is estimated. For our ITM we wanted to define an appropriate *interaction profile* to support statistical testing of component interactions. The resulting interaction profile describes the probabilities of executing component interactions at different states of the ITM.

The example of the feasibility study, an in-car info- and entertainment system consisting of a phone and a CD-player, is shown is Figure 3. The system has to fulfill comfort functions which are provided by the interplay of both components. The interesting parts of the example deal with interrupting and resuming the CD-player during a phone call. For the first version we made the following simple assumptions: Incoming calls (60%) are more frequent than outgoing calls (40%). The CD-player is more often in the playing state (80%) than in the state when it is paused by a radio traffic message (20%). An incoming or outgoing phone call is rejected in 20% and accepted in 80% of the cases. The probabilities of the transitions to the exit state have a high impact on the average length of randomly generated test cases. The probability of the exit transition (γ-path, *stop+deact*) was set to 10% which leads to an average length of randomly test cases of about 21 events in the statistical model analysis. If longer or shorter test cases are required the probabilities of the exit transitions have to be adapted. More information on statistical model analysis and simulation is provided in [WPT95].

Test cases are automatically generated from the test models as arbitrary paths from the start to the exit state. Different strategies for automated test case generation exist, including model coverage for covering all states, transitions, or transition pairs and random tests which randomly traverse the model based on the usage profile. The number of test cases to be generated is linked to the planned test effort. Assuming an average test case length of 21 events (taken from the statistical model analysis), 10 sec for setting up a test case, 5 sec for executing and evaluating one composite transition of the ITM, the average time needed for preparing, executing, and evaluating a random test case is about 120 sec. If we consider a 2 hours slot for the execution of the test case set, about 60 test cases can be executed. In order

to assure the complete coverage of all model states and transitions, we generate the transition coverage set first, followed by 60 random test cases.

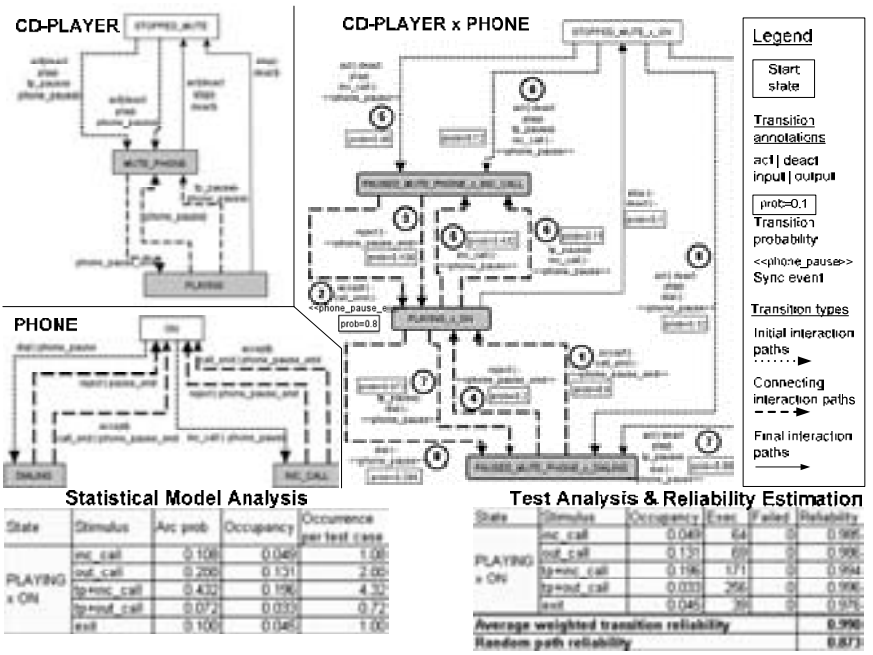


Figure 3: Example for component models and interaction test models

After the execution and evaluation of the test cases, the reliability of test object is estimated based on the test results [SP00, PP05]. The reliability estimation represents the probability that a future system usage will not fail. It represents the confidence of the user in the test object. During the test planning the test manager may set a desired reliability value as test stopping criterion, e.g. 0.99. Based on the interaction profile some composite transitions of the ITM are executed very often in the test, e.g. *[PAUSE_MUTE_PHONExINC_CALL]."accept+call_end"* with 289 occurrences in all test cases. The reliability estimates for this transition is very high assuming zero failures (0.997). Other transitions were generated and executed only few times, e.g. *[STOPPED_MUTExON]."act+play+tp_pause+dial"* with only 7 occurrences which leads to a reliability estimate of 0.889. The statistics from the reliability estimations are used to support the test management decisions, e.g. the effort distribution of regression tests. The overall system reliability is represented by a weighted average of the transition reliabilities. In our example the estimate is 0.990 which satisfies the test stopping criterion.

6 Conclusion and Outlook

We have presented a new approach for the systematic generation of interaction test cases from state-based component test models. In the first step, reduced component test models were constructed, focusing on the component interactions and the connections between different interactions. In the next step, the reduced models were composed into interaction test models. An interaction profile was annotated to enable the statistical generation of representative test cases. Finally, we applied structural coverage criteria and random search to automatically derive

test cases. Based on the test results and the underlying interaction profile, the reliability w.r.t. component interactions could be estimated by applying model-based statistical testing.

The approach was initially applied to a simplified demonstrator from the automotive domain. In the feasibility study, we were able to significantly reduce the complexity of the models used for automated test case generation. The interaction profile allowed the generation representative test cases for the integration testing of a component-based system. The quality of automatically generated test cases especially the failure detection capabilities have to be assessed in a more complex industrial case study.

Future work will comprise the determination of an appropriate integration strategy based on the component models. Furthermore, we aim at improving the algorithms used for the construction of the interaction test models. In the current status of the work we have not considered the reliability of the underlying communication channels. This could be another research topic for the future.

This work was funded by the German Federal Ministry of Education and Research (BMBF) in the context of the project ViERforES (No.: 01 IM08003 B).

7 References

[ABJ07] Ali S., Briand L. C., Jaffar-ur Rehman M., Asghar H., Zafar Z., Nadeem A.: A state-based approach to integration testing based on UML models, Journal of Information and Software Technology (Elsevier), vol. 49(11), 2007

[Be07] Benz S., Combining test case generation for component and integration testing, 3rd International Workshop on Advances in Model-Based Testing 2007, London, pp.23-33, 2007

[BRT03] van der Bijl M., Rensink A., Tretmans J.: Compositional Testing with ioco. 3rd FATES, pp. 86-100, 2003

[DV07] Desmoulin A., Viho C.: A New Method for Interoperability Test Generation. 19th TestCom/FATES, pp.58-73, 2007

[HIM00] Hartmann, J.: Imoberdorf, C.; Meisinger, M.: UML-Based Integration Testing, In: Proc. of the Intl. Symposium on Software Testing and Analysis Portland, USA, pp.60-70, ACM Press, 2000.

[HP05] Huo j., Petrenko A.: Covering Transitions of Concurrent Systems through Queues. 16th ISSRE, pp. 335-345, 2005

[KCT02] Koppol P. V., Carver R. H., Tai K.-C.: Incremental integration testing of concurrent programs. IEEE Transactions on Software Engineering, 28(6):607–623, 2002

[LR01] Lyer S., Ramesh S.: Apportioning: a technique for efficient reachability analysis of concurrent object-oriented programs. IEEE Transactions on Software Engineering, 27(1):1037–1056, 2001.

[Mu93] Musa, J.D., "Operational Profiles in Software Reliability Engineering," IEEE Software Magazine, vol 10(2), p.14-32, 1993

[PP05] S. Prowell, J. Poore, Computing system reliability using Markov chain usage models, Journal of Systems and Software, Vol. 73, No. 2, October 2004, 215 - 225.

[RHP08] Robinson-Mallett C., Hierons R., Poore J., Liggesmeyer P.: Using communication coverage criteria and partial model generation to assist software integration testing. Software Quality Journal 16(2): 185-211 (2008)

[Pr03] Pretschner A.: Compositional Generation of MC/DC Integration Test Suites. Electr. Notes Theor. Comput. Sci. 82(6): (2003)

[SP00] K. Sayre, J. Poore, "Stopping criteria for statistical testing", Information and Software Technology, Vol. 42, No. 12, September 2000

[SWK09] Stefanescu A., Wieczorek S.,Kirshin A.. MBT4Chor: A model-based testing approach for service choreographies. ECMDA-FA 2009, vol.5562, Springer LNCS, pp. 313-324, 2009

[WPT95] Walton G., Poore J., Trammell C.: Statistical Testing of Software Based on a Usage Model. Softw., Pract. Exper. 25(1): 97-108 (1995)

Modelle als Mittel zur Verstärkung von strukturellen Tests sicherheitskritischer Software

Dirk Wischermann, Wolfgang Schröder-Preikschat
Friedrich-Alexander-Universität Erlangen-Nürnberg
Lehrstuhl für Verteilte Systeme und Betriebssysteme
Erlangen, Deutschland
{dw,wosch}@cs.fau.de

Abstract:
Software kann durch eine gezielte Überdeckung des Programmtextes mit Tests systematisch auf Fehler abgesucht werden. Viele Fehler treten aber nur bei der Ausführung ganz *bestimmter Pfade* durch den Programmtext auf. Da die Zahl dieser Pfade mit der größe des Programms schnell ansteigt, sinkt die Wahrscheinlichkeit der Fehleraufdeckung. Mit diesem Aufsatz stellen wir eine Methode vor (Arbeitstitel PERSPECTIVE TESTING), um mit Verhaltensmodellen eine Auswahl besonders relevanter Testfälle zu treffen und diese effektiv zu beurteilen. Die Modelle werden dabei in zweierlei Hinsicht genutzt: Zum gezielten Abtesten der modellierten Verhaltensgesichtspunkte mit starken Überdeckungskriterien und zur Generierung von *Kontrollflussorakeln*, das sind im Code verwobene Prüfpunkte, die die Modellkonformität des Kontrollflusses überwachen. Dadurch unterscheidet sich unser Ansatz erheblich von den meisten modellbasierten Testansätzen, die vor allem einen hohen Automatisierungsgrad oder die Wiederverwendung von Modellen aus dem Entwicklungsprozess zum Ziel haben.

1 Einleitung

Software, die für formale Verifikation oder vollständiges Testen zu komplex ist, wird längst auch schon in sicherheitskritischen Bereichen angewendet. Umfassende funktionale Tests gegen eine formale Spezifikation können dabei sinnvoll von strukturellen Tests[1] ergänzt werden, mit denen eine gewisse Evidenz für die *Abwesenheit* von Fehlern ermittelt wird. Allerdings ist dabei gerade die Aufdeckung von Fehlern mit Kontrollflussbezug nur mit geringer Wahrscheinlichkeit gewährleistet, da die Zahl der zu testenden Pfade einfach zu groß wird. Es gilt also, möglichst repräsentative Pfade für die Ausführung vorzusehen.

Modelle können beim Testen in vielerlei Hinsicht genutzt werden [UPL06]. Oft wird mit Modellen von der komplexen Struktur des Programmes abstrahiert und Testfallauswahl und Testfortschritt anhand von struktureller Überdeckung *des Modells* gemessen und dargelegt ("Grey-Box Testen", z.B. in [PSN09]). Damit lässt sich aber nur bedingt Vertrauen in die Richtigkeit *des Codes* gewinnen, denn oft korrespondiert eine Modellentität zu mehreren Stellen im Code. Genau hier setzt unsere Methode an, indem Verhaltensmo-

[1] Für die Grundbegriffe strukturellen Testens verweisen wir auf einschlägige Literatur, exemplarisch [Bei90].

delle auf den Code projiziert werden und der dadurch ausgezeichnete Code mit starken Überdeckungskriterien getestet wird (reines "White-Box Testen"). So können systematisch Fehler mit Kontrollflussbezug aufgedeckt werden. Der Ansatz eignet sich für Integrationstesten (Test auf die richtige Verwendung von Schnittstellen) sowie allgemein für den Test von Komponenten auf die Einhaltung modellierter Verhalten oder Protokolle.

Übersicht Der Hauptteil des Aufsatzes ist zweigliedrig: In Abschnitt 2.1 werden die von uns verwendeten Modelle vorgestellt und ihre Bedeutung am Beispiel erläutert. Im Abschnitt 2.2 wird ausgeführt, wie daraus Überdeckungsmetriken und Kontrollflussorakel gewonnen werden. Da es sich um eine laufende Arbeit handelt, präsentieren wir noch keine Ergebnisse. Eine kurze Zusammenfassung rundet die Arbeit ab.

2 Testen nach Gesichtspunkten

Bei PERSPECTIVE TESTING setzen wir endliche Automaten ein, die einzelne zu testende Verhaltensaspekte (abgekürzt CUT, Concern under Test) einer Komponente näher beschreiben. Unser Ansatz gewinnt aus den Modellen zwei Mittel, um einen solchen CUT besonders gut nach Fehlern mit Kontrollflussbezug (zu denen mehrere Programmzeilen beitragen) abzusuchen:

1. Eine gezielte Verstärkung der Überdeckungskriterien für den CUT.
2. Automatisch generierte *Kontrollflussorakel*, die solche Fehler durch eine dem Automaten entsprechende Instrumentierung des Codes aufdecken.

Insbesondere decken die Kontrollflussorakel auch Fehler auf, die sich (noch) nicht in Softwareversagen niedergeschlagen haben. Ein einfaches Beispiel eines CUT, der von Unit-Tests häufig nicht ausreichend erfasst wird, ist die Freigabe einer wiederverwendbaren Resource (Speicher, eine Datei oder eine Sperrvariable) nach deren Verwendung. Die funktionale Bewertung der Testläufe sowie die eigentliche Erzeugung der Testfälle sind nicht Thema dieses Aufsatzes.

Wir begleiten unsere Ausführungen mit einem realistischen Beispiel. Das Szenario folgt einem typischen Aufbau sicherheitskritischer Software, gegeben durch ein Bibliotheksbetriebssystem, das (von einem Systemintegrator) zusammen mit Treibern und Anwendungskomponenten übersetzt und auf ein eingebettetes System mit Sensoren und Aktoren verbracht wird. Als CUT betrachten wir die differenzierte Synchronisation von Aktivitätsträgern beim Zugriff auf Sensoren, Aktoren und gemeinsamen Speicher.

2.1 Gestalt und Bedeutung von CUT-Modellen

CUT-Modelle sind endliche Automaten, die eine *gerichtete* Abstraktion vom Code darstellen: Bezüglich der betrachteten Zuständigkeit bleiben sie praktisch auf Ebene des Codes,

andere Belange werden dagegen völlig ausgeblendet. Den direkten Code-Bezug stellt die Menge der *relevanten* Operationen her (ggf. mit Jokerzeichen[2], die von dem Automaten berücksichtigt werden sollen. Wir betrachten solche Modelle in zwei verschiedenen Rollen: Als *Spezifikationsmodelle* geben sie ein grundsätzlich erlaubtes Verhalten an und als *Designmodelle* stellen sie das vom Entwickler intendierte Verhalten der Software dar.[3]

In unserem Beispiel wird ein globales Regelwerk für die Verwendung von Sperrvariablen vorgegeben (siehe Abbildung 1). Das Protokoll verhindert Verklemmungen durch die Vermeidung zirkulärer Wartebedingungen. Dies kann leicht (zumindest für eine feste Zahl an Instanzen) mit einem Modelchecker überprüft werden. Ein Systemintegrator kann auf dieser Grundlage sämtliche Komponenten (Treiber, Anwendungen), die in das System eingebracht werden sollen, überprüfen und Verklemmungen zuverlässig ausschließen.

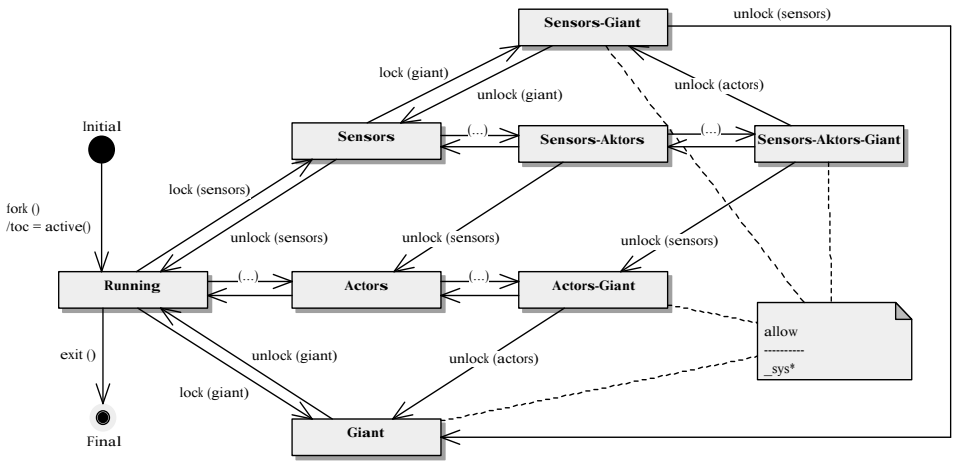


Abbildung 1: Allgemeiner Reihenfolgezwang für Sperrvariablen

In unserem Szenario spezialisiert der Entwickler diese Vorgabe dahingehend, dass je ein Faden[4] für das Auslesen der Sensoren und für die Bedienung von Aktoren verwendet wird. Ausserdem modelliert er für den Datenaustausch zwischen beiden die Verwendung eines beschränkten Puffers mit Semaphoren. Das Design der beiden Fäden (Abbildung 2) schränkt weiter ein, dass jeder Faden nur die tatsächlich benötigten Sperrvariablen auch belegen darf.

[2]Das für PERSPECTIVE TESTING gewählte Mittel ist derzeit die AspectC++ Pointcut Sprache [SGSP02]. Allgemein ist eine *Traceability* zwischen Code und Modell notwendig.

[3]Design Modelle spezialisieren Spezifikationsmodelle, siehe [WSP10]. Bei Design-Modellen kann *zusätzlich* zur Code-Überdeckung auch die Überdeckung von Modell-Entitäten betrachtet werden. Das zeigt, ob das Modell vollständig implementiert wurde.

[4]Faden: Engl. Thread of Control (toc)

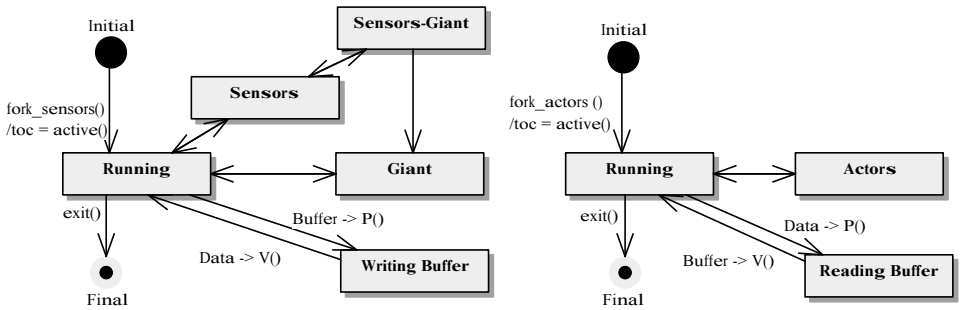


Abbildung 2: Modelle für Aktor- und Sensorthread mit Semaphorenbenutzung

2.2 Anwendung der Modelle beim Testen

Die Modelle werden auf zwei Arten genutzt: Zum einen zeichnen sie Teile des gesamten Programmtextes als *relevant* für einen Belang aus, was eine gezielte Verstärkung der Überdeckungskriterien erlaubt. Zum anderen werden aus ihnen *Kontrollflussorakel* extrahiert, die für jeden Testfall bewerten, ob der beschrittene Pfad dem Modell genügt.

Modellbasierte Überdeckungskriterien Voraussetzung für die Definition der verstärkten Überdeckungskriterien ist die Festlegung, wann Ausführungspfade durch den Programmtext als gleichwertig für einen CUT angesehen werden sollen:

Definition 1 *Zwei Pfade p und p' heissen **äquivalent bezüglich eines CUT-Modelles** $\mathcal{M}_D$ ($p \equiv|_{\mathcal{M}_D} p'$), wenn sie die selben relevanten Operationen von $\mathcal{M}_D$ in gleicher Reihenfolge durchlaufen.*[5]

Dies impliziert auf natürliche Weise eine Vereinfachung des Kontrollflussgraphen, sodass dieser nur noch die relevanten Pfade repräsentiert. Das erlaubt es uns, *beliebige* konventionelle Überdeckungskriterien auf die beschnittene Version des Kontrollflussgraphen zu übertragen. Für die Aufdeckung kontrollflussbezogener Fehler sind vor allem die Pfadkriterien (mit und ohne Schleifenbegrenzung) von Bedeutung; es können aber auch Kriterien der Bedingungsüberdeckung übertragen werden, indem sie nur auf Verzweigungen im beschnittenen Kontrollflussgraphen angewendet werden.

Die Modelle des Beispiels verlangen, die wechselseitige Puffer-Synchronisation mit Semaphoren *gemeinsam* mit der übergreifenden Synchronisation mit Sperrvariablen zu testen. Ein Pfadkriterium würde daher auch die Pfade enthalten, auf denen während der Pufferverwendung noch das `Actors`-Lock gehalten wird. Um einen solchen Fehler (der sich kaum manifestiert) tatsächlich aufzudecken, eignen sich *Kontrollflussorakel*.

[5] Eine abschwächende Definition würde auch Pfade identifizieren, die Abfolgen gleich lautender relevanter Operationen enthalten.

Kontrollflussorakel als zustandsbasierte Selbsttests Kontrollflussorakel sind das Mittel zur tatsächlichen Aufdeckung kontrollflussbezogener Fehler (beispielsweise fehlende Methodenaufrufe, falsche Methodenreihenfolge oder Aufrufe am falschen Objekt). Sie entscheiden für jede relevante Operation,

- ob eine neue Automateninstanz (s.u.) benötigt wird (bei daten- oder threadbezogenen Verhalten),
- ob eine Operation gerade erlaubt ist
- und ob die Operation einen Zustandsübergang auslöst.

Die CUTs im Beispiel werden fadenweise beobachtet, es wird also beim Starten eines neuen Thread eine den Automaten repräsentierende Datenstruktur angelegt und bei relevanten Operationen aktualisiert. Kommt es zu einer Verletzung des Synchronisationsprotokolls, kann eine Fehlermeldung mit der Historie von relevanten Codestellen zur Unterstützung der Fehlersuche ausgegeben werden. Der besondere Vorteil liegt darin, dass die Fehler oft früher und unabhängig von ihrer Manifestation aufgedeckt werden als durch ein konventionelles Orakel. Der Rückschluss auf die eigentliche Fehlerstelle wird dadurch vereinfacht.

Faktorisierung von Testobliegenheiten und Gestaltung für Testbarkeit Modelle dürfen sich ausdrücklich in Zuständen und Kanten überschneiden. Das ist ein Vorteil des Testers gegenüber dem Entwickler, der es ihm erlaubt, Testobliegenheiten entlang von Belangen zu *faktorisieren*, anstatt sie entlang der Programmstruktur zu spalten. Mit der Möglichkeit, beliebige Kriterien für verschiedene CUTs anzugeben, kann eine Komponente *adäquat* in dem Sinne getestet werden, dass wichtige Eigenschaften besser überprüft werden als andere. Bemerkenswert ist, dass der Testaufwand durch die Trennung der Belange erheblich eingedämmt werden kann: Die Anwendung von Pfadkriterien auf mehrere Belange *einzeln* verhält sich diesbezüglich nur additiv (statt multiplikativ).

In der gezielten Verstärkung von Überdeckungskriterien liegt ein Mittel, um die Gestaltung der Software zu Gunsten der *Testbarkeit* sicherheitskritischer Eigenschaften zu gewährleisten. Idealerweise wird eine solche Eigenschaft als Automat modelliert, mit einem Modelchecker formal überprüft und dann mit möglichst starken Überdeckungskriterien, die nur mit entsprechend testbarem Design zu erreichen sind, getestet. Beispielsweise verbietet (im Extremfall) die Forderung nach Pfadüberdeckung, dass relevante Operationen in Schleifen einbezogen werden. Andere Aspekte der Softwaregestaltung (Wiederverwendung, Variabilität) müssen sich der Testbarkeit dann nötigenfalls unterordnen.[6]

Verwandte Arbeiten und Ausblick Es gibt ausgesprochen viele verwandte Arbeiten, beispielsweise aus den Bereichen von modellbasiertem Testen, Integrationstesten, AOP, Code Slicing, Modelchecking oder Policy Enforcement. Um den Rahmen nicht zu sprengen, verweisen wir nur kurz auf zwei Arbeiten: Das SLAM Projekt [BR02] (enthalten

[6] Brückenpfeiler werden auch in erster Linie dick genug gebaut – auch wenn das mehr kostet und vielleicht nicht so gut aussieht.

im *Static Driver Verifier* von Microsoft) arbeitet zwar statisch, ist aber dennoch eng mit unserem Ansatz verwandt. Dort werden Modelle aus dem Code extrahiert und auf (insgesamt 65) temporallogische Eigenschaften untersucht. Die Vorgehensweise ist allerdings auf kleine Protokolle beschränkt. Als Ausblick sei die Idee erwähnt, mit symbolischer Ausführung [CDE08], *gezielt* nach Testfällen zu suchen, die eine Protokollverletzung nachweisen. Symbolische Ausführung könnte sich auch eignen, nicht überdeckte Pfade als nicht ausführbar nachzuweisen, indem die symbolisch aus den Fallunterscheidungen akkumulierten Bedingungen als widersprüchlich offenbart werden.

3 Fazit

In diesem Aufsatz wurden zwei Mittel des modellbasierten strukturellen Testens vorgestellt, die sich zur gezielten Aufdeckung von Fehlern mit Kontrollflussbezug eignen. Ausgehend von endlichen Automaten (idealerweise mit Modelchecking geprüft), wurden (Pfad-)Überdeckungskriterien *auf den Code* projiziert. Sie können aufzeigen, ob eine Zuständigkeit ausgiebig genug getestet wurde. Kontrollflussorakel stellen dabei sicher, dass der Kontrollfluss einem intendierten Muster entspricht. Damit stellt PERSPECTIVE TESTING Mittel zur Verfügung, die dynamischen Abläufe möglichst strikt an die Vorgaben zu binden.

Literatur

[Bei90] Boris Beizer. *Software testing techniques (2nd ed.)*. Van Nostrand Reinhold Co., New York, NY, USA, 1990.

[BR02] Thomas Ball und Sriram K. Rajamani. The SLAM project: debugging system software via static analysis. *SIGPLAN Not.*, 37(1):1–3, 2002.

[CDE08] Cristian Cadar, Daniel Dunbar und Dawson R. Engler. KLEE: Unassisted and Automatic Generation of High-Coverage Tests for Complex Systems Programs. In *8th USENIX Symposium on Operating Systems Design and Implementation, OSDI 2008, December 8-10, 2008, San Diego, California, USA, Proceedings*, Seiten 209–224, 2008.

[PSN09] Florin Pinte, Francesca Saglietti und Achim Neubauer. "Visualisierung überdeckter sowie zu überdeckender Modellelemente im modellbasierten Test". *Beiträge der 39. Jahrestagung der Gesellschaft für Informatik e.V. (GI)/ Workshop Modellbasiertes Testen (MOTES) Vol. P-154 of Lecture Notes in Informatics (LNI)*, 2009.

[SGSP02] Olaf Spinczyk, Andreas Gal und Wolfgang Schröder-Preikschat. AspectC++: An Aspect-Oriented Extension to C++. In *40th (TOOLS Pacific '02)*, Seiten 53–60, Sydney, Australia, Februar 2002.

[UPL06] Mark Utting, Alexander Pretschner und Bruno Legeard. A taxonomy of model-based testing. Bericht, University of Waikato, April 2006.

[WSP10] Dirk Wischermann und Wolfgang Schröder-Preikschat. Separating Testing Concerns by Means of Models. In *1st ECOOP Workshop on Testing Object-Oriented Systems (ETOOS)*, June 2010. to appear.

Test Case Generation for Visual Contracts Using AI Planning

Matthias Schnelte, Baris Güldali
Software Quality Lab, University of Paderborn
Warburger Str. 100, 33098 Paderborn/Germany
{mschnelte,bguldali}@s-lab.upb.de

Abstract: In this paper, we propose a novel approach for computing test case preambles using visual contracts and AI Planning. In unit testing, preambles are required for setting the class under test into a controlled state (*prestate*). The class operation can then be invoked with test inputs. In previous research, we have used model checking for computing preambles. In this paper, we show how preamble computation is conducted by AI Planning and discuss its differences to model checking.

1 Introduction

Model-based testing facilitate abstract models of system under test (SUT) or of its environment for generating test artifacts, e.g. test cases, test oracles or test execution environments. Thereby some interesting research questions arise: (1) how should test cases be selected such that a adequate *coverage* of models is reached; (2) how can we compute *preambles* for test cases such that their requirements are fulfilled; (3) how can we compute expected results for the test cases (*oracle problem*).

In previous work, we have described how visual contracts (VCs) – a visual notation of pre/-post conditions – can be used for model based testing [EGL06]. VCs aim at describing the SUTs behavior by specifying object structures before (*precondition*) and after (*postcondition*) its execution. For testing, we generate a prestate and test inputs fulfilling the precondition of VC and compute the expected test results conform to the postcondition. For computing preambles, we have proposed two techniques: (1) artificial preambles, which simulate a prestate [EEG08]; (2) natural preambles, which construct the prestate in a realistic way [EGL06]. For computing natural preambles, we have applied *model checking* techniques. Other researchers have applied constraint logic programming to solve the same problem [SCP04].

In this paper, we want to discuss the suitability of AI Planning for computing preambles and compare it to graph transformations-based model checking. We want to show that AI Planning outperforms this kind of model checking. In the next section, we give an overview on AI Planning. Sect. 3 introduces the test case generation using AI-Planning in detail. In section 4, we report on early experiences on a case study and compare the result to graph transformation-based model checking.

2 AI Planning

The problem addressed by AI Planning [GNT04] can be described as follows: how can a given initial state be transformed into a desired goal state by using a set of actions? The inputs for this problems are a domain theory and a problem description. The *domain theory* describes the set of actions changing the state. Actions have *preconditions* that must hold before its execution and *effects* that describe the state changes after the execution.

The *problem description* contains a set of conditions that describe the *initial state* and others that must hold for the *goal state*. Given the domain theory and problem description, a planner searches for a sequence of actions whose effects transform the initial state until a state is reached that conforms to the goal state. Having shortly introduced the AI Planning, we explain as next, how test cases can be generated using AI Planning.

3 Test Case Generation using AI Planning

In this paper, we deal with *quasi modal* classes, where their behavior depends on the system state, in which they are tested. While testing such classes, first the system state must be set to a controlled state (*prestate*). Then the operation under test is invoked. The changes in the system state after the test invocation are compared to the specification of the expected state. Particularly, we derive a prestate from the precondition of the visual contract (VC) which specifies the operation under test. After setting the prestate, the operation is invoked with some input parameters conforming to the prestate. Finally, the changes in the system state (*poststate*) is checked against the postcondition of the VC.

As next, we will focus on deriving and setting a prestate. For setting it, we need a *preamble*, which is a sequence of class operations. In [EGL06], we explain how a preamble can be computed by using model checking. In section 3.2, we explain, how preamble computation can be done by using AI Planning. Finally, we compare these two techniques.

3.1 Derivation of Prestates from Visual Contracts

VCs describe the behavior of an operation by specifying requirements on the system state before and after its execution [LSE05]. These requirements are called pre- and postconditions, respectively. The VC in Figure 1.a illustrates the pre- and the postconditions for an operation *commitReservationList* of a hotel booking system. The objects left to the arrow represent the precondition, and the objects right to the arrow represent the postconditions. The objects are typed over a *class diagram*, which is not shown here. The object *hr:HotelReservation* is a *multiobject*, which can address one or many objects of type *HotelReservation*. The objects with a gray underground are called negative application condition (*NAC*), which are not allowed to exist in the prestate. In our example, the customers can add one or many hotels into their reservation lists. After committing the reservation, a receipt is produced and the reservations are deleted from the reservation list.

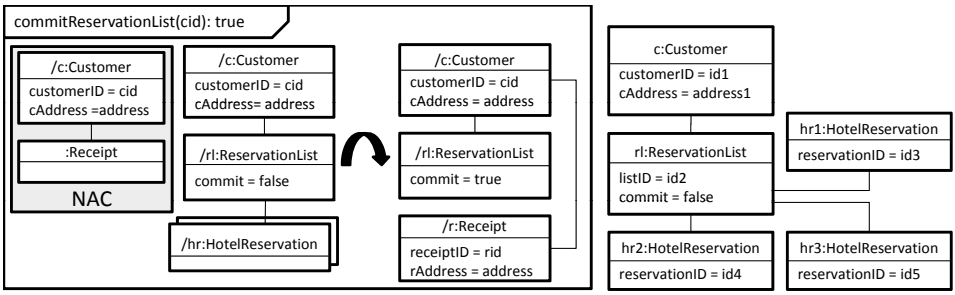


Figure 1: a) VC for operation commitReservationList, b) exemplary prestate

For testing the operation *commitReservationList*, we derive prestates from the precondition. For that, first, the objects and object links from the precondition are instantiated, then the attributes are filled with concrete values. Thereby we use structural coverage criteria [UL07] for the multiobjects and well-known data selection criteria (e.g. boundary values, equivalence classes) for the attributes [EEG08]. Figure 1.b. shows an exemplary prestate.

In [EGL06], we have shown how to compute a preamble for setting the prestate in a natural way by calling other operations. For computing the preamble, we first translated VCs into graph production rules. Then we have conducted a reachability analysis using the model checker GROOVE [KR06]. GROOVE represents the states of the system as graphs and references to graphs, instead of bit vectors as in most explicit state representing approaches. For the analysis, GROOVE first computes a graph transition system and secondly translates it to an ordinary Kripke structure. Then, a counterexample is computed by CTL model checking which represents the preamble for setting the prestate. We have experienced in our case, that GROOVE does not scale very well with the growing number of objects. Next we want to show how this problem can be tackled by using AI planning.

3.2 Translating visual contracts to PDDL

To use AI planning for the computation of preambles, the first step is to translate the VCs into a representation that can be processed by a planning tool. The most popular representation for the description of planning problems is the planning domain definition language (PDDL) [McD98], that serves as the input for most modern planners. In PDDL, a logic formalism, which is similar to first order predicate calculus, is used to describe the world state and actions. In this section, we describe a translation process that transforms VCs into an action representation noted in PDDL.

VCs and the action representation are similar in the sense that both use a pre-/postcondition formalism. Therefore, each VC is translated into one action where the left hand side of the VC is translated into the preconditions in PDDL and the right hand side is translated into the effects of the action.

For representing objects, attributes and links, the translation process introduces several types and predicates that are derived from the classes used in the VCs. For each class, a type with the name of the class is created. To express the fact that two objects *o1* and *o2*, are linked, the predicate `isLinkedTo(?o1 ?o2)` is used. Further predicates are created to represent the attributes of a class. For each attribute of a class, a predicate `attribute_` *className_attributeName* is introduced. If the attribute *a* is of type object, the predicate has two parameters: *a* and *o*. The meaning of this predicate is that *a* is an attribute of object *o*. If the attribute is of type boolean, it is sufficient to just provide the parameter *o*. If the predicate is *true* it means that *o*'s boolean attribute *a* is *true*.

Unfortunately, PDDL does not support the creation and destruction of objects. Therefore, we are using a work-around as proposed in [KB06]. All objects that can potentially be created in the future have to be specified in advance in the problem description. The creation and destruction of an object *o* is than realized by using the predicate `active(?o)`. If this predicate is *true*, the object *o* has been created. To delete the object, the predicate is set to *false*. Listing 1 shows some of the types and predicates for the hotel booking example.

```
(define (domain HotelReservation)
 (:types
   Customer - object ReservationList - object HotelBookingSystem - object
   Hotel - object HotelReservation - object Date - object Receipt - object String)

 (:predicates
   (active ?o - object)
   (isLinkedTo ?o1 - object ?o2 - object)
   (attribute_Customer_customerID ?o - Customer ?a - String)
   (attribute_ReservationList_commit ?o - ReservationList))
```

Listing 1: Type and predicate definition in PDDL.

Using these predicates, the translation process creates one action for each visual contract. As an example we provide an action for the operation `commitReservationList` as shown in Figure 1.a. The action definition can be seen in listing 2. The parameters of the action correspond to the object and attribute names in the visual contract. Lines 4 - 7 encode the positive part of the contract's precondition. Lines 8 - 9 are used to encode the negative precondition. The effects in lines 10 - 11 encode the part of the postcondition, that states that all *HotelReservation* objects that are linked to the *CustomerReservation* object are deleted. Lines 12 - 13 encode the rest of the postcondition.

After the VCs have been translated into PDDL, a planner can be used to compute a sequence of actions (*preamble*) that establish a desired system state.

3.3 Preamble computation using AI-Planning

To compute the preamble, the initial state and the goal state has to be defined. Listing 3 shows a problem definition in PDDL for the prestate shown in Figure 1.b. Lines 3 to 5 define all objects that can possibly be used. The prestate we want to establish is defined in line 9 following the same principles as we used in translating VCs to PDDL, but instead

```
(:action commitReservationList
  :parameters (?c - Customer ?rl - ReservationList ?r - Receipt
    ?cid - String ?lid - String ?address - String)
  :precondition (and  (active ?c) (active ?rl) (not (active ?r))
    (attribute_Customer_customerID  ?c ?cid)
    (attribute_Customer_customerAddress ?c ?address)
    (not (attribute_ReservationList_commit ?rl)) (isLinkedTo ?c ?rl)
    (not (exists (?receipt -  Receipt) ; The NAC
    (isLinkedTo ?c ?receipt)) ))
  :effect (and (forall (?hr - HotelReservation)
    (when (isLinkedTo ?rl ?hr)(and (not (active ?hr))(not (isLinkedTo ?rl ?hr)))))
    (active ?r) (isLinkedTo ?c ?r) (attribute_Receipt_receiptaddress ?r ?address)
    (attribute_ReservationList_commit ?rl)))
```

Listing 2: Action definition for createReservationList.

of using variables for the objects we are using constants. The *init* state is defined in line 8 as a state in which the only existing object is the hotel booking system.

Calling the planner with this problem definition results in a sequence of actions where the action's parameters are replaced with concrete objects. This sequence can be used as the preamble for setting the prestate.

```
(define (problem computePreamble)
  (:domain pddltest)
  (:objects  rl1, rl2, rl3 - ReservationList hr1, hr2, hr3 - HotelReservation
    c1, c2, c3 - Customer re1, re2, re3 - Receipt
    id1, id2, id3, id4, id5  - String hbs - HotelBookingSystem ...  )
  (:init (active hbs))
  (:goal (and (active c1)(isLinkedTo c1 rl1) (active rl1)
    (attribute_ReservationList_listID rl1 id2) ... ))
```

Listing 3: Problem definition for preamble computation.

4 Evaluation and Conclusion

We conducted some performance analysis for our approach using a realistic but small example. We used the planner LAMA [RW08] and searched for preambles with increasing complexity due to the number of objects used and the length of the required preamble. In [EGL06], we used the graph transformation-based model checker GROOVE for computing the preambles. Table 1 compares these two approaches. Column 3 shows the states generated by the planner LAMA followed by the states generated by GROOVE. GROOVE always generates the whole reachable state spaces before searching for the goal state. Therefore, the state space explodes within the number of objects. The used planner on the other hand performs a heuristic search and hence generates only a portion of the whole state space, namely that parts of the state space that are - according to the heuristic - most likely to contain the goal state. It can be seen that the planner scales very well within the problem size, whereas GROOVE does not scale at all.

We conclude that the VCs can be translated very easily and automatically to action repre-

Nr. of objects	Preamble length	States generated (LAMA/GROOVE)
4	4	14 / 10
7	5	27 / 152
10	6	44 / 3248
13	7	65 / > 20000
16	8	90 / n.a.

Table 1: Evaluation

sentation in PDDL due to the mutual use of pre/post paradigm. Furthermore, early results show the good performance of the PDDL-based approach. By stating the preamble computation problem as a planning problem in PDDL, we can reuse the mature technology and tools developed for the area of AI planning. The potential of using AI planning for test case generation has already been identified elsewhere [AZS+02]. In contrast to our work, the planner representation and the test objectives are derived *manually* from an UML model, while our approach is fully automated.

References

[AZS+02] A. Amschler Andrews, C. Zhu, M. Scheetz, E. Dahlman, and A. E. Howe. AI Planner Assisted Test Generation. *Software Quality Journal*, 10(3):225–259, 2002.

[EEG08] Jens Ellerweg, Gregor Engels, and Baris Güldali. Modellbasierter Komponententest mit visuellen Kontrakten. In *INFORMATIK 2008, Band 1, GI*, pages 211–214. GI, 2008.

[EGL06] Gregor Engels, Baris Güldali, and Marc Lohmann. Towards Model-Driven Unit Testing. In *Proc. of the Workshops at MoDELS 2006*, pages 182–192. Springer, 2006.

[GNT04] Malik Ghallab, Dana Nau, and Paolo Traverso. *Automated Planning: Theory and Practice*. Morgan Kaufmann, 2004.

[KB06] A. Korvarsky and M. Buro. A first Look at Build-Order Optimization in Real-Time Strategy Games. In *Proceedings of the GameOn Conference*, 2006.

[KR06] Harmen Kastenberg and Arend Rensink. Model Checking Dynamic States in GROOVE. In *Proc. of SPIN 2006*, LNCS 3925, pages 299–305. Springer-Verlag, 2006.

[LSE05] Marc Lohmann, Stefan Sauer, and Gregor Engels. Executable Visual Contracts. In M. Erwig and A. Schürr, editors, *Proc. of VL/HCC'05*, pages 63–70, 2005.

[McD98] D. V. McDermott. PDDL - The Planning Domain Definition Language. Technical report, Yale Center for Computational Vision and Control, 1998.

[RW08] S. Richter and M. Westphal. The lama planner using landmark counting in heuristic search. In *Proceedings of IPC*, 2008.

[SCP04] B. Legeard S. Colin and F. Peureux. Preamble Computation in Automated Test Case Generation using Constraint Logic Programming. *STVR*, 14:213 – 235, 2004.

[UL07] Mark Utting and Bruno Legeard. *Practical Model-Based Testing: A Tools Approach*. Morgan Kaufmann, 2007.

Evaluierung des Fehlererkennungspotentials modellbasierter Komponenten- und Integrationstestfälle

F. Pinte und F. Saglietti

Lehrstuhl für Software Engineering (Informatik 11)
Universität Erlangen-Nürnberg
Martensstraße 3
91058 Erlangen (D)
pinte@informatik.uni-erlangen.de
saglietti@informatik.uni-erlangen.de

Abstract: Der vorliegende Beitrag präsentiert neueste Ergebnisse einer Evaluierungsstudie über die Effizienz modellbasierter Verifikationsansätze im Hinblick auf Anzahl und Art erkennbarer Fehler. Ein seit kurzem einsatzbereites Werkzeug zur vollautomatischen modellbasierten Testfallerzeugung lieferte hierfür ein wertvolles Forschungsinstrument, das in Bezug auf die Aufdeckung von Modellierungsfehlern und Implementierungsfehlern zur Erlangung weiterer Einsichten auf diesem Gebiet verhelfen konnte.

1 Einführung

Die weitgehend automatische Umsetzung modellbasierter Testverfahren war Gegenstand des mehrjährigen Verbundprojekts UnITeD. Im Rahmen dieses kürzlich abgeschlossenen Vorhabens wurde ein gleichnamiges Werkzeug zur Testfallgenerierung entwickelt. Zu vorgegebenen Überdeckungskriterien [SOP07] erzeugt UnITeD vollautomatisch Testsequenzen mit zugehörigen Eingabedaten sowohl für den Komponenten- [OSS07] als auch für den Integrationstest [PSO08]. Darüber hinaus unterstützt das implementierte Werkzeug auch die Visualisierung der erzielten Testüberdeckung [PSN09] sowie die systematische Erkennung und Ergänzung wieder verwendbarer Testläufe im Regressionstest [PBS08]. Die zu diesen Zwecken entwickelten Verfahren sind vor allem durch multikriterielle Optimierung gekennzeichnet; hierdurch wird sowohl die Maximierung der Testobjektüberdeckung als auch die Minimierung der Testfallanzahl verfolgt.

Das Werkzeug wurde anschließend – neben dem Piloteinsatz in industriellen Umgebungen – auch als wissenschaftliches Forschungsinstrument zur Evaluierung des Fehlererkennungspotentials unterschiedlicher modellbasierter Testverfahren eingesetzt. Ziel der hier vorgestellten Evaluierungsstudie ist es, die Effizienz modellbasierter Verifikationsansätze im Hinblick auf Anzahl und Art dadurch aufgedeckter Fehler sowie den damit verbundenen Aufwand vergleichend zu bewerten. Das Papier ist folgendermaßen gegliedert: In Abschnitt 2 werden unterschiedliche Mutationsklassen zur Simulierung entsprechender Fehlerarten eingeführt. Die anschließend in Abschnitt 3

vorgestellte Mutationsanalyse erlaubt eine experimentelle Evaluierung des Fehlererkennungspotentials im Hinblick auf Modellierungsfehler. Darüber hinaus wird in Abschnitt 4 über die Möglichkeit der Aufdeckung von Implementierungsfehlern berichtet.

2 Mutationsklassen

Zur Untersuchung modellbasierter Testfallmengen hinsichtlich ihres Fehlererkennungspotentials am Modell wurden zunächst unterschiedliche Fehlerarten identifiziert und klassifiziert, wobei unterschieden wurde zwischen:
- intramodularen Fehlern im Entwurf einer Komponente und
- intermodularen Fehlern in der Interaktion zwischen zwei Komponenten.

2.1 Intramodulare Mutationsklassen

Zur Simulation intramodularer Fehlerarten wurden folgende Mutationsklassen für zustandsbasierte Komponentenmodelle betrachtet:
- Trigger-Mutationen durch Entfernen, Hinzufügen bzw. Verfälschen des Auslösers eines Zustandsübergangs,
- Guard-Mutationen durch Entfernen, Hinzufügen bzw. Verfälschen eines Wächters an einem Zustandsübergang (durch Änderung der Vergleichsoperatoren, Boolescher Operatoren bzw. zugrunde liegender Variablen),
- Effekt-Mutationen durch Entfernen, Hinzufügen bzw. Verfälschen des von einem Zustandsübergang ausgelösten Ereignisses (durch Änderung der auszulösenden Variablenzuweisungen),
- Pre-State-Mutationen durch Änderung des Anfangszustandes einer Transition,
- Post-State-Mutationen durch Änderung des Endzustandes einer Transition,
- Transition-Mutationen durch Entfernen bzw. Hinzufügen eines Zustandsübergangs,
- Zustand-Mutationen durch Entfernen bzw. Hinzufügen eines Zustands.

2.2 Intermodulare Mutationsklassen

Zur Simulation intermodularer Fehlerarten wurden folgende Mutationsklassen für zustandsbasierte Modelle aufrufender und aufgerufener Komponenten betrachtet:
- Trigger-Mutationen durch Entfernen, Hinzufügen bzw. Verfälschen des Auslösers eines Zustandsübergangs,
- Guard-Mutationen durch Entfernen, Hinzufügen bzw. Verfälschen eines Wächters an einem Zustandsübergang,
- Effekt-Mutationen durch Entfernen, Hinzufügen bzw. Verfälschen des von einem Zustandsübergang ausgelösten Ereignisses (durch Änderung der auszulösenden Operationsaufrufe),
- Pre-State-Mutationen durch Änderung des Anfangszustandes einer Transition,
- Post-State-Mutationen durch Änderung des Endzustandes einer Transition.

3 Testfallgenerierung und Mutationsanalyse

Zur Analyse des Fehlererkennungspotentials wurden einerseits Testüberdeckungskriterien auf Komponentenebene (Zustands-, Transitions-, Transitionspaarüberdeckung), andererseits zustandsbasierte Integrationstestkriterien ausgewählt. Letztere beruhen auf dem Konzept der Überdeckung interagierender Transitionen [SOP07], d.h. von Paaren (t, t') von Transitionen, so dass als Effekt der Transition t in einer aufrufenden Komponente die Transition t' in der aufgerufenen Komponente traversiert wird. In [SOP07] wurden auf der Basis dieses Konzeptes 8 unterschiedliche Überdeckungsmaße definiert, deren Subsumptionshierarchie in Abb. 1 dargestellt wird.

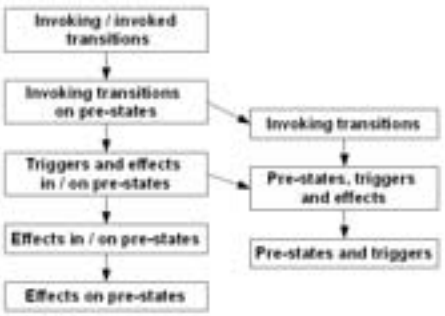


Abbildung 1: Subsumptionshierarchie der zustandsbasierten Integrationstestkriterien

Um das Fehlererkennungspotential eines Überdeckungskriteriums zu bewerten, wurde der Fehlererkennungsgrad von Testfallmengen betrachtet, die besagtes Kriterium erfüllen und mit dem im Projekt UnITeD entwickelten Werkzeug (s. [OSS07], [PSO08]) generiert wurden. Zur automatischen Ermittlung des Fehlererkennungsgrades wurden durch Anwendung unterschiedlicher Mutationsklassen (s. Abschnitt 2) zahlreiche Mutanten des ursprünglichen Modells generiert. Weicht nach Ausführung eines Mutanten mit den zu bewertenden Testfällen dessen von außen beobachtbares Verhalten von dem ursprünglichen Modellverhalten ab, so kann die zugrunde liegende Testfallmenge im Hinblick auf die Fehleraufdeckung der betrachteten Mutation als adäquat betrachtet werden. Die im Folgenden dargestellten Ergebnisse bzgl. der Anzahl erkannter Mutanten sind als konservativ zu betrachten, da sie auch Mutanten umfassen, die zum ursprünglichen Modell funktionsäquivalent und deshalb nicht erkennbar sind.

Die Mutationsanalyse wurde exemplarisch auf das Modell eines Softwaresystems angewendet, das die korrekte Positionierung einer Patientenliege beim Einsatz eines medizintechnischen Gerätes zu überprüfen hat. Zu diesem Zweck kontrolliert und verwaltet die Software einzelne Prüfschritte, die vor Auslieferung jeder Liege durchzuführen sind; darüber hinaus ist sie auch für die Generierung eines Protokolls über die durchgeführten Prüfschritte und deren Ergebnisse zuständig. Neben dem internen Verhalten der Software beschreibt das Modell auch die vom Benutzer wahrgenommene Mensch-Maschine Interaktion, also die Abfolge der dem Benutzer angebotenen Dialoge. Zur Realisierung dieses Verhaltens wurden 2 interagierende Software-Komponenten durch UML-Zustandsmaschinen modelliert.

3.1 Fehlererkennungspotential auf Komponentenebene

Zur Bewertung des Fehlererkennungspotentials auf Komponentenebene wurden sämtliche Mutationsklassen aus Abschnitt 2.1 wiederholt instanziiert und die dadurch erzielten Mutationen auf die komplexere der beiden Komponenten angewendet. Die generierten Testfallmengen erreichten jeweils folgende Modellüberdeckung:

- 100%ige Zustandsüberdeckung durch 8 Testfälle,
- 100%ige Transitionsüberdeckung durch 27 Testfälle,
- 97,8%ige Transitionspaarüberdeckung durch 80 Testfälle.

Abb. 2 zeigt die jeweiligen Anteile an erkannten Mutationsklassen.

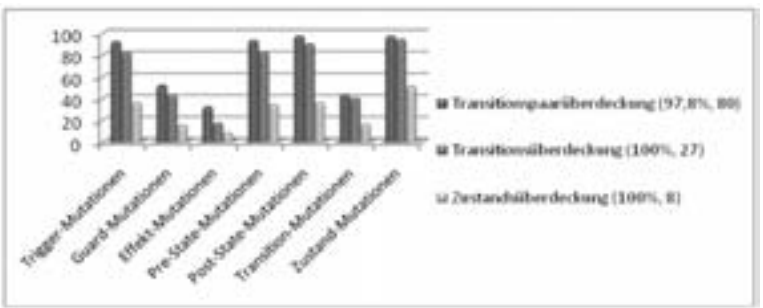


Abbildung 2: Ergebnisse der Mutationsanalyse für Komponententestkriterien

Erwartungsgemäß konnte das stärkste Kriterium, nämlich die Transitionspaar-überdeckung, trotz des höheren Aufwands (und der deshalb mit den vorhandenen Ressourcen nur unvollständig erzielbaren Überdeckung) hinsichtlich des Fehlererkennungsgrads am besten abschneiden. 63% aller Mutanten wurden identifiziert, während durch Transitionsüberdeckung nur 55% und durch Zustandsüberdeckung nur 23% aller Mutanten erkannt werden konnten.

Dabei muss der Fehlererkennungsgrad bzgl. der einzelnen intramodularen Mutationsklassen differenziert betrachtet werden. Bezogen auf Mutationen, die grundsätzliche Strukturänderungen enthalten, wie z.B. geänderte End- oder Anfangszustände von Transitionen, erreichte die Transitionspaarüberdeckung sehr gute Ergebnisse (über 90%). Mutationen, die geringfügige Änderungen im Datenfluss beinhalten, wie z.B. Effekt-Mutationen, wurden allerdings nur zu einem ca. 30%igen Anteil erkannt. Eine Erklärung hierfür liegt darin, dass die hier betrachteten Testkriterien auf strukturellen und nicht auf datenflussbasierten Überdeckungskonzepten basieren.

3.2 Fehlererkennungspotential der Integrationsteststrategien

Während im Komponententest die Funktionsweise innerhalb einzelner Komponenten getestet wird, prüft der Integrationstest speziell die Komponenteninteraktionen. Die implementierte Mutationsanalyse ermöglicht auch die Untersuchung der Fehlererkennung von Testfällen, die zustandsbasierten Interaktionsüberdeckungs-kriterien genügen. Die Ergebnisse sind in Abb. 3 illustriert.

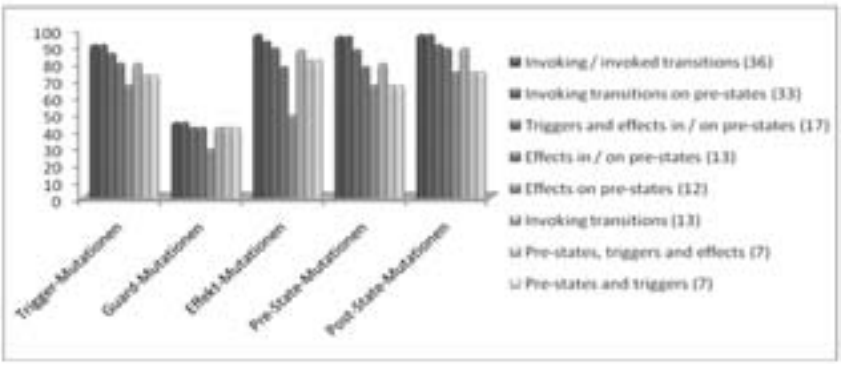


Abbildung 3: Ergebnisse der Mutationsanalyse für Integrationstestkriterien
(pro Kriterium Angabe der zu einer 100%igen Überdeckung benötigte Anzahl an Testfällen)

Wie bereits auf Komponentenebene, erkannte die Testfallmenge, die zum anspruchsvollsten Kriterium generiert wurde, die meisten Mutanten. In diesem Falle handelt es sich um das „Invoking / invoked transitions“ - Kriterium, das 84% der generierten Mutanten zu erkennen erlaubte. Mit Ausnahme der Guard-Mutationen erkannte diese Testfallmenge über 90% aller erzeugten Mutanten. Guard-Mutationen konnten allerdings nur zu einem 44%igen Anteil erkannt werden. Dies lässt sich dadurch erklären, dass ursprünglich generierte Testfälle nach Entfernung vorhandener Wächterbedingungen ihren Testablauf (und damit das Testverhalten) im Wesentlichen beibehalten. Nur geringfügig schlechtere Ergebnisse (83%) erzielte die Testfallmenge zum zweitstärksten Kriterium „Invoking transitions on pre-states“, die aus heuristischen Gründen zum Teil die gleiche Anzahl interagierender Transitionspaare erzielen konnte.

Abschließend sollte nicht unerwähnt bleiben, dass der so gestaltete Integrationstest - über die Erkennung intermodularer Fehler hinaus - auch die Aufdeckung intramodularer, einem vorausgegangenen Komponententest evtl. entgangener Fehler unterstützt. Um auch dieses Fehlererkennungspotential experimentell zu untersuchen, wurden sämtliche Mutationsarten aus Abschnitt 2 auf beide Systemkomponenten angewendet. Erwartungsgemäß konnte die gemäß höchster Integrationstestanforderungen generierte Testfallmenge mehr intermodulare Mutanten als intramodulare Mutanten aufdecken, da das zugrundeliegende Testkriterium ja auf die Überdeckung von Komponenten-interaktionen ausgerichtet war. Dennoch weist der relativ hohe Anteil (59%) an erkannten Komponentenfehlern darauf hin, dass ein automatisierter Integrationstest auch im Hinblick auf Schwächen vorausgegangener (bei vorgefertigten Komponenten u.U. ausgelagerten, kontextfremden) Verifikationsaktivitäten wertvolle Hilfe bieten kann.

4 Erkennungspotential hinsichtlich Implementierungsfehler

An der gleichen Anwendung wurde die beschriebene automatische Testfallerzeugung mit der manuellen Ermittlung (ohne Modellbeschreibung) verglichen. Zunächst wurde der Vergleich in Bezug auf den Aufwand beider Vorgehensweisen durchgeführt: die

Erstellung des Modells erforderte ca. 18 Stunden, während die manuelle Testfallermittlung etwa 12 Stunden in Anspruch nahm. Auf dieser Grundlage lohnt sich das modellbasierte Vorgehen nur unter der Annahme, dass es eine entsprechend höhere Qualität des Endprodukts in Aussicht stellt. Tatsächlich stellte sich beim anschließenden Vergleich beider Verfahren in Bezug auf die Testobjektüberdeckung heraus, dass die automatisch generierten Testfälle eine deutlich höhere Überdeckung interagierender Transitionen als die manuell erzeugten Testfälle erzielen konnten (100% vs. 58%); dementsprechend höher war auch deren Fehleraufdeckung, insbesondere in Bezug auf drei Implementierungsfehler, die die manuell gestalteten Tests nicht erkannten. Dabei handelte es sich im Wesentlichen um inkorrekte Reaktionen auf vom üblichen Nutzungsprofil abweichende Aktionen des Anwenders. Beispielsweise wurde ein anomales Verhalten seitens des Anwenders, wie etwa das Abbrechen eines Prüfschrittes vor Einstufung der Prüfergebnisse, nicht korrekt umgesetzt. Das höhere Fehlererkennungspotential modellbasierter Testfälle lässt sich durch die frühe Modellierung der Mensch-Maschine Interaktion unter Berücksichtigung entsprechender Ausnahmeszenarien erklären.

5 Zusammenfassung

In diesem Beitrag wurden Ergebnisse aus einer Evaluierungsstudie über die Effizienz modellbasierter Verifikationsansätze im Hinblick auf Anzahl und Art erkennbarer Fehler präsentiert. Die experimentelle Untersuchung erfolgte mit Hilfe eines seit kurzem einsatzbereiten Werkzeugs zur vollautomatischen modellbasierten Testfallerzeugung. Die Analyse umfasste die Erkennbarkeit von Modellierungsfehlern und von Implementierungsfehlern.

Literaturverzeichnis

[SOP07] Saglietti, F.; Oster, N.; Pinte, F.: Interface Coverage Criteria Supporting Model-Based Integration Testing. In M. Platzner, K.-E. Großpietsch, C. Hochberger, A. Koch (Eds.): ARCS '07 Workshop Proceedings, VDE Verlag, 2007.

[OSS07] Oster, N.; Schieber, C.; Saglietti, F.; Pinte, F.: Automatische, modellbasierte Testdatengenerierung durch Einsatz evolutionärer Verfahren. In Informatik 2007 - Informatik trifft Logistik, Lecture Notes in Informatics, Vol. 110, Gesellschaft für Informatik, 2007.

[PBS08] Pinte, F.; Baier, G.; Saglietti, F.; Oster, N.: Automatische Generierung optimaler modellbasierter Regressionstests. In Informatik 2008 - Beherrschbare Systeme dank Informatik, Lecture Notes in Informatics, Vol. 133, Gesellschaft für Informatik, 2008.

[PSN09] Pinte, F.; Saglietti, F.; Neubauer, A.: Visualisierung überdeckter sowie zu überdeckender Modellelemente im modellbasierten Test. In Informatik 2009 - Im Fokus das Leben, Lecture Notes in Informatics, Vol. 154, Gesellschaft für Informatik, 2009.

[PSO08] Pinte, F.; Saglietti, F.; Oster, N.: Automatic Generation of Optimized Integration Test Data by Genetic Algorithms. In Software Engineering 2008 - Workshopband, Lecture Notes in Informatics, Vol. 122, Gesellschaft für Informatik, 2008.

Workshop

Zukunftsfähigkeit unternehmensweiter Anwendungssysteme

Prof. Dr.-Ing. Norbert Gronau, Sandy Eggert

Flexibilität für PLM-Services in der Digitalen Fabrik

Karl Flieder

Externer Lektor für Geschäftsprozessmanagement sowie für
Business Process Management & Engineering (BPME)
Academia Nova, Schwechat
eai@karlflieder.at

Abstract: Eine durchgängige digitale Abbildung von Geschäftsprozessen, die sich in der Fertigungsindustrie von der Produktentstehung über die Fertigung bis hin zu Serviceprozessen erstrecken kann, unterstützt die geschäftliche Flexibilität und damit auch die Zukunftsfähigkeit von Anwendungen ganz wesentlich. Durch eine Integration geometrischer Daten und Informationen, wie sie im Rahmen der Digitalen Fabrik und auf der Basis einer serviceorientierten Architektur (SOA) Anwendung finden, können im Engineering Entwicklungs- und Änderungszyklen verkürzt, Prozesse stabilisiert sowie Qualität, Zeit und Kosten optimiert werden. Dieser Praxisbeitrag beantwortet die Frage wie Services für das Product Lifecycle Management (PLM) für mehrere Produktlinien wiederverwendbar gestaltet werden können.

1 Einleitung

Der Begriff Geschäftsprozessmanagement umfasst im Sinne einer allgemeinen Managementdefinition nicht nur die strategische und operative Planung von Prozessen, sondern auch deren Implementierung, Ausführung und Kontrolle [BR07]. Aus technischer Sicht wird damit zumeist eine Assoziation zu bestehenden Technologien der verbindenden oder verknüpfenden Prozessintegration wie Enterprise Application Integration (EAI) und serviceorientierten Architekturen (SOA) hergestellt ([AW09], [FNS06], [He07]). Eine serviceorientierte Architektur steht dabei für ein aktuelles Paradigma, ein Denkmuster, zur Gestaltung von Unternehmensarchitekturen [SA08], damit auch zur Steigerung der Flexibilität, die verschiedene Facetten annehmen kann [HR+10]. Mit einer SOA wird das Ziel einer an die Geschäftsprozesse ausgerichteten Infrastruktur verfolgt, die flexibel und leicht adaptierbar sein soll. Im Rahmen der Implementierung von determinierten, kurz laufenden und datenzentrierten Prozessen kommen typischerweise XML-Technologien, Web Services und BPEL (Business Process Execution Language) auf breiter Ebene zum Einsatz [Fl10]. Fachliche Funktionen werden in Form von lose gekoppelten, dokumentierten und mehrfach nutzbaren Services bereitgestellt. Parallel zu diesem technikzentrierten Ansatz wird eine SOA aber auch als Managementkonzept verstanden. Aus Managementsicht wird mit einer SOA das Ziel einer an den Prozessen ausgerichteten Service-Landschaft verfolgt,

mit deren Hilfe agil und flexibel auf veränderte Anforderungen im geschäftlichen Umfeld reagiert werden soll. Während bei monolithischen Anwendungen (ERP-Systemen), die durch eine vereinigende Integration entstehen, die geschäftliche Funktionalität und die prozessuale Ablauflogik kaum getrennt voneinander betrachtet werden können, ist dies bei der verknüpfenden Integration auf SOA-Basis sehr wohl möglich. Damit können sowohl die Flexibilität als auch die Zukunftsfähigkeit einer Anwendung positiv beeinflusst werden. Die geschäftliche Funktionalität wird in elementaren Services abgebildet, die Implementierung der Prozess- oder Ablauflogik erfolgt durch eine Orchestrierung elementarer Services sowie durch eine variable Gestaltung der Wegwahl auf verschiedenen Ebenen (Prozesse, Module, Dokumente) und mit Hilfe von Variablen. In den folgenden Abschnitten wird ein Lösungsansatz zur flexiblen und modularen Gestaltung einer Integrationslösung im Rahmen des Product Lifecycle Managements (PLM) auf Basis einer SOA skizziert. Die Erhöhung der Prozessflexibilität wurde durch einen besonderen Gestaltungsansatz erreicht.

2 Variabilität durch Servicekontext

Aus technischer Sicht wird mit einer serviceorientierten Integrationsarchitektur das Ziel verfolgt, elementare fachliche Services zu schaffen, um durch eine Kombination dieser Services – je nach Bedarf – Prozesse flexibel gestalten zu können ([AG08], [La06]). Services sind damit die zentralen Elemente einer Integrationsarchitektur, die nach den Prinzipien einer SOA konzipiert wurde. Flexibilität gilt als einer der wichtigsten SOA-Architekturziele, die Bildung von Varianten gilt als eine anerkannte Möglichkeit zur Erhöhung der Flexibilität [AD08]. Ein elementarer Service implementiert in der Regel eine klar abgegrenzte geschäftliche Funktionalität und beinhaltet das gemeinsame Vielfache an „variablen Eigenschaften“ jener Produktlinien, für die er mehrfach verwendet werden soll. Im Rahmen einer SOA erfolgt die Zusammenstellung eines komplexen Dienstes, einer so genannten Composite Application (Abbildung 1), durch eine Orchestrierung elementarer Dienste. Im konkreten Fall wurde diese derart gestaltet, dass sie für verschiedene geschäftliche Projekte nutzbar wird. In ihrer Außenwirkung beruhen Services auf dem Konzept des Versteckens der eigenen Komplexität und geben ihrem Umfeld nicht preis, wie ihre Funktionalität implementiert wurde. In ihrem Inneren jedoch können sie mit einem Kontext versehen werden, der zur Laufzeit gezielt für eine Wiederverwendung genutzt werden kann. Ein nach außen hin statusloser Service gilt in der Regel als agnostisch im Sinne von „nicht wahrnehmend“ bzw. „ohne Wissen“ [Er08]. Innerhalb dieses Service jedoch kann gezielt eine nicht-agnostische Logik eingesetzt werden, um damit den Daten- und Nachrichtenfluss zur Laufzeit variabel steuern zu können. In der Integrationspraxis eines Fahrzeugherstellers ruft zum Beispiel ein Top-Level-Service, dieser erfüllt vornehmlich Steuerungsaufgaben, weitere Services auf, in denen fachliche Funktionalität und Verarbeitungslogik abgebildet wurde. Dabei verarbeitet die Composite Application Logik, Funktionalität und Daten aus verschiedenen Quellsystemen. Durch die Orchestrierung entstehen prozessuale Abläufe, die heterogene Systemgrenzen überwinden und flexible Anwendungen schaffen. Variablen und Parameter spielen dabei eine wichtige Rolle, um der Forderung nach Flexibilität und Agilität nachkommen zu können (vgl. [AT06]).

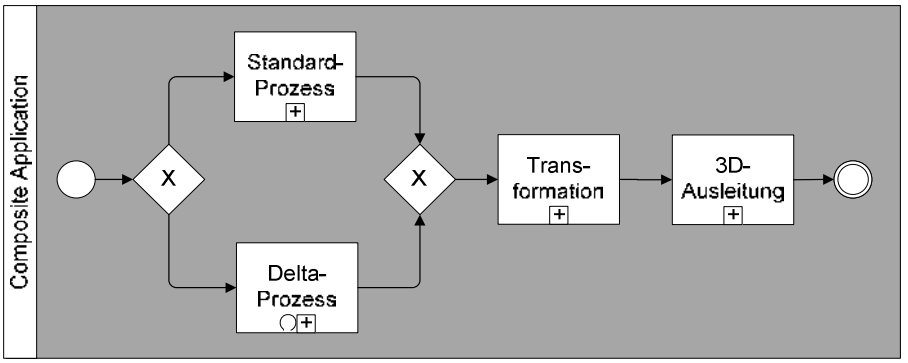


Abbildung 1: Schema einer PLM-Anwendung aus elementaren Services

Abbildung 1 zeigt die Orchestrierung einer PLM-Anwendung aus elementaren Diensten, die für verschiedene Produkte wiederverwendbar ist. Dies wird unter anderem durch eine generische Ablauflogik erreicht. Lediglich das variable Beziehungswissen zwischen den einzelnen Services, welches der Composite Application in Form von Parametern mitgegeben wird, ist für die jeweiligen Projekte unterschiedlich. Durch das über Prozessvariablen hergestellte Beziehungsgeflecht konnte eine flexible Prozesslogik für unterschiedliche fachliche Anforderungen etabliert werden. Die Prozess- bzw. Ablauflogik wird zur Laufzeit von Variablen beschrieben, die an jedem Knotenpunkt über das für die weitere Ausführung benötigte „Wissen“ verfügen. Bei Bedarf werden die variablen Anteile dieser Prozesslogik an die elementaren Services weitergereicht. Dadurch erhalten die nach außen hin statuslosen Services, in ihrem Inneren einen Kontext, mit dem zur Laufzeit der aktuelle Zustand in jedem Knotenpunkt beschrieben werden kann. Dieser Servicekontext wird für die Gestaltung der fachlichen Variabilität für mehrere Produktlinien benötigt. Der Service verhält sich zudem idempotent und ist daher mehrfach aufrufbar, ohne dass sich das Ergebnis der Verarbeitung ändert.

3 Anwendungsfall Digital Mock Up (DMU)

Unter dem Begriff „Digitale Fabrik“ versteht man das virtuelle Abbild einer geplanten oder realen Fabrik mit seinen Gebäuden, Einrichtungen, Anlagen und Fördersystemen sowie deren Wertschöpfungsprozessen in digitaler Form. Computer Aided Engineering (CAE), das die computerunterstützten Ingenieurleistungen Konstruktion und Berechnung umfasst, ist ein typischer Anwendungsbereich im Rahmen der Digitalen Fabrik. Für die Simulation von Engineering-Prozessen werden in Simulationsprogrammen zunächst Daten über die Produktionsabläufe und die zugehörige Logistik hinterlegt. Auf diese Weise können im Vorfeld der physischen Fahrzeugkonstruktion komplette Prozesse, wie zum Beispiel die automatische Geometrieprüfung, mittels Simulation untersucht und optimiert werden. Dabei kommen mathematische Modelle zur Darstellung der Prozesse und Abläufe ebenso zum Einsatz, wie Softwarewerkzeuge zur geometrischen Darstellung. Bei der DMU-Integration (Digital Mock Up), dem so genannten "digitalen Prototypenbau", geht es in erster Linie um die digitale Montage, um Untersuchungen einfacher Bauteilgruppen sowie um die Beherrschung komplexer Produktstrukturen

außerhalb des CAD-Systems (Computer Aided Design) [HHD08]. DMU ist ein Überbegriff für verschiedene Softwarebausteine, die in verschiedenen Phasen des Entwicklungsprozesses eingesetzt werden, um frühzeitig Kontrollen zur Beseitigung von Fehlern und zur Einsparung von Zeit und Kosten durchführen zu können. Diese Untersuchungen erfolgen auf der Basis von 3D-CAD-Modellen und umfassen die Absicherung der geometrischen Stimmigkeit aller Bauteile in verschiedenen Fahrzeugkonfigurationen sowie die Berechnung von Kollisions-, Kontakt- und Abstandsproblemen. Ein wesentliches wirtschaftliches Ziel dieser Simulation ist eine möglichst frühe Kenntnis allfälliger Fehlersituationen, um damit Zeit und Kosten sparen sowie die Qualität erhöhen zu können. Dabei werden mechanische Kräfte simuliert, unterschiedliche Varianten berechnet und Ergebnisse beurteilt [DH09]. Im Zuge der gegenständlichen SOA-Implementierung wurde darüber hinaus das methodische Ziel verfolgt, die Services für verschiedene Produktlinien wiederverwendbar zu gestalten. Dieses Ziel wurde durch den konsequenten Einsatz von Variablen und Parametern auf verschiedenen Abstraktionsebenen erreicht: *Prozessebene* (Service-Orchestrierung), *Modulebene* (fachliche Logik) und *Dokumentebene* (Austauschformat) [Fl09].

3.1 Prozessebene

Auf der Prozessebene wird das Beziehungswissen zwischen den einzelnen Services durch Prozessvariablen hergestellt. Der eingesetzte Enterprise Service Bus kann einzelne Werte, aber auch komplette XML-Dokumente, die zum Beispiel die Beschreibung von einzelnen Teilen, Komponenten oder Baugruppen beinhalten, in Variablen speichern. Deren Inhalte werden zur Laufzeit an die elementaren Services weitergereicht und damit für die Ablaufsteuerung nutzbar gemacht. Die Orchestrierung der Services zu einer Composite Application erfolgt über HTTP-Konnektoren, da keine dynamische Adressierung, wie sie mit BPEL-Controls und Web Service Schnittstellen möglich wäre, notwendig ist. Dabei können wahlweise globale Header-Variablen – zum Beispiel mit dem Programm `curl` – und/oder lokale Prozessvariablen verwendet werden. Die Interaktion zwischen den einzelnen Services erfolgt auf der Basis von Nachrichten. Diese rufen entfernte Funktionen auf und transportieren XML-Dokumente mit fixen Inhalten ebenso wie Ein- und Ausgangsparameter mit variablen Inhalten. Die Auswertung der vom Fachbereich vorgegebenen Konfigurationsdaten erfolgt auf der Basis von Business Rules [Fl10]. Auf diese Weise entsteht eine für verschiedene Produkte verwendbare generische Auflauflogik.

3.2 Modulebene

Die einzelnen Module eines Service, die zum Beispiel für die Transformation und die hierarchische Strukturierung von XML-Dokumenten genutzt werden, enthalten Funktionalität, die durch XSLT (eXtented Stylesheet Language) und XPath (XML Path Language) bereitgestellt wird. Lokale und globale Variablen bilden dabei wiederum das Beziehungswissen zwischen den einzelnen Templates innerhalb der Module ab. Dies geschieht in der Weise, dass die Eingabeparameter von XSLT-Modulen wie Parameter einer Klassenfunktion behandelt werden. Eine Variable stellt damit einen leeren „Container“ für fachliche Inhalte bereit, der je nach Projekt mit unterschiedlichen

Werten gefüllt wird. Für die Konvertierung einer flachen XML-Struktur, in ein in seiner Tiefe unbegrenztes hierarchisches XML-Dokument, wird das notwendige Beziehungswissen zwischen den einzelnen Baugruppen, Komponenten, Einzelteilen und deren Attributen während der Verarbeitung ebenfalls in Form von Parametern und Variablen bereitgestellt.

3.3 Dokumentebene

Ein Fahrzeug besteht auch in seiner digitalen Repräsentation aus verschiedenen Baugruppen, eine Baugruppe besteht aus mehreren Komponenten, die sich wiederum aus einzelnen Teilen zusammensetzen und durch das Geometriemodell dreidimensional (3D) beschrieben werden. Ein PDM-System (Produktdatenmanagement) bildet all diese Beziehungen und geometrischen Informationen ab und stellt sie in Form eines XML-Dokuments für den Datenaustausch bereit. Das Schema PLM-XML ist ein Industriestandard für den PLM-Datenaustausch [AMR10], das die Wiederverwendung gleicher Konstrukte für verschiedene Projekte unterstützt. Dabei verwenden zwei oder mehrere Partner ein und dasselbe Schema, um den Inhalt (Syntax) und die Bedeutung (Semantik) der Daten auf die gleiche Art und Weise interpretieren zu können. In einem „flachen" PLM-XML-Dokument, werden die Zusammenhänge zwischen den Attributen, z.B. `Product` (Sachnummer), `ProductView` (Struktur), `ProductRevision` (Änderungsstand), `RevisionRule`, `ReleaseStatus`, `DataSet` und `ExternalFile` (Pfadangaben) über variable, verbindende Attribute hergestellt. Die Verknüpfung dieser Informationen, wie sie für die Erstellung einer hierarchischen Baumstruktur benötigt werden, übernimmt der SOA-Entwickler nach fachlichen Vorgaben mit Hilfe von XSLT und XPath unter Nutzung der deskriptiven Eigenschaften der genannten XML-Technologien.

4 Zusammenfassung

In diesem Beitrag wurde ein SOA-basierter Lösungsansatz für eine PLM-Integration vorgestellt, bei dem eine wiederverwendbare, generische Ablauflogik im Vordergrund stand. Durch eine konsequente Berücksichtigung von variablem Beziehungswissen über verschiedene Abstraktionsebenen hinweg konnte eine Wiederverwendung für mehrere Produktlinien erreicht werden. Die Orchestrierung der elementaren Services zu einer Composite Application erfolgte mit Konnektoren, die das universelle Protokoll HTTP (Hypertext Transfer Protocol) nutzen. Im Zusammenwirken mit dem verwendeten Enterprise Service Bus (ESB) eröffnete sich dabei die Möglichkeit neben den lokalen Variablen innerhalb der Services auch die globalen Header-Variablen der HTTP-Aufrufe über die gesamte Composite Application hinweg zu nutzen. Weitere Eckpunkte dieser Integrationslösung sind die Nutzung anerkannter Methoden wie Enterprise Integration Patterns (z.B. Canonical Data Format und Event-driven Consumer) [HW04] sowie von domänenspezifischen Austauschformaten wie PLM-XML. Bedingt durch die hohe Anzahl an Bauteilen, die in den XML-Dokumenten hierarchisch abgebildet werden mussten, war bei der Implementierung zudem auf ein günstiges Laufzeitverhalten Bedacht zu nehmen.

Literaturverzeichnis

[Ad08] Adam, S.; Doerr, J.: How to better align BPM & SOA – Ideas on improving the transition between process design and deployment. In Proceedings of BPMDS 2008, http://lams.epfl.ch/conference/bpmds08/program/paper6.pdf (abgerufen am 29.06.2010)

[AMR10] Anderl, R.; Mosch, C.; Rollmann, T.: Neutraler Datenaustausch in der Digitalen Fabrik. In ZWF 105 (2010) 3, Carl Hanser Verlag, München 2010, S. 163-167.

[AG08] Aier, S; Gleichauf, B.: Begründung eines differenzierten Serviceverständnisses und Richtlinien für die Servicekonstruktion. In Proceedings 3. Workshop BSOA 2008, Shaker Verlag, Leinfelden b. Stuttgart 2008, S. 47-58.

[AT06] Ackermann, J; Turowski, K.: Zur Rolle von Parametrisierung bei der fachlichen Anpassung betrieblicher Softwarekomponenten. In Schelp, J.; Winter, R.; Frank, U.; Rieger, B.; Turowski, K. (Hrsg.): Integration, Informationslogistik und Architektur. Proceedings DW2006, Gesellschaft für Informatik, Bonn 2006, S.341-359.

[AW09] Aier, S.; Winter, R.: Virtuelle Entkopplung von fachlichen und IT-Strukturen für das IT/Business-Alignment: Grundlagen, Architekturgestaltung und Umsetzung am Beispiel der Domänenbildung. In Wirtschaftsinformatik 51 (2009) 2, S. 175-191.

[BR07] Brabänder, E.; Erbach, F.: "AVE for SAP" – Eine Vorgangsweise für das Geschäftsprozessmanagement mit SAP. In Informatik Spektrum 30 (6) 2007, S. 419-427.

[DH09] Dietrich, W.; Heilig R.: Effiziente Kollisionsprüfung in der Fahrzeugtechnik. In Digital Engineering Magazin 1/2009, S. 72-74.

[Er08] Erl, T.: Introducing SOA Design Patterns. SOA World Magazin, Vol. 9, 6/2008.

[Fl09] Flieder, K.: Serviceorientierte Softwarearchitekturen in Theorie und Praxis. In e&i Elektrotechnik & Informationstechnik 126 (12) 2009, Springer-Verlag, Berlin, Heidelberg, New York 2009, S. a32-a35.

[Fl10] Flieder, K.: Mit RFID und BPM zum ereignisgesteuerten Unternehmen. In ZWF 105 (2010) 5, Carl Hanser Verlag, München 2010, S. 494-502.

[FNS06] Fischer, D.; Nirsberger, I.; Stelzer, D.: Ein Modell zur Bestimmung des Grades der unternehmensübergreifenden Integration von Informationssystemen. In Schelp, J.; Winter, R.; Frank, U.; Rieger, B.; Turowski, K. (Hrsg.): Integration, Informationslogistik und Architektur. Proceedings DW2006, Gesellschaft für Informatik, Bonn 2006, S. 427-447.

[He07] Heutschi, R.: Serviceorientierte Architektur: Architekturprinzipien und Umsetzung in der Praxis. Springer-Verlag, Berlin, Heidelberg, New York 2007.

[HHD08] Hirz, M.; Hirschberg, W.; Dietrich, W.: Integrated 3D-CAD Design Methods in Early Automotive Development Processes. In Proceedings FISITA 2008 – The Future of Automobiles and Mobility, S. 1-15.

[HR+10] Holschke, O; Rake, J; Offermann, P.; Bub, U.: Steigerung der Softwareflexibilität bei Geschäftsprozessänderungen. In Wirtschaftsinformatik 52 (2010) 1, S. 3-15.

[HW04] Hohpe, G.; Woolf, B.: Enterprise Integration Patterns: Designing, Building, and Deploying Messaging Solutions. Springer-Verlag, Berlin, Heidelberg, New York 2007.

[La06] Laures, G.: Flexibilitätsanalyse service-orientierter Architekturen zur Realisierung von Geschäftsprozessen. In Proceedings EMISA 2006, Springer-Verlag, Berlin, Heidelberg, New York 2006, S. 163-177.

[SA08] Stutz, M.; Aier, S.: Vorgehensmodell zur fachlichen Bewertung serviceorientierter Architekturen. In Bichler, M.; Hess, T.; Krcmar, H.; Lechner, U.; Matthes, F.; Picot, A.; Speitkamp, B.; Wolf, P. (Hrsg.): Proceedings Multikonferenz Wirtschaftsinformatik 2008, GITO-Verlag, Berlin 2008, S. 989-999.

Gestaltung wandlungsfähiger ECM-Referenzprozesse unter Verwendung kartographischer Methoden

Sandy Eggert

Lehrstuhl für Wirtschaftsinformatik und Electronic Government
Universität Potsdam
August-Bebel-Str. 89
14482 Potsdam
sandy.eggert@wi.uni-potsdam.de

Abstract: Enterprise Content Management (ECM) bildet in Unternehmen zunehmend die Basis für die IT-seitige Verwaltung schwach strukturierter Informationen, insbesondere in Form von Dokumenten und Inhalten ohne Dokumentenbezug. Neben dem Einsatz von IT-Systemen, die im Rahmen des Enterprise Content Managements Anwendung finden, liegt eine besondere Herausforderung in der Gestaltung der Prozesse des Enterprise Content Managements, in denen die Informationsobjekte und Dokumente entlang ihres Lebenszyklus verarbeitet werden. Turbulenzen, die bspw. durch Wirtschaftskrisen oder auch interkulturelle Kooperationsvorhaben auf Unternehmen einwirken, erfordern ein hohes Maß an Anpassung und Reaktionsgeschwindigkeit. Wandlungsfähigkeit ist ein Ansatz, diesen Turbulenzen adäquat zu begegnen. Im Rahmen des Beitrags wird Wandlungsfähigkeit als eine über die Flexibilität hinausgehende Fähigkeit zur Anpassung von Systemen und Prozessen verstanden. Als Eigenschaft von Prozessen bezieht sich Wandlungsfähigkeit auf das Erkennen von Änderungsbedarfen und die Bereitstellung von Möglichkeiten zur veränderten Prozessgestaltung im Sinne der Wandlungsfähigkeit. Der Beitrag stellt einen methodischen Ansatz vor, der unter Anwendung von Indikatoren zur Erhöhung der Wandlungsfähigkeit wandlungsfähige Referenzprozesse entwickelt. Diese unter Verwendung der Prozesskartographie entwickelten Referenzprozesse sollen sich durch die Eigenschaft der Wandlungsfähigkeit auszeichnen und werden im Folgenden als Kartenmuster bezeichnet.

1 Entwicklung von Kartenmustern

Im Rahmen einer Forschungsarbeit wurde die Methode zur Erarbeitung wandlungsfähiger Kartenmuster entwickelt. Die Vorgehensweise ist in der Abbildung 1 zusammenfassend dargestellt. Die Phasen „Bildung der Fragmente“ und „Ableitung der Bewertungsindikatoren“ stellen die Grundlage zur anschließenden Ableitung von Gestaltungshinweisen für wandlungsfähige ECM-Prozesse dar.

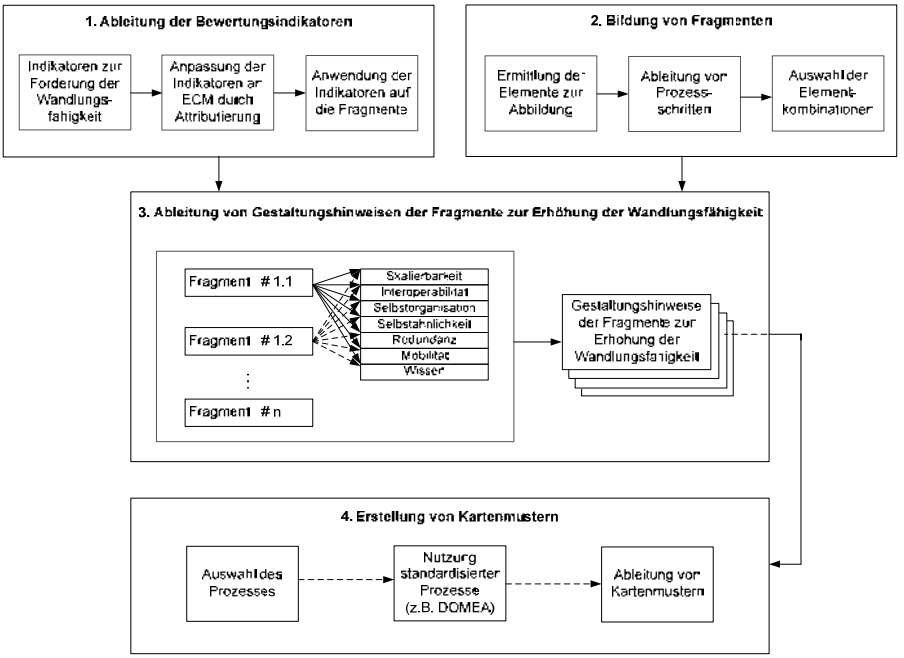


Abbildung 1: Methode zur Entwicklung wandlungsfähiger Kartenmuster

1.1 Bewertungsindikatoren für die Wandlungsfähigkeit von ECM-Prozessen

Die Methode zur Entwicklung wandlungsfähiger Kartenmuster verwendet die Indikatoren zur Förderung der Wandlungsfähigkeit von Anwendungssystemen [Gr06]. Es besteht die Annahme, dass diese Indikatoren vor dem Hintergrund einer systemischen Betrachtungsweise [ALG08] auf ECM-Prozesse angewandt werden können. Als Systembestandteile werden Prozessschritte, organisatorische Einheiten sowie Anordnungsbeziehungen der Phasen und Elemente definiert. Um die Übertragung der Indikatoren auf ECM-Prozesse durchführen zu können, werden zuvor wandlungsfähige ECM-Prozesse definiert. ECM fokussiert das Management von Informationen innerhalb der Unternehmensprozesse. Die Unternehmensprozesse, die durch ein erhöhtes Dokumentenaufkommen gekennzeichnet sind, werden demnach als ECM-Prozesse bezeichnet. Ein Prozess allgemein wird basierend auf den Ergebnissen des im Forschungsprojektes IOSEW² [BGW09] vorgestellten Betrachtungsansatz als soziotechnisches System angesehen. Dieses System kann entlang dieses Ansatzes genau dann als wandlungsfähig angesehen werden, wenn Änderungen des Prozessablaufs möglich sind und das System einen kontinuierlichen Prozess der Verbesserung darstellt. Ein ECM-Prozess kann basierend auf diesem Verständnis als wandlungsfähig bezeichnet werden, wenn die Kriterien der Wandlungsfähigkeit als Gestaltungsprinzipien angewandt wurden und geeignete Parameter zur Überprüfung der Performance definiert sind und kontrolliert werden können. In Tabelle 1 wird die Übertragung der Indikatoren auf ECM-Prozesse mit dem Gestaltungsziel der Wandlungsfähigkeit am Beispiel des Indikators „Interoperabilität“ verdeutlicht.

Tabelle 1: Attributierung der Indikatoren (Auszug)

Indikator	**Attributierung**
Interoperabilität	Informations- und Dokumentenaustausch, Prozessschnittstellen, Modul- und Systemschnittstellen, Transfer von Prozesselementen
Modularität	Aufbau der Systeme, Aufbau und Abgrenzung der Prozessschritte, Austauschbarkeit
Skalierbarkeit	Quantitative Anpassbarkeit, Prozesssteuerung
Mobilität	Unabhängigkeit, örtliche und zeitliche Verfügbarkeit
Redundanz	Effizientes Ressourcenmangt., Vermeidung unerwünschter Redundanzen
Selbstähnlichkeit	Aufbau von ECM-Modulen, Gestaltung der Prozessschritte
Selbstorganisation	Prozessbasierte Kontrollparameter, Diagnosemöglichkeiten, Gestaltungsoptionen (Ordnungsparameter)
Wissen	Prozess- und Systemwissen, Anforderungsprofile, Informationsbereitstellung

Die Anpassung der Indikatoren wird mit Hilfe einer ECM-spezifischen Attributierung zum Zweck einer Operationalisierung zur Ableitung von Gestaltungsempfehlungen für ECM-Prozesse durchgeführt.

1.2 Bildung von Fragmenten mit Gestaltungshinweisen

Um als Referenzprozesse dienende Kartenmuster für ECM-Prozesse unter dem Aspekt der Wandlungsfähigkeit zu entwickeln, werden zuvor einzelne Prozessbausteine zur Betrachtungsgrundlage entwickelt. Ziel ist es, alle möglichen und relevanten Prozessbausteine in den Kartenmustern abzudecken, um schließlich auf Basis der Indikatoren Empfehlungen zur Verbesserung der Wandlungsfähigkeit für einen ECM-Gesamtprozess ableiten zu können. Die zu ermittelnden Prozessbausteine werden im Folgenden als Fragmente bezeichnet, da sie einzelne, spezifische Teile eines ECM-Prozesses verkörpern. Zur Bildung von Fragmenten werden die Lebenszyklusphasen von Dokumenten und Content herangezogen. Es gilt hierbei die Annahme, dass mit Hilfe der Lebenszyklusphasen alle Prozessschritte in allen möglichen ECM-Prozessen abgebildet werden können [Eg10]. Da die aktuell im Bereich des ECM eingesetzten Systeme in ihrer Mehrheit eher Dokumenten- als Content-orientiert sind [EG08] und meist der Dokumentenverwaltung dienen [MKG08], liegt der Schwerpunkt der Betrachtung zunächst auf den Dokumentenlebenszyklusphasen. Der typische Lebenszyklus eines Dokuments reicht von der Erstellung über die Nutzung und Archivierung bis hin zur Vernichtung [KSR08]. Jeder Lebenszyklusphase werden Prozesse zugeordnet, die Aktivitäten dieser Phase wiedergeben. Die dafür genutzten Prozesse sind Prozesse des ECM, die bei mehreren Praxispartnern erhoben wurden. Das mit diesem Vorgehen verbundene Ziel ist die Ermittlung der Prozessschritte, die im Fragment abgebildet werden. Nach der Filterung der für die Lebenszyklusphase relevanten Prozessschritte können diese Prozessschritte zur Abbildung in den Fragmenten aufgenommen werden. Die weiteren Visualisierungselemente (z.B. System, Verbindungen) zur Darstellung in den Fragmenten können nun anhand der zugeordneten Prozesse ergänzt werden. Entlang dieser Vorgehensweise wurden insgesamt sieben Fragmente erstellt: Digitalisierung, Erfassung, Bearbeitung, Veröffentlichung, Wiederverwendung, Archivierung und Löschung.

Zu jedem Fragment wurden in einem weiteren Schritt Empfehlungen für eine wandlungsfähigere Gestaltung entlang der Indikatoren generiert.

1.3 Erstellung von Kartenmustern

Basierend auf den erarbeiteten Fragmenten wird nun die Entwicklung von Kartenmustern aufgezeigt, die im Sinne einer Referenzmodellbildung typische ECM-Prozessbausteine unter dem Aspekt der Wiederverwendung darstellen. Referenzmodelle stellen grundsätzlich Artefakte mit Empfehlungscharakter dar, die Modellierer bei der Konstruktion individueller Prozesse anleiten sollen [BD08]. Im Bereich des ECM besteht hinsichtlich der Referenzmodellierung enormer Handlungsbedarf. Derzeit existieren nur Hinweise zur Prozessgestaltung im DOMEA®-Organisationskonzept (teilweise auch im MoReq2-Standard) sowie in einigen Spezifikationen konkreter Anwendungssysteme. Die grundlegende Idee der Kartenmuster ist die Beschreibung eines ECM-Prozessschrittes in wandlungsfähiger Form. Die Gestaltung eines Prozessschrittes basiert dabei auf einem standardisierten Prozess, welcher einen typischen ECM-Prozess abbildet. Die Beschreibung eines Kartenmusters besteht aus einem initialen Kontext, der einen allgemeinen, wiederkehrenden Ablauf innerhalb der Dokumentenbearbeitung beschreibt, einem Vorschlag einer wandlungsfähigeren Gestaltung sowie der visualisierten Abbildung in Form einer Prozesskarte. Die entwickelten Kartenmuster sollen als Referenzprozesse im Sinne einer wandlungsfähigen Gestaltung von ECM-Prozessen eingesetzt werden können. Ein sinnvoller Einsatzbereich stellt hierbei die Potenzialanalyse im Rahmen der Aufdeckung von Prozessverbesserungen dar. Die Ableitung von Kartenmustern, die sich durch die Eigenschaft der Wandlungsfähigkeit auszeichnen, erfolgt entlang der Gestaltungsempfehlungen der erarbeiteten Fragmente. Ein Kartenmuster beschreibt hierbei einen bestimmten Prozessbaustein, der häufig innerhalb von ECM-Landschaften auftritt. Kartenmuster stellen neben der eigentlichen Prozessmodellierung Möglichkeiten für eine wandlungsfähige Gestaltung dar. Die Erstellung der Kartenmuster beruht auf einer 3-stufigen Vorgehensweise:

1. Auswahl des Prozesses

Die Entwicklung eines Kartenmusters beginnt mit der Auswahl des Prozesses, der als Kartenmuster im Sinne der Referenzmodellierung abgebildet werden soll. Hierzu werden die innerhalb einer Umfrage erhobenen typischen ECM-Prozesse [EG05] genutzt. Nach der Auswahl des ECM-Prozesses wird dieser mit Hilfe standardisierter Prozessvorgaben abgebildet. Der Prozess, dessen Ableitung im Folgenden aufgezeigt wird, ist die „Übergabe von Archivgut“.

2. Prozessdarstellung

Die Prozessdarstellung, bestehend aus einer an der standardisierten DOMEA-Spezifikation zur Übergabe von Archivgut orientierenden Beschreibung [Dom05] und einer Darstellung in Form einer Prozesskarte, dient als Basis für die Anwendung der Fragmente.

3. Entwicklung des Kartenmusters

Zur Ableitung des Kartenmusters „Übergabe von Archivgut“, welche wandlungsfähige Gestaltungsaspekte berücksichtigt, erfolgt zunächst eine Zuordnung von Fragmenten. Der Prozesskarte „Übergabe von Archivgut“ wird das „Fragment 5.1: Archivierung“ zugeordnet. Entlang des Fragments werden mögliche Potenziale aufgedeckt, die mit Hilfe der Gestaltungsempfehlungen konkretisiert werden. Die Umsetzung dieser Gestaltungsempfehlungen mündet in die Ableitung eines Kartenmusters für die Übergabe von elektronischem und papierbasiertem Archivgut“.

2 Anwendung innerhalb der Prozessanalyse

Die auf Basis der Fragmente entwickelten Kartenmuster wurden innerhalb mehrerer Fallstudien evaluiert. Dazu wurden Unternehmen ausgewählt, die sich in einem turbulenten Umfeld befinden und einen Bedarf an Prozessverbesserung im Bereich ihrer ECM-Prozesse identifiziert haben.

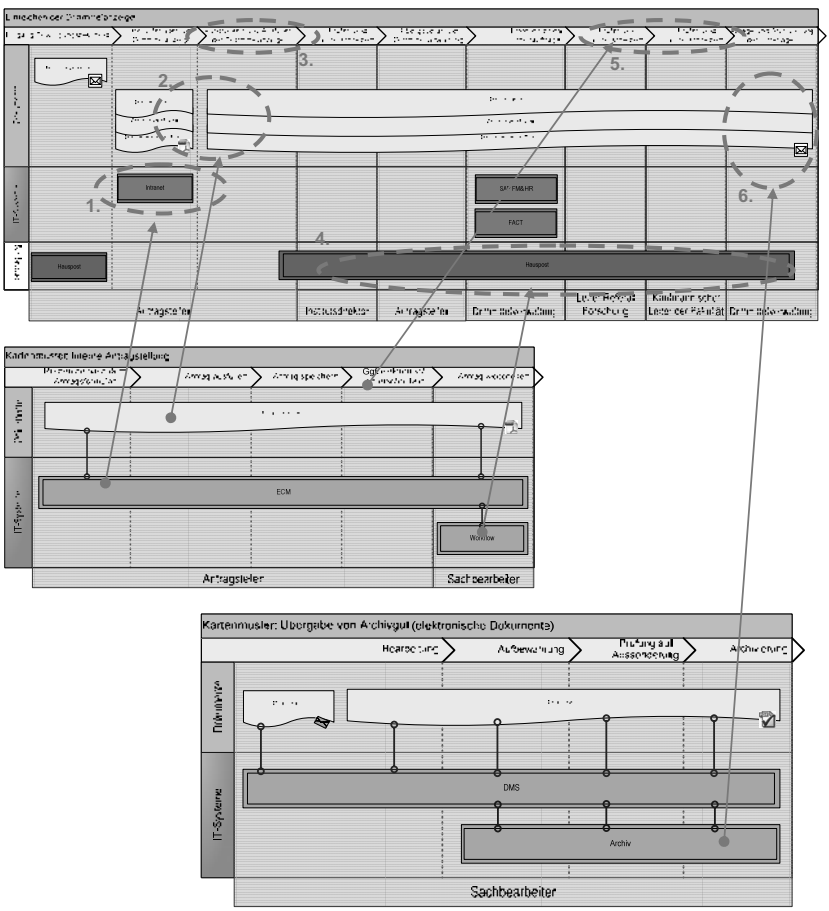


Abbildung 2: Potenzialanalyse – Prozess: Einrichten der Drittmittelanzeige

Zunächst wurden die IST-Prozesse aufgenommen und visualisiert. Beispielhaft wird die Ableitung der Potenziale anhand des Beispielprozesses „Einreichen der Drittmittelanzeige“ aufgezeigt. Für diesen Prozess werden die Kartenmuster „Interne Antragstellung“ und „Übergabe von Archivgut“ verwendet. Abbildung 2 zeigt die Analyse der Potenziale. Anhand der Potenziale konnten SOLL-Prozesse abgeleitet werden. Ziel der Sollmodellierung ist es, alternative Szenarien für wandlungsfähigere Prozesse aufzuzeigen, die die identifizierten Potenziale beinhalten.

3 Zusammenfassung

Innerhalb des Beitrags wurde eine Methode aufgezeigt, wie unter Nutzung der Lebenszyklusphasen von Dokumenten Fragmente entwickelt werden können. Fragmente stellen kleine Prozessbausteine von ECM-Prozessen in Form einer kartographischen Visualisierung dar und enthalten Gestaltungsempfehlungen entlang der ermittelten Indikatoren zur Förderung der Wandlungsfähigkeit. Diese Indikatoren wurden zudem durch eine ECM-spezifische Attributierung gekennzeichnet. Um die erarbeiteten Ergebnisse leichter in die Praxis überführen zu können, wurde basierend auf den erarbeiteten Fragmenten ein Verfahren zur Entwicklung von Kartenmustern aufgezeigt. Durch dieses Verfahren wurden im Sinne einer Referenzprozessbildung typische ECM-Prozessbausteine unter dem Aspekt der Wiederverwendung erstellt. Diese Referenzprozesse werden als Kartenmuster bezeichnet und beschreiben einen bestimmten Prozessbaustein, der häufig innerhalb von ECM-Landschaften auftritt. Sie stellen neben der eigentlichen Prozessmodellierung Möglichkeiten für eine wandlungsfähige Gestaltung dar. Der Einsatz der Kartenmuster fand im Rahmen der Potenzialanalyse ihrer in den Betrachtungsbereich gezogenen ECM-Prozesse statt. Hierbei wurden SOLL-Prozesse unternehmensspezifisch unter dem Aspekt einer wandlungsfähigeren Gestaltung generiert und zur Umsetzung vorgeschlagen.

Literatur

[ALG08] Andresen K, Levina O, Gronau N (2008) Design of the evolutionary process model for adaptable software development processes. In: Proceedings of the EMCIS.

[BGW09] Broy, M., Gronau, N., Wildemann, H. : Gestaltung interorganisationaler Software-Entwicklung - Herausforderungen durch Wandlungsfähigkeit und Wiederverwendung, GITO-Verlag, Berlin 2009.

[BD08] Becker J; Delfmann P (2008) Referenzmodellierung 2008. In: Proceedings of the MKWI. M. Bichler et al.: 339-340.

[EG05] Eggert S, Gronau N (2005) Auswertung einer Marktstudie zu Enterprise Content Management Systemen. In: ERP Management, 4: 48-59.

[Gr06] Gronau, N.: Wandlungsfähige Informationssystemarchitekturen - Nachhaltig-keit bei organisatorischem Wandel. 2. Aufl., GITO-Verlag, Berlin 2006.

[KSR08] Kaiser M G, Smolnik S, Riempp G (2008) Konzeption eines Information-Lifecycle-Management-Frameworks im Dokumenten-Management-Kontext. In: Proceedings of the MKWI. M. Bichler et al.: 483-494.

[MKG08] Mertins K, Kohl H, Görmer M (2008) Benchmarking-Studie Dokumentenlogistik. Fraunhofer IPK Eigenverlag, Berlin.

Workshop Informationssysteme mit Open Source

Prof. Dr. Frank Rump, Prof. Dr. Bernd Müller

Software-Entwicklung mit Open Source Werkzeugen – Der GeneSEZ-Ansatz

Tobias Haubold, Wolfgang Golubski, Oliver Arnold, Georg Beier, Gerrit Beine

Westsächsische Hochschule Zwickau
Fakultät Physikalische Technik / Informatik
Dr.-Friedrichs-Ring 2a
08056 Zwickau
toh@fh-zwickau.de
golubski@ fh-zwickau.de

Abstract: Bei der Softwareentwicklung werden häufig eine Vielzahl von Werkzeugen z.B. zum Anforderungsmanagement und zur Modellierung oder auch Bug-Tracker sowie Build-Server eingesetzt. Zur effektiven Softwareentwicklung ist ein gutes Anforderungsmanagement unverzichtbar. Bei agilen Vorgehensweisen werden häufig Modelle zur Architekturdokumentation verwendet, die auch als Ausgangspunkt für die teilautomatisierte Softwareentwicklung dienen können. Dieses Paper stellt den GeneSEZ-Ansatz vor, mit dem durch entwickelte Open Source Werkzeuge Anforderungsbeschreibungen erfasst und Modelle teil-automatisiert in Quellcode transformiert werden können.

1 Motivation

Die Kosten der Softwareentwicklung machen heute immer noch einen großen Teil der Gesamtkosten der Entwicklung von Softwareprodukten aus. Seit mehr als drei Jahrzehnten werden enorme Anstrengungen in Forschung und Entwicklung unternommen, um diesen Kostenfaktor weiter zu reduzieren.

Ziel des hier vorgestellten GeneSEZ-Projektes [GEN] (Generative Software Engineering Zwickau) ist, die Softwareentwicklung als Ganzes effizient zu gestalten. Dabei wurde eine agile Vorgehensweise zur Anforderungsermittlung mit geeigneter Toolunterstützung sowie ein Framework zur Codegenerierung aus UML-Modellen entwickelt. Aus den Modellen wird teilweise oder vollständig ausführbarer Programmcode (z.B. Java, C#, PHP etc.) generiert und so die Entwicklung komplexer Software vereinfacht.

Die im Projekt entstandenen Werkzeuge und Methoden, die allesamt als Open-Source-Werkzeuge zur Verfügung stehen, ermöglichen eine schnelle, qualitativ hochwertige und wirtschaftliche Entwicklung von Systemen für verschiedene Anwendungsbereiche.

Eine Herausforderung bei der Anforderungsermittlung ist das Verwalten der Anforderungen. Bei komplexeren Fragestellungen wird eine Vielzahl von Dokumenten

erstellt, die meistens auch noch miteinander verknüpft sein müssen. Daher ist eine elektronische Unterstützung unabdingbar. Dies ist durch ein geeignetes Wiki-System [ABG08], das auf der Basis des bekannten Volere-Prozesses [RR06] und der Volere-Templates entworfen wurde, realisiert.

Die Object Management Group (OMG) [OMG] propagiert eine Vorgehensweise zur rationellen, teilautomatisierten Softwareentwicklung auf der Basis formaler UML-Spezifikationen, die sogenannte Model-Driven Architecture (MDA) [MDA]. Im Gegensatz zu dem Ziel der OMG, Interoperabilität zwischen Softwareentwicklungswerkzeugen und Standardisierung von Modellen für Anwendungsbereiche zu erreichen, zielt der hier dargestellte Ansatz (MDSD – Model-Driven Software Development) auf die Bereitstellung von praktisch einsetzbaren Komponenten und Werkzeugen für den Softwareentwicklungsprozess.

Kapitel 2 betrachtet die Analyse und das Management von Anforderungsbeschreibungen, während Kapitel 3 auf die teilautomatisierte Softwareentwicklung mit Modellen eingeht. Anschließend werden verwandte Arbeiten sowie die bisherigen Erfahrungen betrachtet.

2 Anforderungsermittlung und -Management

Das korrekte Ermitteln der Anforderungen eines zu entwickelnden Produktes stellt heute immer noch eine sehr anspruchsvolle und kreative Tätigkeit dar. Basierend auf langjährigen Erfahrungen in der Praxis haben sich verschiedene Herangehensweisen entwickelt, die situationsbedingt erfolgreich eingesetzt werden können. Unabhängig von der Art der Erfassung sowie der Struktur von Anforderungen ist eine elektronische Unterstützung durch ein geeignetes Werkzeug unerlässlich. Dies kann mit Hilfe eines Wiki-System realisiert werden.

Wiki-Systeme als einfach zu bedienende und oftmals frei verfügbare Werkzeuge für die kollaborative und (oftmals räumlich) verteilte Arbeit an der evolutionären Erarbeitung von Wissen erfreuen sich immer größerer Beliebtheit. Dies schließt immer häufiger auch den professionellen Einsatz durch Unternehmen ein. Weit verbreitet ist auch die Nutzung von Wiki-Systemen im Rahmen von Softwareentwicklungsprojekten, allerdings eher in der Realisierungsphase. Auf der Basis des Volere-Prozesses und der Volere-Templates [RR06] ist ein Twiki entstanden. In der Twiki-Lösung [ABG] wird für jede Anforderungsspezifikation ein eigenes Web (eine Verwaltungseinheit in einer TWiki-Installation bestehend aus einer Reihe verlinkter Webseiten) angelegt. Diese basieren jeweils auf einem so genannten Template Web, das die grundlegenden Strukturen enthält, die für eine auf dem Volere-Template basierende Anforderungsspezifikation notwendig sind.

Wichtig für die Anwendbarkeit von TWiki für die Anforderungsermittlung ist die Möglichkeit, strukturierte Daten miteinander zu verknüpfen. Dies geschieht, indem die Inhalte einer TWiki Form dynamisch über Abfragen an das TWiki-System generiert werden.

Basierend auf den Ergebnissen der Anforderungsermittlung wird die Realisierung des Software-Produktes durchgeführt.

3 Das GeneSEZ-Generator-Framework

In der MDA/MDSD-Vorgehensweise zur Softwareentwicklung werden erstellte Anwendungsmodelle automatisch durch Modelltransformationen in Programmcode überführt. Bei den Modellierungssprachen für MDSD-Plattformen sind UML [UML09] sowie Domain Specific Languages (DSLs) weit verbreitet. Die UML stellt objektorientierte Modellierungskonzepte zur Verfügung und bietet eine gute Abstraktionsebene zur Programmcodegenerierung. Leider entwickelte sich die UML mit einer steigenden Anzahl von Modellierungskonzepten zu einer aus folgenden Gründen für Modelltransformationen eher ungeeigneten Modellierungssprache:

- sie wurde sehr komplex [Th03,WM07a,WM07b]
- sie enthält für die Programmcodegenerierung uninteressante Informationen
- die UML Spezifikation enthält keinen Überblick des UML-Metamodells, aus dem hervorgeht, wie auf bestimmte Informationen zugegriffen werden kann
- sie enthält Informationen über die grafische Darstellung von Modellelementen, z.B. den Bildern von Stereotypen
- u.a. [MB02,Ba07] zeigen auf, dass nur ein Teil der UML sinnvoll für MDA/MDSD-Plattformen nutzbar ist.

Neben UML gibt es DSLs, die Konzepte eines Anwendungsbereiches oder technische Aspekte zur Beschreibung von Modellen nutzen. Dabei sind vor allem die fachlichen DSLs interessant, da die Modelle mit den Experten der Anwendungsbereiche diskutiert werden können. Jedoch vergrößern diese DSLs den Unterschied zwischen Modellierungskonzepten und den Konzepten der Programmiersprachen.

Die Idee bei GeneSEZ ist, nicht das Metamodell der Anwendungsmodelle für die MDSD-Plattform zu nutzen, sondern ein knappes und für die objektorientierte Quellcodegenerierung geeignetes Metamodell zu erstellen. Die MDSD-Plattform wird somit zu einer stabilen, wiederverwendbaren und investitionssicheren Basis für Modelltransformationen zur Quellcodegenerierung, die unabhängig von Modellierungswerkzeugen ist. Abbildung 2 zeigt die MDSD-Plattform [Ha09] schematisch mit den folgenden charakteristischen Konzepten:

- das GeneSEZ-Metamodell [HBG09] als zentraler Bestandteil stellt Informationen in aufbereiteter Form für Modelltransformationen zur Verfügung
- Modelladapter zur Anbindung von Modellierungswerkzeugen, indem Anwendungsmodelle in GeneSEZ-Modelle transformiert werden
- die GeneSEZ-Komponenten als gemeinsame Basis für die Erzeugung von Quellcode für verschiedene Programmiersprachen
- Plattformprojekte mit Skripten und Templates für die jeweiligen Programmiersprachen sowie Bibliotheken und Frameworks zur Quellcodegenerierung

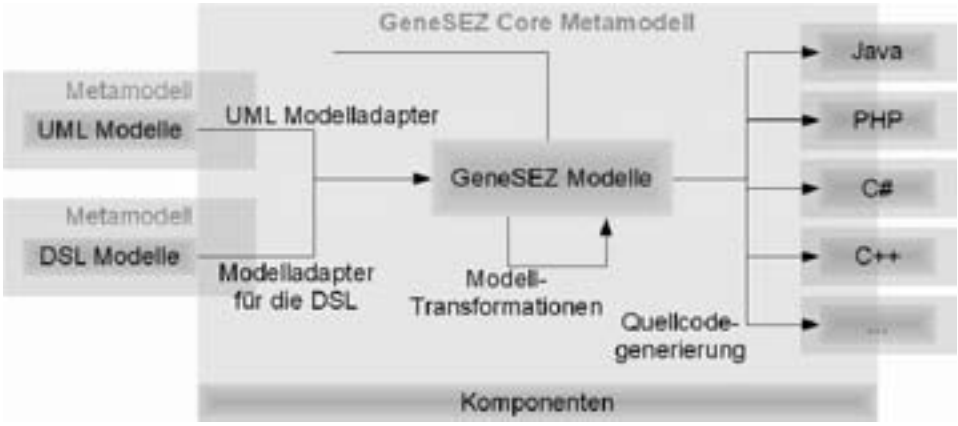


Abbildung 2: Schematischer Überblick des GeneSEZ MDSD Ansatzes

Anwendungsmodelle sind unabhängig von der MDSD-Plattform und können separat entwickelt werden. In iterativen Zyklen können die Modelltransformationen ausgeführt und Quellcode generiert werden. Da meistens der Quellcode nicht sinnvoll vollständig generiert werden kann, gibt es zur Vervollständigung der Implementierung geschützte Bereiche Der manuelle Quellcode bleibt beim erneuten Generieren erhalten. Durch das iterative Vorgehen sind Informationsrückflüsse aus der Implementierung auf das Anwendungsmodell möglich.

Für die Entwicklung von Modelltransformationen ist es hilfreich, dass das GeneSEZ-Metamodell vollständig auf zwei Seiten als Navigationsübersicht ausdruckbar ist.

4 Verwandte Arbeiten

Stanislaw Wrycza et al [WM07a,WM07b] beschreiben die Komplexität der UML2 aus der Perspektive der Lehre. Sie konzentrieren sich auf die Definition einer reduzierten und beschränkten Menge an Modellierungskonzepten und Diagrammen, die sie Light UML nennen. Während Stanislaw Wrycza et al sich auf das Wissen zur Erstellung robuster UML-Modelle konzentrieren, vereinfacht der hier vorgestellte Ansatz die Auswertung von UML-Modellen zur Quellcodegenerierung.

AndroMDA [AND] ist ein MDA-Open-Source-Framework mit Unterstützung für weit verbreitete Architekturen, u.a. Spring, EJB, Hibernate, Struts und .NET. Es wird ein proprietäres, in Java implementiertes UML-Metamodell sowie Velocity als Template-Sprache zur Codegenerierung eingesetzt, welche keine AOP-Fähigkeiten bietet.

Im Gegensatz zu AndroMDA unterstützt das Open-Source-Framework openArchitectureWare (oAW) [OAW] neben UML auch Java- oder EMF-basierte Metamodelle. Für Modelltransformationen stehen die Sprachen Xtend (Modell-zu-Modell) und Xpand (Modell-zu-Text) bereit, die beide Unterstützung für die AOP bieten. Unterstützung zur Quellcodegenerierung bietet oAW nicht. Die beiden Open-Source-

Projekte fornax-Plattform [FOR] und Sculptor [SCU] stellen Modelltransformationen auf Basis des UML-Metamodells bzw. einer DSL für Webanwendungen bereit. Der hier vorgestellte Ansatz basiert auf oAW, da dessen Sprachen für Modelltransformationen gut geeignet sind.

5 Praktische Anwendung und Ergebnisse

Der GeneSEZ-Ansatz wurde im Rahmen eines BMBF-Forschungsprojektes mit zwei kleinen- und mittelständischen Unternehmen entwickelt und in deren realen Projekten eingesetzt. Darüber hinaus wurde er in mehreren Forschungs- und Studentenprojekten sowie in der Lehre weiterentwickelt und angewandt. Die Einsatzgebiete umfassen:

- Entwicklung von Java- sowie EJB3 und Seam-basierter Anwendungen
- Eingebettete Systeme basierend auf Java (Java ME)
- Reengineering und Portierung einer C++ -basierten Software auf C#
- Entwicklung PHP-basierter Webanwendungen

Der Anteil des generierten Quellcodes lag zwischen 50% und 80% Dieser Wert ist abhängig von der Art der Software und der vorhandenen Unterstützung des Generators. Bei eingebetteten Systemen liegt der Wert an der oberen Grenze, während bei Webanwendungen der Wert an der unteren Grenze liegt. Die Ursache ist die aktuell noch fehlende Unterstützung für View-Technologien. Bisherige Tests zeigen auch hier Automatisierungspotenzial.

Der Ansatz ist eine pragmatische Realisierung der MDA/MDSD-Konzepte und entwickelte sich durch Anwendungen in der Industrie, Forschung sowie Studium und Lehre zu einem praktischen Ansatz mit vielversprechenden Ergebnissen:

- Trennung von Modellierungswerkzeugen und MDSD-Prozess durch Einführung eines einfachen Metamodells
- Nutzung von UML sowie DSLs zur Anwendungsmodellerstellung
- Knappes und präzises Metamodell für objektorientierte Software
- Wiederverwendbare und anpassbare Modelltransformationen
- Hohe Investitionssicherheit in Transformationen durch unabhängiges Metamodell
- Hohe Quellcodequalität durch Generierung von sauberen, kompilierbaren, fehlerfreien und geprüften Quellcode gemäß Quellcode-Konventionen / Code-Style
- Unterstützt iterative und agile Softwareentwicklung

Die Aktualität der Modelle zum Quellcode ist ständig gegeben. Da diese Softwaremodelle auch ein Teil der Dokumentation sind, wird die Aktualität dieser nachhaltig verbessert.

Durch das vom Generator erzeugte Assoziationshandling sind die Objektmodelle stets konsistent. Bei bidirektionalen Assoziationen wird die Gegenseite automatisch gesetzt. Bei den Plattformen, bei denen eine selbst erstellte Bibliothek zum Assoziationshandling

eingesetzt wird, ist diese unter LGPL lizensiert, um auch in kommerziellen Projekten den Einsatz zu ermöglichen.

Das einfach strukturierte Metamodell sowie die Unterstützung aspektorientierter Programmierung (AOP) für Modelltransformationen, erlauben eine einfache projektspezifische Anpassung des GeneSEZ-Frameworks.

6 Ausblick

Aktuell werden Strukturmodelle und Zustandsautomaten unterstützt. An der Unterstützung bei der Entwicklung von Tests wird aktuell gearbeitet sowie auch an der stärkeren Einbeziehung von Anforderungsbeschreibungen.

Der GeneSEZ-Ansatz basiert auf Open Source (openArchitectureWare sowie verschiedenen Eclipse-Modeling-Projekten) und ist selbst Open Source. Er ist unter der GPL lizenziert mit Ausnahme der LGPL-lizenzierten Assoziationshandlingbibliotheken. Dadurch kann auch kommerzielle Software entwickelt werden. Lediglich Änderungen am Framework selbst müssen durch die GPL veröffentlicht werden.

Literaturverzeichnis

[ABG08] Arnold, O.; Beier, G.; Golubski, W.: Requirements Management 2.0. In: Proceedings of the IASTED International Conference on Software Engineering (SE 2008), Innsbruck, Austria, ACTA Press, Calgary, Canada, 2008.

[AND] Homepage von AndroMDA, http://www.andromda.org/

[Ba07] Baker, P.; et.al.: Model-Driven Testing: Using the UML Testing Profile. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2007.

[FOR] Homepage der fornax Platform, http://fornax-platform.org/

[GEN] Homepage des GeneSEZ-Projektes, http://www.genesez.de/

[Ha09] Haubold, T.; et.al.: The Technical Foundation of the GeneSEZ MDSD Approach. In: New Trends in Software Methodologies, Tools and Techniques: Proceedings of the Eighth SoMeT_09 (Frontiers in Artificial Intelligence and Applications), IOS Press, September 2009.

[HBG09] Haubold, T.; Beier, G.; Golubski, W.: A Pragmatic UML-based Meta Model for Object-oriented Code Generation. In: Proceedings of the 21st International Conference on Software Engineering & Knowledge Engineering (SEKE'2009), Juli 2009.

[MB02] Mellor, S.; Balcer, M.: Executable UML: A Foundation for Model-Driven Architectures. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 2002.

[MDA] Überblick der OMG MDA-Spezifikationen, http://www.omg.org/mda/specs.htm.

[OAW] Homepage von openArchitectureWare, http://www.openarchitectureware.org/

[OMG] Homepage der Object Managment Group (OMG). http://www.omg.org/.

[RR06] Robertson, S.; Robertson, J.: Mastering the requirements process, second edition Pearson Education, Inc. 2006.

[SCU10] Sculptor, http://fornax.itemis.de/confluence/display/fornax/Sculptor+(CSC)

[Th03] Thomas, D.: Uml - Unified or Universal Modeling Language? Journal of Object Technology, 2(1):7–12, January-February 2003.

[UML09] Die UML Superstructure Spezifikation der OMG 2.2, Februar 2009.

[WM07a] Wrycza, S.; Marcinkowski, B.: Towards a light Version of UML 2.x: Appraisal and Model. Organizacija, 40(4), 2007.

[WM07b] Wrycza, S.; Marcinkowski, B.: A light Version of UML 2: Survey and Outcomes. In Proceedings of the 2007 Computer Science and IT Education Conference, 2007.

Sensordatenverarbeitung mit dem Open Source Datenstrommanagementframework Odysseus

André Bolles[1], Dennis Geesen[1], Marco Grawunder[1],
Jonas Jacobi[1], Daniela Nicklas[2], H.-Jürgen Appelrath[1]

Abteilung Informationssysteme[1] / Datenbanken & Internettechnologien[2]
Department für Informatik
Universität Oldenburg
26129 Oldenburg

Abstract: Die Verarbeitung von kontinuierlichen Datenströmen, bspw. aus Sensoren, gewinnt zunehmend an Bedeutung. Flexible Möglichkeiten zur Verarbeitung solcher Daten werden durch Konzepte des Datenstrommanagements (DSM) bereitgestellt. Bisherige prototypische Datenstrommanagementsysteme (DSMS) wurden jeweils zur Evaluation bestimmter Verarbeitungskonzepte entwickelt. Der Vergleich oder die kombinierte Betrachtung dieser verschiedenen Konzepte ist bisher nur unter großem Aufwand möglich, da hierzu verschiedene DSMS integriert werden müssen. An der Universität Oldenburg wird daher ein Datenstrommanagementframework (DSMF) namens Odysseus entwickelt, welches es ermöglicht, nahezu beliebige Aspekte des DSM zu implementieren und zu evaluieren. Dieses Framework wird noch 2010 als Open Source DSMF veröffentlicht. Die vorliegende Arbeit stellt Odysseus und dessen Erweiterbarkeit vor und erläutert Fragestellungen, die bei der Veröffentlichung auftreten.

1 Einleitung

Sensoren sind in immer mehr Anwendungsfeldern, wie der Überwachung von Energienetzen oder zur Objektverfolgung in Fahrerassistenzsystemen, verbreitet. Die von Sensoren produzierten Datenströme müssen kontinuierlich verarbeitet werden. Konzepte des Datenstrommanagements (DSM, auch als complex event processing (CEP) bekannt) bieten hierzug analog zur Anfrageverarbeitung in Datenbanken flexible Möglichkeiten. Auf Grund der Heterogenität verschiedener, in den vergangenen Jahren entwickelter DSMS ist eine Kombination oder auch ein Vergleich der verschiedenen, dort implementierten Konzepte nur unter großem Aufwand möglich. Häufig ist eine Neuimplementierung in einem anderen DSMS nötig. An der Universität Oldenburg wurde daher mit Odysseus [BG+09] ein Datenstrommanagementframework (DSMF) entwickelt, welches als Plattform zur Evaluation nahezu beliebiger Konzepte des DSM dient. Odysseus wird noch 2010 als Open Source DSMF veröffentlicht, um auch anderen Forschungs- und Entwicklungseinrichtungen die Umsetzung ihrer Konzepte zur kontinuierlichen Verarbeitung von Sensordaten zu ermöglichen. In dieser Arbeit werden die Vorteile von Odysseus vorgestellt und Fragestellungen erläutert, die bei der Veröffentlichung als Open Source DSMF auftreten.

Hierzu werden im Kapitel 2 verwandte Arbeiten im Bereich DSM vorgestellt. Odysseus selbst wird mit seiner modularen Architektur in Kapitel 3 vorgestellt. Aufbauend darauf wird in Kapitel 4 beispielhaft beschrieben, wie eigene Konzepte der kontinuierlichen Datenverarbeitung in Odysseus integriert werden können. Fragestellungen, die bei der Veröffentlichung von Odysseus als Open Source DSMF auftreten, werden in Kapitel 5 erläutert, bevor abschließend diese Arbeit in Kapitel 6 zusammengefasst wird.

2 Verwandte Arbeiten

Obwohl DSM ein relativ junges Forschungsfeld ist, existieren bereits einige prototypische DSMS wie STREAM [AB+03], Borealis [AA+05], TelegraphCQ [CC+03] oder auch PIPES [KS09]. Diese DSMS dienen in erster Linie dazu bestimmte Verarbeitungskonzepte von Datenströmen zu evaluieren. So werden in STREAM Datenströme für die Verarbeitung zunächst in Relationen und anschließend zurück in Datenströme transformiert. Auch die Umsetzung des sogenannten Fensterkonzeptes unterscheidet sich in den einzelnen Systemen. Fenster werden dazu eingesetzt Ausschnitte auf Datenströmen zu definieren, um nicht alle Daten des Datenstroms für die Verarbeitung einlesen zu müssen. Diese Fenster werden in [AA+05, CC+03] bspw. an die Operatoren angehängt, damit diese entscheiden können, welche Elemente des Datenstroms aktuell zur Verarbeitung betrachtet werden müssen. In PIPES existieren dagegen eigene Fensteroperatoren, die sogenannte Gültigkeitsintervalle an die Elemente eines Datenstroms anhängen. Diese Gültigkeitsintervalle können von den nachfolgenden Operatoren ausgelesen werden, so dass auch hier entschieden werden kann, welche Elemente aktuell gültig sind. Ein alternativer Ansatz zu den Gültigkeitsintervallen wird wiederum in Aurora verwendet, wo die Gültigkeit mit sogenannten positiven Elementen beginnt und durch entsprechende negative Elemente beendet wird. Neben diesen Aspekten unterscheiden sich die Systeme auch in den verwendeten Datenmodellen. Während die zuvor genannten Systeme auf das relationale Modell setzen, werden in [GG+04] XML-Datenströme verarbeitet. Natürlich gibt es weitere Unterschiede zwischen den Systemen, die an dieser Stelle aber nicht alle genannt werden können.

Im Vergleich zu den genannten DSMS wird mit Odysseus jedoch nicht das Ziel verfolgt, bestimmte Konzepte des DSM evaluieren zu können, sondern vielmehr eine Plattform bereitzustellen, mit der eben nahezu beliebige Konzepte evaluiert werden können. So können mit Odysseus bspw. verschiedene Datenmodelle untersützt werden oder aber auch beide zuvor genannten Fensterrepräsentationen. Die Architektur, die für ein solches DSMF notwendig ist, wird im folgenden Kapitel beschrieben.

3 Odysseus

Odysseus ist ein Datenstrommanagementframework (DSMF), welches vollständig in Java implementiert ist. Es besitzt eine komponentenorientierte, OSGi-basierte Architektur (Abbildung 1), die es erlaubt DSMS für verschiedene Anwendungsdomänen zu entwickeln.

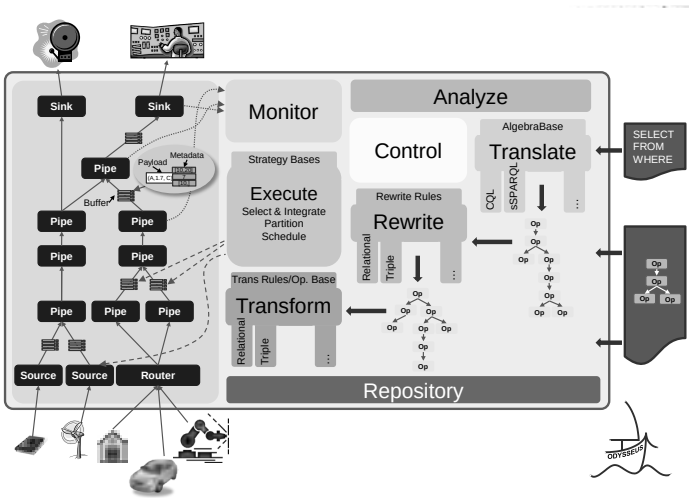


Abbildung 1: Architektur von Odysseus

Odysseus' Architektur kann dabei in sogenannte Fix- und Variationspunkte unterteilt werden. Fixpunkte sind die Schnittstellen, die Odysseus für verschiedene Aufgaben zur Verfügung stellt. Mit den Variationspunkten ist eine Anpassung der einzelnen Komponenten an verschiedene Bedürfnisse möglich.

In Abbildung 1 ist rechts zu erkennen, dass verschiedene Möglichkeiten existieren, Anfragen in Odysseus zu installieren. Neben verschiedenen Anfragesprachen wie Stream-SQL oder Stream-SPARQL [BG+08] (*SELECT, FROM, WHERE*) können Anfragepläne auch direkt über eine prozedurale Anfragesprache in Form eines Operatorbaums beschrieben werden. Über eine einheitliche in Odysseus implementierte Parser-Schnittstelle (*TRANSLATE*) (Fixpunkt) können diese Anfragesprachen in logische Anfragepläne übersetzt werden. Auch andere Anfragesprachen, wie die Esper Query Language (http://esper.codehaus.org), können so in Odysseus integriert werden.

Logische Anfragepläne stellen Informationen über eine Anfrage, wie bspw. Join-Prädikate bereit. Odysseus stellt abstrakte logische Operatoren als Fixpunkt bereit, die die Verknüpfung mehrerer Operatoren zu ganzen Anfrageplänen erlauben. Als Variationspunkt können eigene logische Operatoren von diesen abstrakten Operatoren ableiten und so spezifische Metadaten zur Verfügung stellen.

Logische Operatoren enthalten nur die Metadaten, aber noch keine Algorithmen zur Ausführung einer Anfrage. Daher werden logische Anfragepläne durch eine Regelmaschine in physische, ausführbare Anfragepläne übersetzt. Als Fixpunkt stellt Odysseus hier eine Schnittstelle zu einer Regelengine[1] bereit. Mit eigenen Regeln kann die Übersetzung in verschiedene physische Anfragepläne gesteuert werden (*REWRITE, TRANSFORM*, Variationspunkt).

[1] Drools, http://www.jboss.org/drools, zuletzt besucht: 04.05.2010

Für physische Anfragepläne stellt Odysseus ein Metadatenframework bereit, über das beliebige Metadaten an einzelne Datenstromelemente angehängt werden können. So kann bspw. sowohl der in Kapitel 2 genannte Positiv/Negativ-Ansatz als auch der Intervallansatz umgesetzt werden. Auch abstrakte, physische Operatoren sind zunächst datenmodellunabhängig, so dass eigene Operatoren zur Verarbeitung neuer Datenmodelle entwickelt werden können.

Es gibt weitere Komponenten, mit denen die Ausführung und Überwachung von Anfragen gesteuert werden kann. Die *EXECUTE*-Komponente bietet eine Schnittstelle (Fixpunkt) zur Ausführung physischer Anfragepläne, wie sie links in der Abbildung dargestellt sind. Verschiedene Ausführungsstrategien bilden einen Variationspunkt. Mit *CONTROL* kann bei Veränderungen der Anfrageverarbeitung, die über *MONITOR* und *ANALYZE* identifiziert werden, korrigierend eingegriffen werden. Weitere Details sind in [BG+09] dargestellt.

Durch die OSGi-basierte Architektur werden alle genannten Komponenten in eigene Bundles ausgelagert. Auch die Entwicklung neuer Konzepte bspw. mit neuen Metadaten, die an den Datenstromelementen angehängt werden, kann in eigene Bundles ausgelagert werden, so dass der Rest des Frameworks unverändert bleibt. Dies erlaubt zudem, nur bestimmte Bundles für bestimmte Anwendungsdomänen bereitzustellen, so dass Entwickler, die Odysseus für ihre Anwendung nutzen wollen, nicht das vollständige Framework mit allen spezialisierten Bundles installieren müssen.

4 Erweiterbarkeit von Odysseus

Dieses Kapitel illustriert die Erweiterbarkeit von Odysseus anhand des Beispiels der Objektverfolgung in einem Fahrerassistenzsystem.

Objektverfolgung Bei der Objektverfolgung sollen Objekte die bspw. durch eine Bildverarbeitung in einem Videobild erkannt wurden, in späteren Videobildern wiedererkannt werden. Beispielhaft passiert hierbei Folgendes: Ein Objekt o_0 sei zum Zeitpunkt t_0 an Position x_0 erkannt worden. Zum Zeitpunkt t_1 sei ein ähnliches Objekt o_1 an Position x_1 erkannt worden. Jetzt soll festgestellt werden, ob $o_0 = o_1$ ist. Hierzu wird mit einem Bewegungsmodell die Position $\hat{x}_0$ von o_0 zum Zeitpunkt t_1 geschätzt. Dann werden $\hat{x}_0$ und x_1 verglichen. Liegen diese dicht beieinander, kann davon ausgegangen werden, dass $o_0 = o_1$ gilt. Jetzt muss jedoch noch die tatsächliche Position x_{neu} aus $\hat{x}_0$ und x_1 geschätzt werden. Hierzu werden Filter, wie der Kalman-Filter eingesetzt. Anschließend beginnt dieser Prozess von vorne, diesmal jedoch mit der neuen Position x_{neu}.

Zyklische Anfragen Will man diesen Prozess der Objektverfolgung in einem DSMS abbilden, so muss die Möglichkeit geschaffen werden, zyklische Anfragen formulieren und ausführen zu können. Weiterhin müssen immer die aktuellen Zustände der Objekte (die Position im zuvor genannten Beispiel) zwischengespeichert werden.

Zur Formulierung zyklischer Anfragen bedarf es der Erweiterung einer Anfragesprache. Dies sei hier am Beispiel von StreamSQL gezeigt. Will man einen Zyklus ausdrücken, so muss es möglich sein, Daten aus einer Quelle auszulesen und auch wieder in diese Quel-

le hineinzuschreiben. Während mit SELECT $*$ FROM <name> bereits Daten aus einer Quelle ausgelesen werden können, wird mit SELECT $*$ INTO <name> FROM <name> eine Erweiterung der Grammatik vorgenommen, die auch das Schreiben in eine Quelle ausdrückt.

Eine solche Quelle soll die Speicherung und Auslieferung von Objektzuständen übernehmen und wird als sogenannter Broker-Operator in Odysseus integriert. Die Integration dieses Operators erfolgt an verschiedenen Stellen im System. Ein logischer Broker erlaubt den Aufbau zyklischer Anfragepläne. In einem physischen Broker ist der Algorithmus zur Verteilung der Objektzustände enthalten. Bei der Verteilung muss u.a. darauf geachtet werden, dass Objektzustände nicht weitergeleitet werden, bevor eine Aktualisierung abgeschlossen ist.

Die Komponenten für zyklische Anfragen befinden sich in einem eigenen OSGi-Bundle, so dass die Unabhängigkeit des Odysseus-Kerns von diesem neuen Bundle gewährleistet wird.

5 Open Source Release

Es ist noch innerhalb dieses Jahres eine Veröffentlichung von Odysseus als Open Source geplant. Es ist das Ziel anderen Forschergruppen ein Framework für DSMS zur Verfügung zu stellen. Hierbei wird nicht eine Weiterentwicklung des Odysseus-Kerns angestrebt, sondern vielmehr die Bereitstellung neuer Module durch andere Forschergruppen. Außerdem erhoffen wir uns eine Verbreitung des Themas Datenstrommanagement. Wir prüfen aktuell noch unterschiedliche Alternativen, wie die Art der Veröffentlichung aussehen kann und welche Lizenzmodelle auf Grund von verwendeten Bibliotheken notwendig sein werden. Als mögliche Verbreitungsmechanismen sind geplant:

Die Veröffentlichung von Odysseus als Eclipse Rich Client Anwendung Dabei wird ein Teil der Software direkt zum Download zur Verfügung gestellt und weitere Module und Aktualisierungen können mit Hilfe des in Eclipse eingebauten Update-Mechanismus vertrieben werden. Es ist auf diese Weise sehr einfach möglich, unterschiedliche Arten von Paketen zu definieren, von einem einfachen relationalen Odysseus bis hin zu einem kompletten Odysseus, ohne die einfache Updatemöglichkeit zu verlieren.

Der Zugriff auf das SVN Derzeit wird Odysseus durch das Versionsverwaltungssystem SVN verteilt. Solange nur eine beschränkte und bekannte Anzahl von Nutzern zugreift, ist dies die beste Lösung. Bei einer Veröffentlichung würden wir hier vermutlich zunächst nur lesenden Zugriff auf das Repository erlauben.

Download von Releases Die letzte angedachte Möglichkeit besteht darin, den vollständigen Sourcecode in Form eines Archivs zum Download zur Verfügung zu stellen. Dies hätte den Nachteil, dass nicht immer die aktuellste Version zur Verfügung stehen würde und Fehler u.U. erst später korrigiert werden.

Odysseus verwendet eine Reihe von externen Bibliotheken. Vor der finalen Veröffentlichung ist hier zunächst noch einmal zu testen, welche Lizenzmodelle in Frage kom-

men können. Da wir möglichst wenig einschränken und eine weite Verbreitung fördern möchten, präferieren wir aktuell Lizenzmodelle wie sie z.B. von der Apache Foundation angewendet werden. Da einige Bibliotheken (u.a. JEP und Drools) von uns speziell für Odysseus erweitert wurden muss hier zunächst der rechtliche Rahmen geklärt werden.

6 Zusammenfassung

In dieser Arbeit wurde Odysseus als ein Framework zur Entwicklung von Datenstrommanagementsystemen vorgestellt. Die Architektur und die Erweiterungsmöglichkeiten von Odysseus wurden sowohl abstrakt beschrieben als auch anhand der Erweiterung um zyklische Anfragen dargestellt. Die Integration neuer Konzepte in Odysseus kann bei Einhaltung der vorgegebenen Schnittstellen mit relativ geringem Aufwand erfolgen, da nur neue Konzepte angefasst werden müssen und bestehende Konzepte wie z. B. der Scheduler wiederverwendet werden können. Abschließend wurde aufgezeigt, welche Fragestellungen bei der Veröffentlichung von Odysseus als Open Source Framework auftreten. Verschiedende Veröffentlichungsmechanismen wie eine Eclipse Rich Client Anwendung, ein SVN-Zugriff oder Release-Downloads sind vorstellbar. Außerdem müssen Lizenzmodelle, insbesondere hinsichtlich der Verwendung von Bibliotheken untersucht werden.

Literatur

[AA+05] Daniel J. Abadi, Yanif Ahmand et al. The Design of the Borealis Stream Processing Engine. In *Second Biennial Conference on Innovative Data Systems Research (CIDR 2005)*, Januar 2005.

[AB+03] Arvind Arasu, Brian Babcock et al. STREAM: The Standford Stream Data Manager. *IEEE Data Engineering Bulletin*, 26(1), 2003.

[BG+08] Andre Bolles, Marco Grawunder et al. Streaming SPARQL - Extending SPARQL to Process Data Streams. In *ESWC 2008*, 2008.

[BG+09] Andre Bolles, Marco Grawunder et al. Odysseus: Ein Framework für maßgeschneiderte Datenstrommanagementsysteme. In *Informatik 2009 - Im Fokus das Leben*, 9 2009.

[CC+03] Sirish Chandrasekaranand, Owen Cooper et al. TelegraphCQ: Continuous Dataflow Processing for an Uncertain World. In *Proceedings of the Conference on Innovative Data Systems Research (CIDR)*, 2003.

[GG+04] Todd J. Green, Ashish Gupta et al. Processing XML streams with deterministic automata and stream indexes. *ACM Trans. Database Syst.*, 29(4), 2004.

[KS09] Jürgen Krämer und Bernhard Seeger. Semantics and implementation of continuous sliding window queries over data streams. *ACM Trans. Database Syst.*, 34(1), 2009.

Java und Open-Source
— Ein Geben und Nehmen —

Bernd Müller
Ostfalia
Hochschule für angewandte Wissenschaften
38302 Wolfenbüttel
bernd.mueller@ostfalia.de

Abstract: Wir zeigen auf der Basis aktueller Java-Spezifikationen und Open-Source-Systeme, dass Java und Open-Source sich gegenseitig befruchten und verschiedene Open-Source-Frameworks im Java-Bereich einen regen gegenseitigen Austausch und eine beachtliche Verwendung anderer Open-Source-Systeme betreiben. Ferner zeigen wir, dass auch Nicht-Open-Source-Systeme im Java-Umfeld von der Verwendungsmöglichkeiten freier Software lebhaften Gebrauch machen.

1 Java und Open-Source

Im Mai 2006 gab Sun bekannt, dass das Java Development Kit (JDK) zukünftig unter einer GPL-Lizenz stehen wird. Im November 2006 wurden zunächst der Compiler und die JVM als Open-Source veröffentlicht. Die Klassenbibliothek wurde schrittweise auf Open-Source umgestellt, da sie proprietären Code enthielt, für den Sun nicht die notwendigen Rechte besaß. Mittlerweile stellt das Projekt OpenJDK [JDK] ein vollständiges Open-Source JDK bereit, so dass auch Linux-Distributionen wie Fedora, die ausschließlich frei Software enthalten (wollen), es verwenden können.

Die Öffnung des Java JDK ist allgemein bekannt. Weniger bekannt ist der Umstand, dass Open-Source-Implementierungen von Java-Frameworks Java-Spezifikationen richtungsweisend und maßgeblich befruchtet haben. Ebenfalls weniger bekannt ist der sehr gut funktionierende Austausch bzw. die Verwendung von Open-Source Systemen sowohl in anderen Open-Source aber auch kommerziellen Systemen. Wir wollen den Bekanntheitsgrad dieser Fakten erhöhen und belegen die

- Beeinflussung von Java-Spezifikationen durch Open-Source
- Verwendung von Open-Source Java-Systeme in kommerziellen Systemen
- Verwendung von Open-Source Java-Systemen in anderen Open-Source Java-Systemen

durch einige ausgewählte Beispiele.

2 Beeinflussung von Java-Spezifikationen durch Open-Source-Implementierungen

Die Java 2 Platform Enterprise Edition Specification 1.4 (J2EE 1.4) besteht aus 21 einzelnen Spezifikationen, die in Tabelle1 aufgezählt sind. Allein die Anzahl der Spezifikationen

Enterprise JavaBeans	Java API for XML Processsing
JavaServer Pages	SOAP with attachements API for Java
Servlets	Java API for XML registries
Java Database Connectivity	Java Enterprise Edition Management Spec.
Java Naming and Directory Interface	Enterprise Edition Deployment Specification
Java Message Service	Java Management Extensions
Java Transaction API	Java Authorization Contract for Containers
Java Transaction Service	Java Authentication and Authorization Service
JavaMail	Web Services for J2EE
JavaBeans Activation Framework	Java API for XML-based RPC
J2EE Connector Architecture	

Tabelle 1: J2EE 1.4 Spezifikationen

macht deutlich, wie komplex die Entwicklung von Unternehmensanwendungen auf Basis von J2EE 1.4 war. Besondere Kritik wurde an der Spezifikation von Enterprise JavaBeans laut, die dem DRY-Prinzip widersprach und mit mehreren Java- und XML-Artefakten pro einzelner EJB als extrem aufwändig und damit praktisch als nicht verwendbar galt.

2.1 Hibernate und JPA

Als Reaktion auf die oben angesprochene Komplexität entstanden z.B. das Spring- und das Hibernate-Framework, wobei letzteres die Persistenz von Java-Objekten in Relationalen Datenbanken realisiert. Wir wollen Hibernate verwenden, um das Umdenken in der Java-Spezifikationsentwicklung zu verdeutlichen. Während früher (J2EE 1.4 und davor) Spezifikationen auf der grünen Wiese entwickelt wurden und zu Spezifikationsungetümen führten, wurden danach existierende Systeme, häufig Open-Source-Systeme, als funktionierendes Beispiel und Blaupause für die Spezifikation verwendet. Die Java Platform Enterprise Edition 5 (Java EE 5) wurde als erste nach diesem Prinzip entwickelt. Die Persistenz wurde in das Java Persistence API (JPA), definiert als Java Specification Request 220 (JSR-220), ausgelagert. Wir wollen diesen JSR im Folgenden als Vergleichssystem für Hibernate heranziehen.

Der folgende Code-Ausschnitt zeigt, wie ein Entity mit Hibernate persistent gemacht wird. Der Code ist intuitiv verständlich, so dass wir auf Erläuterungen verzichten.

```
SessionFactory factory =
    new Configuration().configure().buildSessionFactory();
Session session = factory.openSession();
```

```
Transaction tx = session.beginTransaction();
session.save(<ein Entity>);
tx.commit();
...
```

Während die Persistierung von Entities mit J2EE 1.4 nach einem völlig anderen Muster erfolgte, nahm sich die JPA-Expertengruppe Hibernate als Vorbild und definierte eine leichtgewichtige Persistenz. Der folgende Code-Ausschnitt zeigt das obige Beispiel unter Verwendung von JPA.

```
EntityManagerFactory factory =
    Persistence.createEntityManagerFactory("<name>");
EntityManager em = factory.createEntityManager();
EntityTransaction tx = em.getTransaction();
tx.begin();
em.persist(<ein Entity>);
tx.commit();
...
```

Man erkennt deutliche Parallelen, die durch konsistente Umbenennung von *Session* in `EntityManager` besonders hervortreten. Die Inspiration der JPA-Expertengruppe durch Hibernate wird durch dieses einfache Beispiel deutlich und ist auch in vielen anderen Code-Mustern, aber auch auf konzeptioneller Ebene erkennbar. Hibernate ist mittlerweile als JPA-Implementierung zertifiziert. Gavin King, der konzeptionelle Vater von Hibernate, war Mitglied im JPA-Spezifikationsgremiums.

2.2 Hibernate Validator, Seam und Bean Validation

Hibernate Validator [HV] ist ein Projekt, das die Validierung von Objekt-Properties zunächst für Hibernate-Entites, später aber auch für gewöhnliche POJOs realisiert. Die Validierung erfolgt durch die Annotation der Properties z.B. mit `@Min`, `@Max`, `@NotNull`, `@NotEmpty`, `@Future` und `@Past`, deren Namen für sich selbst sprechen. In geschichteten Anwendungen erfolgt häufig eine Validierung auf der GUI- aber auch auf der Persistenz-Ebene, was dem DRY-Prinzip widerspricht. JBoss Seam ist ein Anwendungs-Framework, das für die Oberfläche JSF und für die Persistenz JPA verwendet. Seam zeigte, wie mit Hilfe von Hibernate Validator das DRY-Prinzip respektiert und die auf der Persistenz-Schicht erfolgte Annotation von Properties mit Validierungs-Constraints bis zur Validierung von Eingaben in der Oberfläche hochgehoben werden kann.

Java EE 6 enthält mit *Bean Validation* eine Spezifikation, die Konzepte von Hibernate Validator aufnimmt und in Java EE 6 Containern sowohl für die Oberfläche (JSF) als auch die Persistenz (JPA) verwendet werden kann. Bean Validation ist als JSR-303 spezifiziert und Hibernate Validator ist die Referenzimplementierung. Emmanuel Bernard, ein JBoss-Mitarbeiter und maßgeblicher Kopf von Hibernate Validator, war Leiter des Spezifikationsgremiums.

2.3 Seam und CDI

Das allgemein Martin Fowler zugeschriebene Konzept der *Dependency Injection* (DI) [Fow] wurde in Java EE 5 erstmalig in einer Java-Spezifikation verwendet, um Container-spezifische Objekte und EJBs zu injizieren. Seam erweiterte die dortige Verwendung von DI auf beliebige Objekte, insbesondere die Verwendung derartiger Objekte in der JSF-basierten GUI-Schicht und verallgemeinerte das Prinzip des DI. Durch den Erfolg motiviert, initiierte JBoss den JSR-299, der nach zweimaliger Umbenennung schließlich den Namen *Contexts and Dependency Injection for the Java EE platform* (CDI) [CDI] erhielt. Er führt die durch Seam begonnenen Innovationen fort und erweiterte sie z.B. um Typsicherheit. CDI ist in Java EE 6 enthalten. Gavin King war als maßgeblicher Kopf hinter Seam auch Leiter des Spezifikationsgremiums für CDI. Die Referenzimplementierung von CDI wird unter dem Code-Namen Weld [Wld] als Seam-Teilprojekt von Seam-Entwicklern als Open-Source realisiert.

2.4 Guice, Spring und Dependency Injection for Java

Kurz vor Fertigstellung des JSR-299 wurde von Google und SpringSource ein weiterer Vorschlag zur Spezifikation von DI in Java eingereicht. Google und SpringSource bieten mit *Google Guice* [GG] und *Spring* [Spr] zwei weit verbreitete DI-Frameworks auf Open-Source-Basis an. Der Vorschlag von *Dependency Injection for Java* als JSR-330 führte zu etlichen Unruhen im Java-Bereich, da eine Konkurrenz zum JSR-299 gesehen wurde. Mittlerweile ist die Spezifikation jedoch verabschiedet und Basis des JSR-299. Bob Lee von Google und Rod Johnson von SpringSource waren die Leiter des Spezifikationsgremiums.

3 Kommerzielle Systeme nutzen Open-Source

Neben der Motivation von Java-Spezifikationen durch Open-Source-Systeme ist die häufige Verwendung von Open-Source-Systemen in kommerzieller Java-Software ebenfalls wenig bekannt. Wir wollen diesen Zustand hier ändern.

Die kommerziellen Java-Entwicklungssysteme (JDK) von Sun [SJK], Oracle [OJK] und IBM [IJK] verwenden eine ganze Reihe von Open-Source-Systemen. Alle genannten Systeme verwenden für die Verarbeitung von XML Systeme der Apache Software Foundation, die unter Open-Source-Lizenzen stehen. Es sind dies Xalan [Xal], ein XSLT-Prozessor, Xerces [Xer], ein XML-Parser und -Generator und XPath [Xpt], die XPath-Implementierung innerhalb Xalans.

Weiterhin verwenden die Entwicklungssysteme von Sun und Oracle BCEL [BCL] (Byte Code Engineering Library), ein System zur Analyse und Manipulation von Java-Byte-Code, während IBM hier eigene, nicht Open-Source-Wege, geht.

Besonders bemerkenswert ist der Umstand, dass Suns JDK die *JavaDB* enthält, ein relationales Datenbankmanagementsystem, das in Java implementiert ist. Das Interessante an JavaDB ist seine Geschichte, die der Geschichte von Firefox, OpenOffice und Eclipse ähnelt, d.h. eine kommerzielle Entwicklungs- und Vermarktungsphase mit anschließender Änderung der Lizenzbedingungen hin zu Open-Source. Die 1996 gegründete Firma Cloudscape entwickelte ein relationales Datenbankmanagementsystem, das ebenfalls den Namen Cloudscape trug. 1999 übernahm Informix Software Inc. die Firma Cloudscape. 2001 übernahm IBM Informix und das Datenbanksystem wurde unter dem Namen *IBM Cloudscape* vermarktet. Im Jahre 2004 übertrug IBM Cloudscape an die Apache Software Foundation unter den neuen Namen *Derby* [DRB], wo es seitdem weiterentwickelt wird. Sun legt Derby seit Version 6 dem JDK bei. Die entsprechenden Dateien im JDK tragen weiterhin den Namen Derby und enthalten die Apache-Lizenz-Datei.

4 Firmenübergreifende Verwendung von OS-Implementierungen

In diesem Abschnitt untersuchen wir die beiden populärsten Open-Source Application-Server. Obwohl Open-Source, werden beide Application-Server zu großen Teilen von Mitarbeitern der jeweiligen Herstellerfirmen implementiert und nur zu geringen Anteilen von anderen Open-Source Entwicklern. Bei den Application-Servern handelt es sich zum einen um den *JBoss Application Server* [JBA], der nach Auskunft von JBoss meistgenutze Java-EE-Application-Server, für den JBoss auch kommerziellen Support anbietet [JEM]. Zum anderen ist dies *GlassFish* [GLF], die Referenzimplementierung von Java-EE 6, für den Sun ebenfalls kommerziellen Support anbietet [GFS]. Da uns keine offiziellen Aussagen der Entwickler über die Verwendung von Fremdsystemen bekannt sind, basieren unsere Ausführungen auf der Analyse der Jar-Dateien der Systeme, die wir zu Dokumentationszwecken angeben. Tabelle 2 zeigt die von GlassFish und JBoss-AS verwendeten OS-Implementierungen.

Durch die Aufstellung wird offensichtlich, wie lebhaft die OS-Gemeinschaft agiert. Sun verwendet JBoss-Systeme, JBoss verwendet Sun-Systeme und beide Hersteller bedienen sich bei OS-Implementierungen von Apache, Codehaus, GNU und anderen.

5 Zusammenfassung

Wir haben an Beispielen belegt, dass Open-Source und Java keine getrennten Welten darstellen, sondern die Entwicklung von Java an einem sehr aktiven Open-Source-Ökosystem partizipiert. Dies ist durch die Motivation von Java-Spezifikationen durch erfolgreiche OS-Frameworks zu belegen, von denen wir lediglich beispielhaft JPA, Bean Validation und CDI aufgeführt haben. Weiterhin haben wir Beispiele genannt, in denen Firmen Software unter nicht Open-Source-Lizenzen entwickeln, die Open-Source-Software enthält. Außerdem haben wir gezeigt, dass bei bekannten OS-Application-Servern eine firmenübergreifende Verwendung von OS-Systemen stattfindet. Suns GlassFish verwendet von JBoss

Jar	System	Entwickler(-gruppe)
ant.jar	Ant, Build-Tool [Ant]	Apache
antlr-repackaged.jar	ANTLR, Parser Generator [ATL]	Terence Parr
bean-validator.jar	Hibernate Validator [HV], RI JSR-303	JBoss
asm-all-repackaged.jar	ASM Byte-Code Analyse und Manipulation [ASM]	OW2 Consortium
groovy-1.7.2.jar	Groovy Skriptsprache [GRY]	Codehaus
org.apache.felix.*.jar	Felix OSGI-Platform [FLX]	Apache
weld-osgi-bundle.jar	Weld [Wld], RI JSR-299	JBoss
org.eclipse.persistence.*.jar	EclipseLink [EL], RI JSR-317	Eclipse
web-ajp.jar	AJP-Protokoll-Implementierung [AJP]	Apache
web-core.jar	Apache Tomcat (Catalina) [CAT]	Apache
jettison.jar	Java-API für JSON [JSN]	Codehaus
antlr-repackaged.jar	ANTLR, Parser Generator [ATL]	Terence Parr
bcel.jar	Byte Code Engineering Library [BCL]	Apache
bsf.jar	Bean Scripting Framework	Apache
bsh.jar	BeanShell [BSH]	Pat Niemeyer
commons-collections.jar	Collection-Framework	Apache
commons-logging.jar	Logging-Framework	Apache
dtdparser121.jar	DTD-Parser [DTD]	Wutka Consulting
jsf-impl.jar	Mojarra JSF RI [JSF]	Sun
hsqldb.jar	HSQL-DBMS [HQL]	HSQL Dev-Group
log4j.jar	log4j Logging [L4J]	Apache
mail.jar	Java Mail [JM]	Sun
quartz.jar	Quartz Enterprise Job Scheduler [QTZ]	Terracotta
slf4j.jar	Simple Logging Facade for Java [S4J]	Quality Open Software
dom4j.jar	XML-Bibliothek [D4J]	dom4j
getopt.jar	Zugriff auf Kommandozeilenparameter	GNU
trove.jar	High Performance Collections for Java	GNU
wstx.jar	High-performance XML Processor	Codehaus
concurrent.jar	Concurrent Implementation	SUNY Oswego
jaxb-*.jar	JAXB-Implementierung	Sun
wsdl4j.jar	WSDL-Implementierung	IBM und Sun
xmlsec.jar	Java XML Security	Apache

Tabelle 2: Von GlassFish (erster Teil) und JBoss-AS (zweiter Teil) verwendete OS-Systeme

entwickelte Software, JBoss-AS verwendet von Sun entwickelte Software.

Aus Platzgründen haben wir in dieser Arbeit nur wenige Beispiele anführen können, die die Verbindung von Java mit Open-Source belegen. Die Liste ließe sich deutlich erweitern. Wir hoffen den Leser davon überzeugt zu haben, dass Java und Open-Source in einer Beziehung stehen, die mit „*Ein Geben und Nehmen*“ treffend beschrieben ist.

Literatur

[AJP] Apache Tomcat Connector. Web. http://tomcat.apache.org/connectors-doc/.

[Ant] Ant Java-Build-Tool. Web. http://ant.apache.org/.

[ASM] ASM. Web. http://asm.ow2.org/index.html.

[ATL] ANTLR Parser Generator (ANother Tool for Language Recognition. Web. http://www.antlr.org/.

[BCL] BCEL (Byte Code Engineering Library). Web. http://jakarta.apache.org/bcel/.

[BSH] BeanShell Lightweight Scripting for Java. Web. http://www.beanshell.org/.

[CAT] Apache Tomcat, Catalina Servlet-Container. Web. http://tomcat.apache.org/.

[CDI] JSR 299, Contexts and Dependency Injection for the Java EE platform. Web. http://jcp.org/en/jsr/summary?id=299.

[D4J] dom4j XML-Bibliothek. Web. http://www.dom4j.org/dom4j-1.6.1/.

[DRB] Apache Derby. Web. http://db.apache.org/derby/index.html.

[DTD] Java DTD Parser, Wutka Consulting. Web. http://www.wutka.com/dtdparser.html.

[EL] EclipseLink — Eclipse Persistence Services Project. Web. http://www.eclipse.org/eclipselink/.

[FLX] Apache Felix – OSGi R4 Service Platform. Web. http://felix.apache.org/site/index.html.

[Fow] Fowler Martin. Dependency Injection, Inversion of Control. Web. http://www.martinfowler.com/articles/injection.html.

[GFS] Sun GlassFish Enterprise Server Support. Web. http://developers.sun.com/appserver/support/index.jsp.

[GG] Google Guice. Web. http://code.google.com/p/google-guice/.

[GLF] GlassFish Open Source Application Server. Web. https://glassfish.dev.java.net/.

[GRY] Groovy – An agile dynamic Language for the Java Platform. Web. http://groovy.codehaus.org/.

[HQL] HyperSQL Pure Java Database. Web. http://www.hsqldb.org/.

[HV] Hibernate Validator. Web. http://www.hibernate.org/subprojects/validator.html.

[IJK] JDK von IBM. Web. https://www.ibm.com/developerworks/java/jdk/.

[JBA] JBoss Application Server. Web. http://www.jboss.org/jbossas.

[JDK] OpenJDK. Web. http://openjdk.java.net/.

[JEM] JBoss Enterprise Middleware Subscriptions. Web. http://www.jboss.com/services/subscriptions/.

[JM] JavaMail. Web. http://java.sun.com/products/javamail/.

[JSF] Project Mojarra. Web. https://javaserverfaces.dev.java.net/.

[JSN] Jettison — A JSON StAX Implementation. Web. http://jettison.codehaus.org/.

[L4J] Apache Log4j. Web. http://logging.apache.org/log4j/.

[OJK] JDK von Oracle (JRockit). Web. http://www.oracle.com/technology/products/jrockit/index.html.

[QTZ] Quartz Enterprise Job Scheduler. Web. http://www.quartz-scheduler.org/.

[S4J] Simple Logging Facade for Java (SLF4J). Web. http://www.slf4j.org/index.html.

[SJK] JDK von Sun. Web. http://java.sun.com/javase/downloads/index.jsp.

[Spr] Spring. Web. http://www.springsource.com/products/enterprise.

[Wld] JSR-299 und Weld Overview. Web. http://seamframework.org/Weld.

[Xal] The Apache Xalan Project. Web. http://xalan.apache.org/.

[Xer] The Apache Xerces Project. Web. http://xerces.apache.org/.

[Xpt] Using the JAXP XPath APIs. Web. http://xml.apache.org/xalan-j/xpath_apis.html.

Toolunterstützte Transformation fachkonzeptioneller Verwaltungsprozesse mit der bflow*Toolbox

Beate van Kempen[1], Frank Hogrebe[2], Astrid Schranz[2]

[1]ITK Rheinland
Fachbereich Kundenmanagement
Michaelstraße 16, 41456 Neuss
beate.vankempen@itk-rheinland.de

[2]Landeshauptstadt Düsseldorf
Competence Center eGovernment (CC eGov)
Burgplatz 1, 40213 Düsseldorf
{frank.hogrebe|astrid.schranz}@duesseldorf.de

Abstract: Verwaltungsleistungen werden zunehmend über das Internet nachgefragt. Diese Entwicklung wird durch die EU-Dienstleistungsrichtlinie weiter verstärkt, die eine elektronische Abwicklung behördlicher Genehmigungsverfahren auch „aus der Ferne" fordert. Durchgehend elektronische Prozesse sind jedoch in der öffentlichen Verwaltung die Ausnahme; Medienbrüche innerhalb der Prozessabläufe dagegen die Regel. Die Integration komplexer Verwaltungsprozesse in dynamische Workflows erfordert bereits auf der Ebene der fachkonzeptionellen Modellierung die Erhebung von transformationsrelevanten Informationen, die für eine spätere (Teil-)Automation der Prozesse erforderlich sind. Der Beitrag beschreibt einen Ansatz, wie auf Basis objektorientierter ereignisgesteuerter Prozessketten (oEPK), unter Einbeziehung von WSDL-Datenkonzepten, die Grundlage für eine durchgängige Prozessautomation gelegt werden kann. Auf Basis der bflow*Toolbox werden die dazu notwendigen Transformationsschritte dargestellt und prototypisch angewandt.

1 Ausgangslage

Dienstebasierte Informationssysteme sollen die Anpassung von IT-Systemlandschaften an sich ändernde Anforderungen kostengünstig, zeitnah und flexibel, durch eine weitestgehende Entkopplung von betriebswirtschaftlichen und technischen Aspekten ermöglichen. Für den öffentlichen Sektor stellt die EU-Dienstleistungsrichtlinie eine solche geänderte Anforderung dar [EU06]. Die Richtlinie erfordert neben der Optimierung und elektronischen Verfügbarkeit von Verwaltungsdienstleistungen den Aufbau dienstebasierter Architekturen [DOL08, S. 122]. Im Abschlussbericht des Deutschland-Online-Projektes zur IT-Umsetzung der Dienstleistungsrichtlinie wird insbesondere der Abbau von Medienbrüchen für die elektronische Verfahrensabwicklung mit durchgehenden Prozessen gefordert.

In diesem Kontext stellt der Beitrag einen Ansatz vor, elektronische Verwaltungsprozesse bereits bei der fachkonzeptionellen Modellierung auf eine Transformation hin zu einem Workflowsystem vorzubereiten. Die Arbeit ist wie folgt gegliedert: Im zweiten Abschnitt werden zunächst die wesentlichen Anforderungen an die Modellierung zur Integration von SOA- in Workflowsysteme beschrieben. Anschließend werden die Erfordernisse an eine durchgehende Transformation dargestellt. An einem Anwendungsfall aus der Verwaltungspraxis wird dargestellt, wie durch die Transformation, unter Einbeziehung von WSDL-Datenkonzepten, die Prozessabläufe automatisiert werden können. Die Arbeit schließt mit einer Zusammenfassung und einem Ausblick.

2 Integration dienstebasierter Konzepte in Workflowsysteme

Der Aufbau dienstebasierter Architekturen im Kontext der EU-DLR [EU06] fordert eine zweckgerichtete Konkretisierung des klassischen SOA-Ansatzes, insbesondere bezogen auf die lose Kopplung der monolithischen Systeme. Da eine dienstebasierte Architektur die vorhandene Heterogenität (durch bestehende Softwarelandschaften) akzeptiert, müssen vorrangig fehlende integrative SOA-Services identifiziert werden [Jo08]. Technische Services bilden damit die Grundlage für eine dynamische Interaktion innerhalb einer SOA-Architektur. Verbindet man diese Dienste durch einen Workflow, so stellt dieser die Geschäftslogik zur Automation dar [Bu07]. Im vorliegenden Fall wird die Entwicklungsumgebung der bflow* Toolbox [Bf09] gewählt, der Ansatz ist jedoch werkzeugunabhängig. Durch einen weitgehenden Aufbau von Workflowsystemen innerhalb einer SOA-Infrastruktur besteht die Möglichkeit, sowohl die Geschäftsprozessmodellierung, als auch lose gekoppelte (Teil-)Automationen durch SOA-Services gemeinsam in dynamische Ablaufsteuerungssysteme zu integrieren. Sofern das gewählte Modellierungswerkzeug - wie im vorliegenden Fall auf Basis der oEPK-Notation [SNZ97] - sowohl UML-Klassendiagramme [Oe06] als auch den korrespondierenden Java-Code bereits (teil-)automativ generiert, wird auf diese Weise eine weitestgehende Ausgangsbasis für die technische Realisierung der ausgewählten Dienste gelegt.

3 Erfordernisse an eine durchgehende Transformation

Betrachtet man die einzelnen Schritte vom Prozessdiagramm zum (teil-)automatisierten Workflow, so kann bereits zu Beginn dieser Kette – auf der Ebene der fachkonzeptionellen Modellierung – diese Transformation durch Anpassung oder Erweiterung der gewählten Modellierungsnotation begünstigt werden. Dies stellt aber auch neue Anforderungen an das eingesetzte Modellierungswerkzeug, sowohl in Bezug auf die Integration in eine SOA-Infrastruktur, als auch die Entwicklung neuer Transformations- und XML-Exportschnittstellen. Zur technischen Umsetzung sind unterschiedliche Transformationskonzepte denkbar. Weit verbreitet ist die Webservice Description Language (WSDL) [Br08], die auch hier als Komponente eingesetzt werden soll.

3.1 Rahmenbedingungen im Modellierungsumfeld

Im konkreten Anwendungsfall wird die Prozessmodellierung mit objektorientierten ereignisgesteuerten Prozessketten (oEPK) durchgeführt. Als Beispielprozess wird, aufgrund seiner Bedeutung für die Wirtschaftsförderung im öffentlichen Sektor, der Prozess zur Gewerbe-Anmeldung gewählt. Das Geschäftsobjekt „Gewerbe-Anmeldung" repräsentiert hierbei das bundeseinheitliche Formular zur Gewerbe-Anmeldung [Gw05], welches sich im Laufe des Verwaltungsverfahrens in unterschiedliche Zustände wandelt. Zum jeweiligen Prozesszustand sind korrespondierende Prozessinformationen wie Attribute, Attributgruppen und Methoden – in angelehnter UML-Notation – bereits mit modelliert. Die jeweiligen Arbeitsschritte werden als Methoden dargestellt. Weitergehende Ausführungen hierzu finden sich bei [*Quelle anonymisiert*].

3.2 Anforderungen an Transformationsparameter

Auf dem Weg vom Geschäftsprozessmodell zum (teil-)automatisierten Workflow müssen die erforderliche Anreicherungen (vgl. Abschnitt 4.2) den jeweils dargestellten Informationen zugeordnet werden. So werden Typinformationen zu Formularfeldern – repräsentiert durch Attribute – benötigt, als auch Ein- und Ausgabewerte für die SOA-Dienste. Die Konkretisierung des Webservices sowie die notwendigen Ein- und Ausgabeparameter werden in der Webservice Definition Language (WSDL) formalisiert. Ist die Zielausgabe ein Webservice, so können die erforderlichen Rumpfbestandteile einer WSDL-Datei unmittelbar als Ergänzung im Geschäftsprozessmodell generiert und durch die Transformation eingeleitet werden.

4 Anwendungsfall eines SOA-Services für die öffentliche Verwaltung

4.1 Medienbrüche als Service-Kandidaten

Als mögliche Service-Kandidaten bieten sich insbesondere Medienbrüche im Prozessablauf an. Müssen beispielsweise erforderliche Prozessinformationen in nicht elektronisch verfügbaren Quellen (Gesetzessammlungen, Register etc.) nachgeschlagen werden, ist der elektronische Prozess regelmäßig unterbrochen. Medienbrüche sind folglich potentielle Service-Kandidaten, die im Rahmen eines Workflows als Webservices angeboten werden können.

Im konkreten Anwendungsfall stellte die „Ermittlung der zuständigen Finanzämter" bisher im Prozessablauf der Gewerbe-Anmeldung einen Medienbruch dar. Die Anschrift der Betriebsstätte kann in größeren Städten dabei theoretisch fünf möglichen Finanzämtern zugeordnet sein. Das zuständige Finanzamt wurde in der Vergangenheit durch manuelle Suche in Straßenverzeichnissen, Onlinediensten oder anderen Quellen ermittelt. Der Projektauftrag im Rahmen einer wissenschaftlichen Begleitforschung [Quelle anonymisiert] war nun, diese Ermittlung durch einen Webservice zu automatisieren. Der Webservice sollte dabei auch gewährleisten, dass dieser auch in Verwaltungsprozessen anderer Akteure im öffentlichen Aufgabenspektrum eingesetzt werden kann (design with reuse [MMY02]).

4.2 Transformation in die Workflowdomain mit WSDL

Wie Geschäftsprozessmodelle für eine weitergehende Nutzung in Workflowsystemen vorbereitet werden können, soll nachfolgend beispielhaft dargestellt werden. Die Ermittlung des zuständigen Finanzamtes wurde als Medienbruch und damit als möglicher Webservice identifiziert. Der Webservice ermittelt das Finanzamt durch Abfrage einer vorhandenen Straßentabelle mit zugeordneten Finanzämtern (Abbildung 1). Die Anschrift der Betriebsstätte ergibt sich dabei aus dem Formular bzw. den bei der Anmeldung eingegebenen Grunddaten des Dienstleisters.

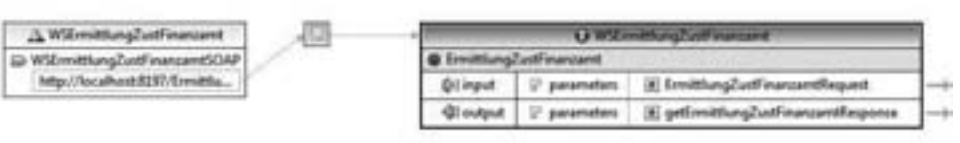


Abbildung 1: Webservice zur Ermittlung des zuständigen Finanzamtes

Durch Definition von Attributsparametern können Bestandteile der WSDL-Datei bereits im oEPK-Modell, dass die fachkonzeptionelle Ausgangsbasis für die Visualisierung der Verwaltungsprozesse legt, definiert werden (Abbildung 2). Hierzu zählen Typdefinitionen wie „String“ für Textfelder, als auch komplexe Klassentypen wie „Anschrift“ oder „Person“, die den Formularfeldern zugeordnet werden.

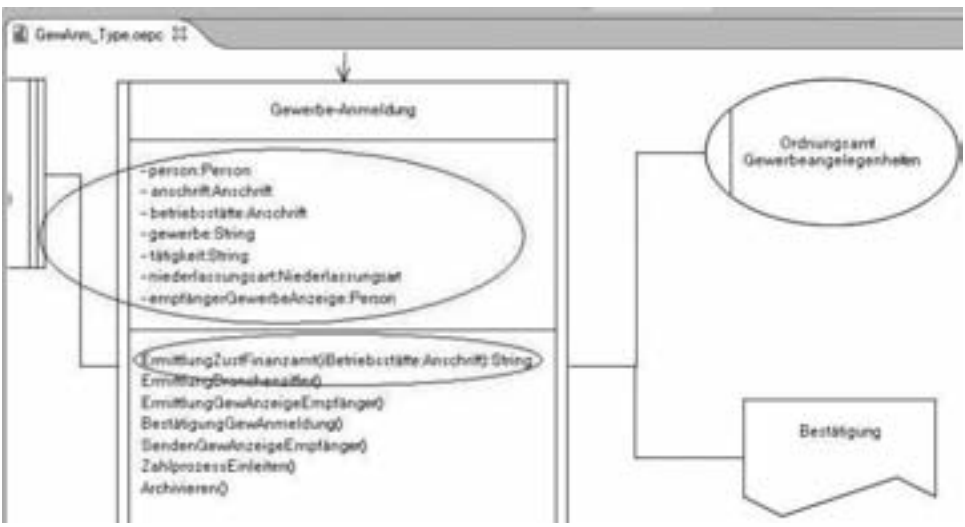


Abbildung 2: Erweiterung der oEPK-Notation um Attributstypen und Parameter

Die Anreicherung der Methoden beinhaltet die Angabe der ein- und ausgehenden Werte mit Hilfe eines sog. In-Parameters, dessen Typ sowie der Ausgabe-Parameter. Dargestellt ist dies in obiger Abbildung durch: Ermittlung *ZustFinanzamt(iBetriebsstätte: An*schrift):String. Der Eingabewert ist die Betriebsstätte vom Typ Anschrift (*betriebsstät-*

te:Anschrift), der Ausgabewert ist vom Typ String. Damit kann für den Webservice der Wert „*Finanzamt Düsseldorf-Süd*“ - also eine Textinformation - zurück geliefert werden. Abbildung 3 verdeutlicht die Verbindung der oEPK-Attributs- und Methodenergänzungen zu den Element-Definitionen im WSDL-Dokument.

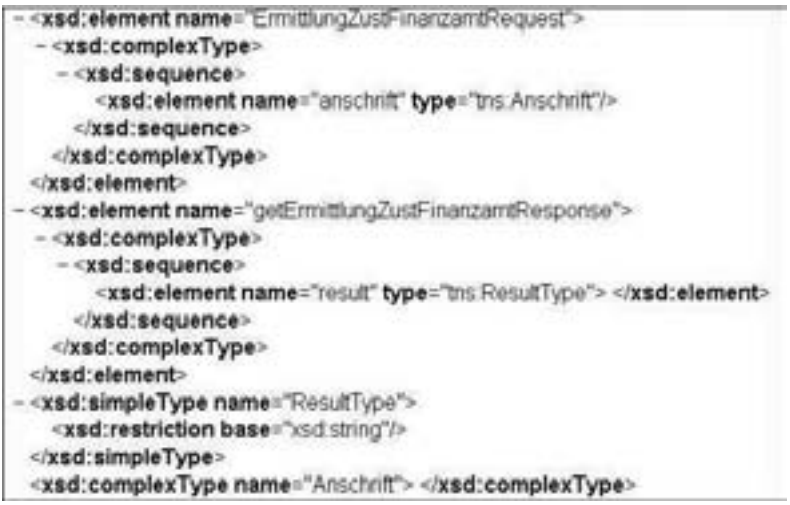


Abbildung 3: WSDL Definition Elemente und Typen (Ausschnitt)

5 Zusammenfassung und weiterer Forschungsbedarf

Der aufgezeigte Ansatz, Transformationskonzepte bereits auf der Ebene der fachkonzeptionellen Modellierung mit zu berücksichtigen, kann den Aufbau und die Integration von SOA-Infrastrukturen und Workflowsystemen wirksam unterstützen. Erste Ergebnisse aus der prototypischen Implementierung sind vielversprechend, gleichwohl aber noch singulär, so dass hier weiteren Untersuchungen folgen, um die Erkenntnisse zu festigen und zu validieren.

Die dargestellte Erweiterung von fachlicher und technischer Modellierung systematisiert die Identifikation und (Teil-)Automation von SOA-Diensten und zeigt Möglichkeiten auf, wie mit erweiterten Transformations- und XML-Exportschnittstellen eine „Brücke“ zwischen fachlichen Geschäftsprozessen und dynamischen Workflowsystemen implementiert werden kann. Derzeit wird der gezeigte Ansatz im Rahmen eines Forschungsprojektes unter Beteiligung von zwei Forschungseinrichtungen methodisch abgesichert und mit Partnern aus Wirtschaft und Verwaltungen unterschiedlicher Größe geprüft und weiter entwickelt [*Quellen anonymisiert*].

Perspektivisch könnten die WSDL-Konzepte auch in der Business Process Execution Language (BPEL) als nachfolgende „Transformationsbrücke“ zum Workflow genutzt werden. Entsprechende konzeptionelle Überlegungen sind in Vorbereitung. Dabei wird mit untersucht, inwieweit die bflow*Toolbox auch hier zum Einsatz kommen kann.

Literatur

[Bf09] Bflow* Toolbox (2009): http://bflow.org, zuletzt geprüft am 25.10.2010.

[Br08] Brüssau, K. (2008): Eclipse Web tools Platform – Java EE, Webanwendungen und Web Services mit WTP, S 120 ff. Unterhaching.

[Bu07] Burbiel, H. (2007): SOA & Webservices in der Praxis, S. 481 ff. Franzis Verlag. Poing

[DOL08] Deutschland-Online-Vorhaben. IT-Umsetzung der Europäischen Dienstleistungsrichtlinie. Projektbericht (Stand: 24.09.2008). http://www.deutschland-online.de/DOL_Internet/broker.jsp?uMen=90040b06-56c2-3021-688d-941ac0c2f214, zuletzt geprüft am 25.10.2010.

[EU06] Europäisches Parlament und Europäischer Rat (2006): Richtlinie 2006/123/EG über Dienstleistungen im Binnenmarkt (EU-Dienstleistungsrichtlinie) vom 12.12.2006, in: Amtsblatt der Europäischen Union, Brüssel.

[Gw05] Gewerbeordnung des Bundes (GewO) (2005): http://www.gesetzesweb.de/GewO.html, zuletzt geprüft am 25.10.2010.

[Jo08] Josuttis, N. (2008): SOA in der Praxis. System-Design für verteilte Geschäftsprozesse. Heidelberg.

[MMY02] Mili, H., Mili, A., Yacoub, S., Addy, E. (2002); Reuse-Based Software Engineering. Techniques, Organization, and Controls, John Wiley & Sons., New York

[Oe06] Osterreich, B. (2006): Analyse und Design mit UML 2.1 – Objektorientierte Softwareentwicklung. München.

[SNZ97] Scheer, A.W.; Nüttgens, M.; Zimmermann, V. (1997): Objektorientierte Ereignisgesteuerte Prozeßketten (oEPK) – Methode und Anwendung. Institut für Wirtschaftsinformatik, Heft 141, Universität des Saarlandes. Saarbrücken.

Konzeption und Implementierung einer konfigurierbaren Komponente zum Modelldatenaustausch am Beispiel der bflow* Toolbox

Lucas Wirmer, Nick Gehrke, Markus Nüttgens, Martina Peris

Fakultät Wirtschafts- und Sozialwissenschaften, Wirtschaftsinformatik
Universität Hamburg
Max-Brauer-Allee 60
22765 Hamburg
[vorname].[nachname]@wiso.uni-hamburg.de

Abstract: Der vorliegende Beitrag entwickelt ein Konzept zur interaktiven Transformation von Formaten zur Abbildung von Geschäftsprozessmodellen. Die Validierung des Konzepts erfolgt in der Form einer prototypischen Implementierung. Hierdurch werden exemplarisch konfigurierbare Modelltransformationen zwischen den Formaten der bflow-Toolbox (einer Open Source Modellierungssoftware) und marktgängigen Formaten wie der ARIS Markup Language (AML) und der Microsoft Visio Formate (VDX) unterstützt. Die Relevanz des Konzepts und der prototypischen Entwicklung zeigt sich im Wesentlichen in zwei Aspekten. Einerseits wird dem Nutzer ermöglicht, in den Transformationsprozess einzugreifen und Modellmanipulationen vorzunehmen. Diese Eingriffsmöglichkeit soll ein erhöhtes Nutzenpotenzial für den Endnutzer erschließen. Andererseits kann eine Erweiterung um Transformationsvorschriften und Formate ohne Eingriff in den originären Quellcode erfolgen. Diese Flexibilität und Konfigurierbarkeit soll die Produktivität und Akzeptanz in der praktischen Anwendung maßgeblich erhöhen.

1 Einleitung

Seit der Entstehung in den 80er Jahren hat das Leitbild der (Geschäfts-) Prozessorientierung und damit auch der Markt für sog. Business Process Management (BPM)-Werkzeuge an Bedeutung gewonnen [BMW09, S.2]. Im Laufe der Zeit wurden zahlreiche Methoden zur Modellierung von Geschäftsprozessen propagiert und in verschiedenen Softwaresystemen implementiert. Nutzerbefragungen belegen jedoch immer wieder die Bedeutung und bislang ungelöste Problematik des Austausches von Modelldaten zwischen verschiedenen BPM-Werkzeugen [De05]. Bis heute konnte sich trotz erster Initiativen kein neutrales Standardformat am Markt etablieren [MNN05]. Dem praktischen Problem der Interoperabilität im BPM-Kontext liegt das theoretische Konzept der Modelltransformation zugrunde [CH06]. Ein Geschäftsprozess in einem bestimmten Format kann abstrakt als ein Modell aufgefasst werden, das in eine andere Form transformiert werden soll.

Gegenstand dieses Beitrages ist die Konzeption und prototypische Implementierung einer konfigurierbaren Komponente für die Transformation von BPM-Modellen. Die

Eigenschaft der Konfigurierbarkeit stellt das Alleinstellungsmerkmal des hier entwickelten Ansatzes dar und soll den Anwender in die Lage versetzen, die Transformation an individuelle Anforderungen anzupassen. Eine Validierung des Konzeptes erfolgt in Form einer prototypischen Implementierung als Plugin für die bflow*Toolbox (einer Open Source Modellierungssoftware).

2 Kontext und Motivation

Ein *Metamodell* ist ein Modell, welches zur Beschreibung anderer Modelle benutzt werden kann[AK03, Be06]. Unter einer *Modelltransformation* versteht man die „Überführung von einem Quellmodell in ein Zielmodell in Abhängigkeit von einer Transformationsdefinition“ [KKF06]. Es existiert eine Vielzahl von Ansätzen zur Durchführung von Modelltransformationen und deren Evaluation [CH06, Mv06]. Dennoch ist keine allgemeingültige Aussage über den besten Ansatz möglich, dies hängt stark vom Einsatzkontext ab.

Gleichermaßen existieren zahlreiche Methoden zur Geschäftsprozessmodellierung [vgl. Re09]. Ebenso fragmentiert zeigt sich der Markt für Software zur Geschäftsprozessmodellierung [vgl. Bö07]. Hinzu kommt das wachsende Interesse der Unternehmen Open Source Software einzusetzen [Re06]. Vor dem Kontext der Geschäftsprozessmodellierung wurden zahlreiche Intermediärformate zum Modelldatenaustausch entwickelt und miteinander verglichen [MNN05]. Bisher konnte sich jedoch kein Format durchsetzen. Zur eigentlichen Durchführung der Transformationen werden in der Literatur zwei verschiedene Ansätze verfolgt und diskutiert. Der eher als pragmatisch einzuordnende Ansatz ist die Nutzung von Transformationssprachen, die auf der Ebene der Serialisierung ansetzen, wie z.B. die XSLT [MN04, VZS05]. Zwar lässt sich hierdurch relativ schnell eine funktionierende Transformation erreichen, allerdings wird die mangelnde Eignung für komplexe Transformationen bemängelt. Oft genannte Schwachpunkte sind die schlechte Lesbarkeit und Wartbarkeit des Codes sowie die Performanz bei großen Transformationen [CH03, SK03, KKF06]. Für den vorliegenden Beitrag wurde daher der Schwerpunkt auf eine höhere Abstraktionsstufe gelegt, bei der die Transformationen auf Ebene der Metamodelle beschrieben und damit zunächst unabhängig vom Serialisierungsformat vorgenommen wird [Gr06, MK07].

Das Ausgangsproblem für diesen Beitrag ist die in der Praxis bestehende Herausforderung, Modelldaten aus verschiedenen Werkzeugen und in teilweise unterschiedlichen Formaten zu transformieren. Derzeit existiert keine Lösung, um eine zufriedenstellende Interoperabilität im Kontext der Geschäftsprozessmodellierung zu erreichen. Im Rahmen dieses Beitrags soll daher ein Konzept zum Austausch von Modelldaten zwischen verschiedenen BPM-Werkzeugen entwickelt werden. Eines der Hauptziele dabei ist die Sicherstellung einer größtmöglichen Flexibilität bezüglich der unterstützten Notationen. Dazu wurde die Definition eines intermediären Formats auf Ebene eines Metamodells als sinnvollster Ansatz identifiziert. Es soll daher ein Metamodell entworfen werden, das über die nötige Ausdrucksmächtigkeit zur Abbildung einer Vielzahl von im Kontext der Geschäftsprozessmodellierung verwendeten Konstrukten verfügt. Als wesentliches Merkmal der Abgrenzung und Alleinstellung soll eine Konfigurierbarkeit der Transfor-

mationskomponente dienen. In der Praxis sind zahlreiche Szenarien denkbar, bei denen durch eine solche Konfigurierbarkeit der vom Anwender empfundene Nutzen der Transformationskomponente stark gesteigert werden könnte. Mit Hilfe eines zu entwickelnden Prototyps soll die grundsätzliche Umsetzbarkeit des Konzeptes demonstriert werden.

3 Konzept

Jegliche im BPM-Kontext auftauchenden Modelle können als Form eines Graphen aufgefasst werden können. Hieraus kann gefolgert werden, dass das Metamodell die allgemeine Form eines Graphen repräsentieren, also im Grunde nur die Elemente Knoten und Kanten enthalten sollte. Damit bleibt die Semantik der Modellelemente zunächst unbestimmt. In Abbildung 1 ist der Aufbau des *generischen Metamodells* in Form eines UML-Klassendiagramms dargestellt. Es handelt sich um das Kompositum Pattern [Ga09]. Knoten und Kanten werden durch die Klassen *Node* bzw. *Edge* repräsentiert. Hinzu kommt die Klasse *Model*, die eine Menge von Knoten und Kanten als eine Art Container zusammenfasst und als Basiselement dient. Jede weitere Spezialisierung des intermediären Modells würde zwingend die Ausdrucksmächtigkeit und Flexibilität einschränken und ist daher nicht wünschenswert. Insbesondere ein Vergleich mit dem integrierten Metamodell von Murzek & Kramler [MK07] zeigt, dass weitere Elemente im hier propagierten Modell durch Attributierung auch ohne explizite Erweiterung des Metamodells hinzugefügt werden können.

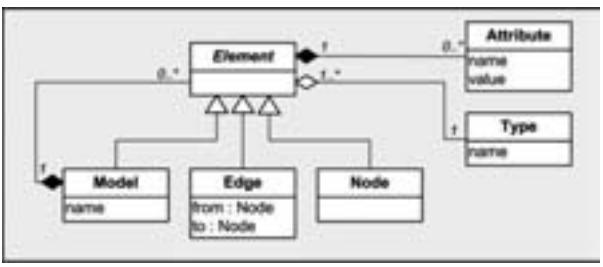


Abbildung 1: Generisches Metamodell

In Abbildung 2 werden die einzelnen Schritte einer Transformation dargestellt. Aus der Anforderung der Konfigurierbarkeit resultiert die Konzeption zweier Modi, dem sogenannten Standard- und dem Expertenmodus. Im Expertenmodus kann an zwei Stellen in die Transformation eingegriffen werden. Im Rahmen der Vorverarbeitung kann der Nutzer ausgewählte Elemente aus dem eingelesenen Modell löschen bzw. bearbeiten. Die zweite Form der Konfiguration ergibt sich bei der Prüfung der Constraints. Dabei wird dem Nutzer bei jeder fehlgeschlagenen Constraint-Prüfung eine vorab definierte Auswahl von Handlungsalternativen präsentiert, wodurch eine individuelle Gestaltung der Transformation ermöglicht wird. Die beiden Schritte des Expertenmodus können unabhängig voneinander gewählt werden. Zur Realisierung sieht das Konzept eine modulartige Architektur der Komponente vor, die entsprechende Eingriffe ermöglicht.

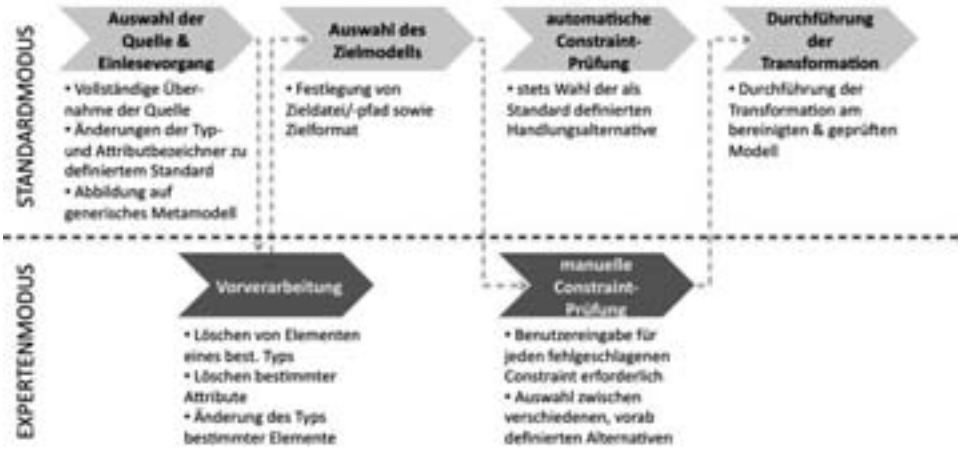


Abbildung 2: Ablauf der Transformation aus konzeptioneller Sicht

4 Prototypische Implementierung und Validierung

Eine wesentliche Rahmenbedingung für die Technologieauswahl ist die geplante Integration des Prototypen in die bflow* Toolbox. Dies führte technologisch dazu, dass die Komponente als ein Plugin für die Eclipse-Plattform entwickelt wurde. Zur technischen Realisierung der Modelltransformationen stehen verschiedene Ansätze und Frameworks zur Wahl. Den Vorzug erhielt *openArchitectureWare* (oAW) [Ha07].

Abbildung 3 zeigt die Kernelemente der Softwarekomponente mit ihren wichtigsten Methoden und Assoziationen. Zur Erhöhung der Verständlichkeit werden ausgewählte relevante Inhalte dargestellt. Das Package *converter.internal* enthält die eigentliche Funktionalität. Als zentrale Konfigurationsdatei dient die *registry.xml*, über die sich sämtliche implementierten Modellformate bei der Anwendung registrieren. In der aktuellen Version unterstützt der Prototyp die Transformation von EPK-Modellen in den Formaten der Werkzeuge ARIS Toolset, bflow* Toolbox und Microsoft Visio. Dabei wurden Abbildungen für ca. 15 der wichtigsten Elementtypen aus den einzelnen Metamodellen definiert. Zudem werden bei der Transformation als Attribute die Beschriftungen und wesentliche Grafikinformationen übernommen, was den Wiedererkennungswert der transformierten Prozessmodelle durch den Anwender deutlich erhöht.

Der entwickelte Prototyp dient der Demonstration der Umsetzbarkeit des hier entwickelten Konzepts. Trotz der erfolgreich umgesetzten Grundfunktionalität, existieren Optimierungspotenziale in den Bereichen Ausnahmebehandlung durch die Benutzeroberfläche und Behebung verschiedener identifizierte Fehler. Auch wenn sich rein subjektiv keine nennenswerten Wartezeiten bei der Durchführung der Transformationen ergaben, sollte die Performanz vor einer Weiterentwicklung der Komponente näher untersucht werden. Ansatzpunkte für eine Weiterentwicklung bieten die Formatunterstützung, der Transformationsumfang sowie die Modell-Vorverarbeitung.

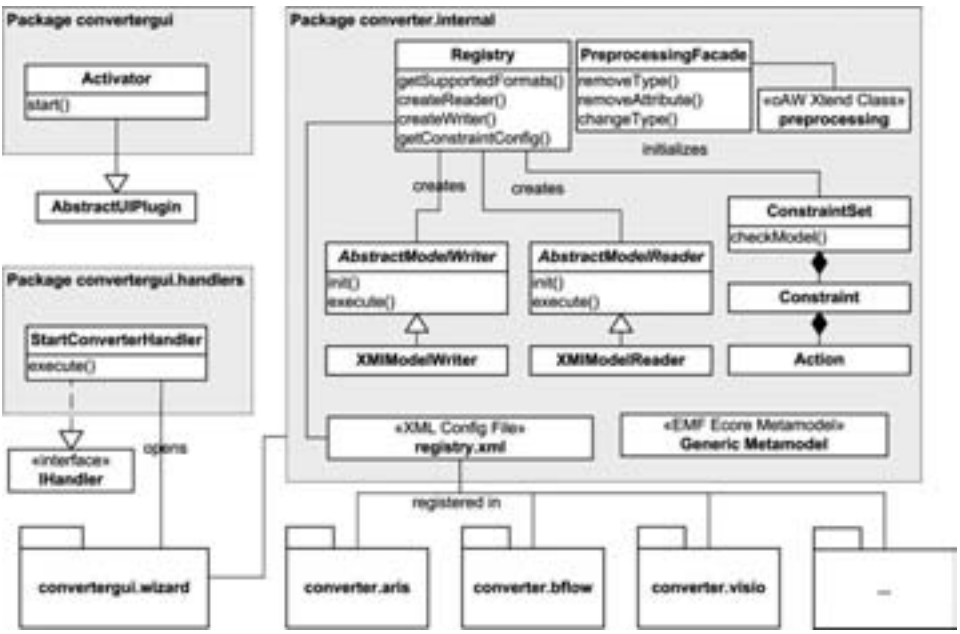


Abbildung 3: Architektur der prototypischen Umsetzung

5 Zusammenfassung

Im Rahmen dieses Beitrags wurde der Grundstein für eine konfigurierbare Komponente zur Durchführung von Modelltransformationen im BPM gelegt. Dabei wurde der Schwerpunkt auf eine höhere Abstraktionsstufe gelegt, bei der die Transformationen auf Ebene der Metamodelle beschrieben und mit Hilfe spezieller Frameworks durchgeführt werden. Das entwickelte Konzept zeichnet sich durch die Verwendung eines generischen Metamodells aus, das Prozessmodelle in einer mit Attributen angereicherten Form eines allgemeinen Graphen repräsentiert. Eine Konfigurierbarkeit des Transformationsprozesses ist an zwei Stellen vorgesehen (Experten-Modus).

Bedingt durch das Ziel einer prototypischen Entwicklung existieren noch verschiedene Optimierungspotenziale, die einem Praxiseinsatz im Wege stehen. Diese beziehen sich vornehmlich auf die Behandlung von Ausnahmezuständen und die Benutzbarkeit der grafischen Oberfläche. Die Software könnte dennoch als unabhängiger Vermittler zwischen der Vielzahl existierender Modellierungsformate im BPM-Kontext dienen. Gewisse Einschränkungen des Konzepts, wie der hohe Abstraktionsgrad des generischen Metamodells führen zu einer aufwendigen Plausibilitätsprüfung von Modellinstanzen und damit zu einer großen Zahl benötigter Constraints. Es existieren verschiedene Ansatzpunkte für eine Weiterentwicklung des Prototyps. In erster Linie bietet sich eine Implementierung der Unterstützung weiterer Formate an, verbunden mit einer Erhöhung des Umfangs der transformierten Modellelemente. Zusätzlich könnte die Übertragung des Konzepts auf andere Anwendungsdomänen Inhalt weiterer Arbeiten sein.

Literaturverzeichnis

[AK03] Atkinson, C.; Kühne, T.: Model-Driven Development: A Metamodeling Foundation. In IEEE Software, 2003, 20; S. 36–41.

[Be06] Berio, G.: UEML 1.0 and UEML 2.0: Benefits, Problems and Comparison. In (Bussler, C.; Haller, A. Hrsg.): Business process management workshops. BPM 2005 international workshops BPI, BPD, ENEI, BPRM, WSCOBPM, BPS, Nancy, France, September 5, 2005; revised selected papers. Springer, Berlin, 2006; S. 245–256.

[BMW09] Becker, J.; Mathas, C.; Winkelmann, A.: Geschäftsprozessmanagement. Springer, Berlin, 2009.

[Bö07] Böhn, M. et al.: Prozessmodellierungswerkzeuge. 14 Systeme für Dokumentation, Entwurf, Simulation und Analyse im Vergleich; eine Studie des Business Application Research Centers. Oxygon-Verl., München, 2007.

[CH03] Czarnecki, K.; Helsen, S.: Classification of Model Transformation Approaches. http://www.swen.uwaterloo.ca/~kczarnec/ECE750T7/czarnecki_helsen.pdf, 15.09.2009.

[CH06] Czarnecki, K.; Helsen, S.: Feature-based survey of model transformation approaches. In IBM Systems Journal, 2006, 45; S. 621–645.

[De05] BPM 2005. Market Milestone Report. A Delphi Group White Paper, Boston, Mass., 2005.

[Ga09] Gamma, E.. et. al.: Entwurfsmuster: Elemente wiederverwendbarer objektorientierter Software. Addison Wesley, München, 2009.

[Gr06] Grangel, R. et al.: Exchange of Business Process Models Using the POP* Meta-model. In (Bussler, C.; Haller, A. Hrsg.): Business process management workshops. BPM 2005 international workshops BPI, BPD, ENEI, BPRM, WSCOBPM, BPS, Nancy, France, September 5, 2005; revised selected papers. Springer, Berlin, 2006; S. 233–244.

[Ha07] Haase, A. et al.: Introduction to openArchitectureWare 4.1.2. http://www.dsmforum.org/events/MDD-TIF07/oAW.pdf, 15.09.2009.

[KKF06] Kern, H.; Kühne, S.; Fötsch, D.: Merkmale und Werkzeugunterstützung für Modelltransformationen im Kontext modellgetriebener Softwareentwicklung. In (Fähnrich, K.-P.; Kühne, S.; Speck, A.; Wagner, J. Hrsg.): Integration betrieblicher Informationssysteme. Problemanalysen und Lösungsansätze des Model-Driven-Integration-Engineering. LIV, Leipzig, 2006.

[MK07] Murzek, M.; Kramler, G.: The Model Morphing Approach - Horizontal Transformations between Business Process Models, Tampere, Finland, 2007.

[MN04] Mendling, J.; Nüttgens, M.: Transformation of ARIS Markup Language to EPML. In (Nüttgens, M.; Rump, F. J. Hrsg.): Proceedings of the 3rd GI Workshop on Event-Driven Process Chains (EPK 2004), Luxembourg, October 2004.

[MNN05] Mendling, J.; Neumann, G.; Nüttgens, M.: A Comparison of XML Interchange Formats for Business Process Modelling. In (Fischer, L. Hrsg.): Workflow Handbook 2005. Future Strategies Inc, Lighthouse Point, Florida, 2005; S. 185–198.

[Mv06] Mens, T.; van Gorp, P.: A Taxonomy of Model Transformation. Proceedings of the International Workshop on Graph and Model Transformation (GraMoT 2005). In Electronic Notes in Theoretical Computer Science, 2006, 152; S. 125–142.

[Re06] Renner, T. et al.: Open Source Software: Einsatzpotenziale und Wirtschaftlichkeit. Fraunhofer IAO, Stuttgart, 2006.

[Re09] Recker, J. et al.: Business Process Modelling - A Comparative Analysis. In Journal of the Association for Information Systems, 2009, 10; S. 333–363.

[SK03] Shane Sendall undWojtek Kozaczynski. Model Transformation: The Heart and Soul of Model-Driven Software Development. IEEE Software, 20(5):42–45, 2003.

[VZS05] Vanderhaeghen, D.; Zang, S.; Scheer, A.-W.: Interorganisationales Geschäftsprozessmanagement durch Modelltransformation, Saarbrücken, Februar 2005.

Workshop
Web Science
–Social Computing Applications–

Sören Auer, Claudia Müller-Birn, Steffen Staab

Workshop Web Science (WSW 2010) –Social Computing Applications–

Sören Auer[(1)], Claudia Müller-Birn[(2)], Steffen Staab[(3)]

[(1)]University of Leipzig
AKSW/Institute for Computer Science
auer@informatik.uni-leipzig.de
[(2)]Carnegie Mellon University
Institute for Software Research
clmb@cs.cmu.edu
[(3)]University of Koblenz-Landau
Institute for Web Science and Technologies
staab@uni-koblenz.de

Preamble

Web Science is often referred to as the "science of decentralized information systems" [BLHH+06a], [Shn07]. *[It] [...] aims to map how decentralized information structures can serve [...] scientific, representational and communicational requirements, and to produce designs and design principles governing such structures.* [BLHH+06b] While technologies such as Semantic Web, Web Services, and Cloud Computing are germane to the broad proliferation of Web technologies, we also need to understand the human side of the Web, in order to retain its usefulness and benefit to people. This is at the center of attention of Web Science and includes, in addition to the aforementioned technological approaches, research related to online communities, information diffusion, Web governance, global network structures beyond the individual communities on the Web, and incentive and monetization systems [HSH+08].

Because the Web itself is socially embedded, a particular focus of this first GI workshop on Web Science is Social Computing applications, such as Wikipedia, Facebook, or del.icio.us, which are analyzed with regard to the complex interdependencies between constraints imposed by the technical system as well as with regard to their use in social actions and interactions under varying social contexts. Understanding the dynamics and evolution of these systems as they depend on inherent social and informational structures is of particular interest, because it is the dynamics of such communities that determines their final success or failure. In order to analyze Social Computing applications, the interplay between aspects of Computer Science, including computational network analysis, visualization techniques, graph theoretic models, semantic web techniques, and machine learning techniques, with social science and psychology is required.

In this interdisciplinary workshop we aim to bridge the gap between paradigms and encourage interdisciplinary collaborations as well as advance and deepen our understanding of web science. Joint efforts are needed to take advantage of the state-of-the-art research from multiple disciplines, such as Computer Science, information systems, sociology, and psychology. Moreover, this workshop provides a platform for researchers and practitioners to exchange preliminary results, new concepts and methodologies in this area.

For this first GI Workshop on Web Science we selected four contributions to be presented.

Laura Hollink and Mark van Assem are specifically interested in on how people search or navigate in the so-called "Culture Web". In their contribution *"Estimating the relevance of search results in the Culture-Web: a study of semantic distance measures"*, they present a recent study in which the relevance of search results is evaluated based on semantic distance measures. Wolfgang Orthuber and Stefan Dietze propose in their contribution *"Proposal: Standardized Vectorial Resource Descriptors on the Web"* a new similarity-based representation of resources, so-called Vectorial Resource Descriptors (VRDs). Existing approaches such as Linked Data could be augmented by this technique. Christian Thurau and Christian Bauckhage study the temporal evolution of different player roles in a multiplayer online game. In their contribution *"Group Evolution Patterns in World of Warcraft"* they specifically investigate different formation patterns. Robert Tolksdorf argues in his contribution *"Towards Quality Measures for Web Network Extraction"* how the quality of data that are obtained by information extraction techniques can be measured. In an exemplary case study, he shows how such a quality measure can be used.

References

[BLHH+06a] Tim Berners-Lee, Wendy Hall, James Hendler, Nigel Shadbolt, and Daniel J. Weitzner. Creating a Science of the Web. *Science*, 313(5788):769–771, 2006.

[BLHH+06b] Tim Berners-Lee, Wendy Hall, James A. Hendler, Kieron O'Hara, Nigel Shadbolt, and Daniel J. Weitzner. A Framework for Web Science. *Web Science*, 1(1):1–130, 2006.

[HSH+08] James Hendler, Nigel Shadbolt, Wendy Hall, Tim Berners-Lee, and Daniel Weitzner. Web science: an interdisciplinary approach to understanding the web. *Commun. ACM*, 51(7):60–69, 2008.

[Shn07] Ben Shneiderman. Web Science: A Provocative Invitation to Computer Science. *Communication of the ACM*, 50(6):25–27, June 2007.

Towards Quality Measures for Web Network Extraction

Robert Tolksdorf
Networked Information Systems, Freie Universität Berlin, www.ag-nbi.de
tolk@ag-nbi.de

Abstract: Within the Web information networks of various kinds are of interest for specific applications. These networks evolve in a decentralised manner. As such, they represent alternatives to centralised information sources such as databases.

While crawlers can try to extract these networks from the Web, there is currently no systematic work on how to measure the quality of this extraction. Quality measures are of central importance to justify efforts and investments in Web network mining as an alternative to licence models of centralised information sources.

In this paper we present a use case of such an interesting information network, that of the art market. We then present a simple model of a quantifiable quality measure of network extracted from the Web.

1 Introduction

[XC08] reports at the end of an article analysing networks of terrorist and criminal activities: "Please also note that care is needed when interpreting these findings. Because dark networks are covert and largely unknown, hidden links may be missing in the elicited networks". The authors seem to be unsure that the extraction of a hidden network provided to them is complete enough to generate insights based on statistical analysis. They seem to be concerned that they worked on a subset of data whose quality was not evaluated and that therefore their analysis could be questioned.

Necessary is some *quality measure* that helps to assess the extent to which an analysed network is a relevant subset of the network to study and – in turn – how valid the results of that analysis can be.

2 Use-Cases of Web Network Extraction

For various applications areas, the Web contains topic specific networks. These are spanned as an overlay on the usual Web page and link structure. To discover them, one needs to crawl the Web and to analyse whether a page should be considered a node in the overlay network and whether a link represents a relation in the overlay.

As an example application area we chose the *art market*. Information about it is a quite loosely coupled and distributed on the Web. Figure 1 shows a simple model of the participating roles and their relations. At the core is the artwork and the artist who has created

it. Around that we find the three common levels of the art market with galleries, auction houses and art dealers. “Consumers” of artworks are private and public collections which may show them in exhibitions.

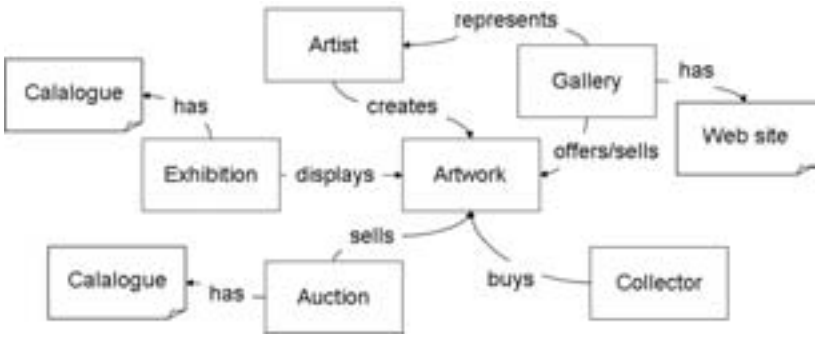


Figure 1: An example network on photography collections

A crawler could try to extract the network of galleries and represented artists from the Web. In order to do so, it has to classify whether a Web page represents a gallery (ie. is its homepage) and whether links in the site can be considered the “represents” relation. In addition, one has to determine the nodes that are considered as artists. These might be simply their names or their homepages, with the later being the more difficult case since the descriptions of artists on the galleries’ sites have to be mapped to the artists homepages.

The number of artists can be estimated as several millions and the number of artworks about two to three dimensions higher. Each artwork appears at least one time the art market, namely when it is sold by a gallery to a collector.

Other cultural areas, eg. cinematography have managed to generate more or less complete archives on their field. An example is the Internet Movie Database IMDB (*http://www.imdb.com/*) which collects information about films, actors, directors etc. with a quite similar data model to the one shown above. For the art market, there is nothing comparable, so the mentioned crawler could collect valuable information.

Networks are subject to various studies, for an overview see [New03]. Across various disciplines such as biology, medicine, computer science, earth sciences, logistics, sociology and lots more, the study of the structure of network structures and their dynamics leads to additional insights with most practical consequences. A good example is the detection of “super spreaders” in infection networks that should be isolated in case of a pandemia, or the analysis on traffic networks answering which airports should then be closed.

Quite often, these networks exhibit specific statistical distributions on which nodes are connected within them. Scale free networks [BB03] are found in various areas like traffic networks or the Web. Here, the likelihood of a node have a huge number of links to others can be described by a single relation, the so-called power law. In essence, it means at only very few nodes have a very high linkage and that the huge majority populates the *long tail* of this nodes with a low linkage. The *preferential attachment* characteristic models for example, that popular Web sites attract even more links – the rich get richer.

Considering the art market, we establish the hypothesis that it exhibits similar characteristics that can be used to provide additional information. For example, we might expect that artists whose works are offered often in auctions will be offered also in future auctions frequently. Or, we are certain that there is a huge number of artists with low presence in collections while a very few ones reach a high probability to be collected.

As a small experiment we used information from [Aig96] to perform a visualisation on which photography artists are part of the seven of eight European collections on photography art listed in the referenced journal. We scanned the pages containing the lists of artists for each collection (one could not be scanned in a satisfactory manner) and did some manual data cleansing. Then, this text data was given to a script that generated a representation of the network in GraphML [BEH+01]. A visualisation tool then was used to generate the depiction of the network as in figure 2. At the centre of the seven clusters are the collection, the other nodes are artists and links between them represent that an artwork created by the artist is displayed in the collection.

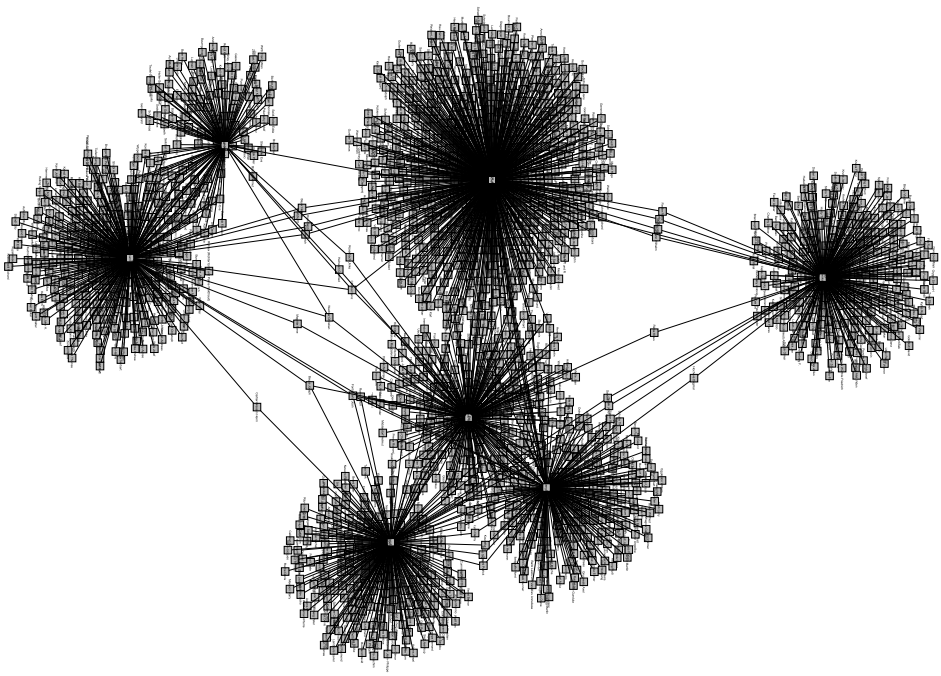


Figure 2: An example network on photography collections

Of the about 1600 artists, 120 were displayed in two or more of the seven collections. About 93% of the artists are collected only once. As indicated by the example, network visualisations exhibit an own aesthetic quality. Considering the amount of possibilities for network visualisation (sites like *VisualComplexity* [1] illustrate this), we could consider self-referential art on the art market based on the data collected.

[1] *http://www.visualcomplexity.com/*

There are centralised and decentralised ways get information about the art market:

- We use a database that is maintained manually and centrally and which offers a measurable set of information on arts in our case.
- We use an automated Web crawler to extract the decentralised information from relevant Web sites, like galleries, auctioneers and collections in our use-case.

If we were able to have an automated approach that leads to a result of *measurable quality*, we could also establish a cost/benefit model for our approach.

For our use-case of art information, there are databases like the *The Union List of Artist Names (ULAN)* which contains around 375,000 names and biographical and bibliographic information about artists and architects (*http://www.getty.edu/research/conducting_research/vocabularies/ulan/*). It has licence costs in the low 4-digit range per year. Assuming that we want could reach a information comparably quality by crawling, one could assume about half the costs per year for running a crawler and a Web site on a rented server. To make that case, we need a quantifiable measure on quality.

3 Quality Measures for Web Network Extraction

In the field of information retrieval, two quality measures are central [BB99]:

- *Precision* represents the internal quality of the result set. It expresses how many wrong results lower the quality and is calculated by dividing the amount of relevant documents found by the number of documents found.
- *Recall* expresses the quality of finding relevant results calculated by the relation of the number relevant found documents to the total number of existing relevant documents.

We use this common notion as a basis to define a similar measure for the "retrieval" of a subgraph from the Web that is relevant for a specific domain of interest.

3.1 Recall and Precision of Network Extraction

The Web is a network W consisting of pages as nodes P_W and links L_W. For a given domain D – art in our example case –, the network of relevant information is $W_D = (P_D, L_D)$. For any other network $W_X = (P_X, L_X)$, we are interested in quantifying, how good it captures W_D. Assuming that W_D does exist[2], we can define the following measures.

[2]In [TH09] we have argued that the notion of a relevance set is problematic since any user has a subjective expectation of an information search, thus no objective relevance set can exist. In our can, we do have databases such as ULAN as subsets of that relevance set.

- The precision measure Pprec of W_X relative to W_D and relating to the pages contained is the proportion of how many pages in P_X are also in P_D to the size of P_X: $\#\{p|p \in P_D \wedge p \in P_X\}/\#P_X$.
- The recall measure Prec of W_X relative to W_D and relating to the pages contained is the proportion of how many pages from P_X are in P_D: $\#\{p|p \in P_D \wedge p \in P_X\}/\#P_D$.
- The precision measure Lprec of W_X relative to W_D and relating to the links contained is the proportion of how many links in L_X are also in L_D to the size of L_X: $\#\{l|l \in L_D \wedge l \in L_X\}/\#L_X$.
- The plain recall measure Lrec_p of W_X relative to W_D and relating to the links contained is the proportion of how many links from L_X are in L_D: $\#\{l|l \in L_D \wedge l \in L_X\}/\#L_D$.

The discovery of links depends on the extraction of their source pages. The reason to miss a link can be its wrong classification or it is a followup error to wrongly classify the page in which it originates. The above recall measure on links would therefore take a double penalty which does not reflect the quality of link extraction. Therefore it has to be extended to only consider links from pages that are in P_X. We write source(l) to denote the page where the Web link l starts, ie. its source anchor.

$$\text{Lrec} = \#\{l|l \in L_D \wedge l \in L_X\}/\#\{l|l \in L_D \wedge \text{source}(l) \in P_X\}$$

3.2 Measuring the Quality

In information retrieval, recall and precision can be combined into a single value, the F-measure [vR79]:

$$F_\beta = \frac{(1+\beta^2) \cdot \text{precision} \cdot \text{recall}}{\beta^2 \cdot \text{precision} + \text{recall}}$$

β can be used to further weight the importance of them. The common value of 1 yields the weighted harmonic mean of both factors, thus being closer to the worse performance.

For our purposes, we have to extend the measure to account for both page- and link-quality. While we keep a β to adjust whether the focus is on precision or recall, we also have to express how much the quality of links is important in relation to the quality of pages extracted. We do so with a second weight γ. We calculate a the weighted arithmetic mean on link and pages qualities and use these to determine our F-measure on the quality of a retrieved subgraph:

$$F_{\beta,\gamma} = \frac{(1+\beta^2) \cdot \frac{\text{Pprec}+\gamma^2\text{Lprec}}{1+\gamma^2} \cdot \frac{\text{Prec}+\gamma^2\text{Lrec}}{1+\gamma^2}}{\beta^2 \cdot \frac{\text{Pprec}+\gamma^2\text{Lprec}}{1+\gamma^2} + \frac{\text{Prec}+\gamma^2\text{Lrec}}{1+\gamma^2}}$$

4 Conclusion

In this paper we have presented a use case of Web network extraction and argued that quality measures for the results are necessary to enable cost/benefit analysis. We then have described a first model of such a measure and given a rationale.

With this as a first step, we see two main future work items. First, the model has to be calibrated and validated. This can be done in the small with a tool and manual analysis. In a second step, one would have to generate a community effort like the TREC competitions [VH05] in which Web graph miners would compete for best results.

References

[Aig96] Carl Aigner, editor. *EIKON - International Zeitschrift für Photographie & Medienkunst*. Number 18/19/20. Turia & Kant, Wien, 1996. Collected Photography, Private und öffentliche europäische Photokunstsammlungen.

[BB99] Michael W. Berry and Murray Browne. *Understanding search engines: mathematical modeling and text retrieval*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 1999.

[BB03] A.-L. Barabási and E. Bonabeau. Scale-Free Networks. *Scientific American*, (288):60–69, 2003. http://www.nd.edu/ networks/Publication Categories/01 Review Articles/ScaleFree_Scientific Ameri 288, 60-69 (2003).pdf.

[BEH+01] U. Brandes, M. Eiglsperger, I. Herman, M. Himsolt, and M.S. Marshall. GraphML Progress Report: Structural Layer Proposal. In *Proc. 9th Intl. Symp. Graph Drawing (GD '01)*, number 2265 in LNCS, pages 501–512. Springer, 2001.

[New03] M. E. J. Newman. The structure and function of complex networks. *SIAM Review*, 45:167–256, 2003. http://aps.arxiv.org/abs/cond-mat/0303516/.

[TH09] Robert Tolksdorf and Thomas Hoppe. Quantitative Analyse von Ergebnissen Semantischer Suche. In *11. Kongress zum IT-gesttzten Wissensmanagement in Unternehmen und Organisationen, KnowTech 2009*, 2009.

[VH05] Ellen M. Voorhees and Donna K. Harman. *TREC: Experiment and Evaluation in Information Retrieval (Digital Libraries and Electronic Publishing)*. The MIT Press, 2005.

[vR79] C. J. van Rijsbergen. *Information Retrieval*. Butterworths, London, 2 edition, 1979.

[XC08] Jennifer Xu and Hsinchun Chen. The topology of dark networks. *Commun. ACM*, 51(10):58–65, 2008.

Group Evolution Patterns in World of Warcraft®

Christian Thurau[1] and Christian Bauckhage[1,2]
[1]Fraunhofer IAIS, 53754 St. Augustin, Germay
[2]B-IT, University of Bonn, 53113, Bonn Germany

Abstract: We study the temporal evolution of different types of guilds in the massively multiplayer online role playing game WORLD OF WARCRAFT® . Given a large corpus of observations of online activities of players, we apply convex-hull non-negative matrix factorization to cluster our data of about 1,400,000 guilds into well interpretable prototypes. Given these prototypes, we analyze guild formation patterns on American and European servers. We find growth patterns according to power laws that result in similar structures in both cases.

1 Introduction and Related Work

In this paper, we present first results of a comprehensive study of patterns of group formation processes in WORLD OF WARCRAFT® . As of this writing, this game has attracted more than 12,000,000 active players. Because of its immense fan-base, it offers unique possibilities for the study of human interaction on a very large scale. In our work, we evaluate 192,000,000 observations of in-game player activity. Based on this data, we provide an interpretable categorization of different types of guilds, visualize general guild evolution patterns, and compare the development of US and EU based guilds.

Web-scale game data mining is of growing interest to academia and industry alike. In earlier work, we analyzed in-game recordings of QUAKE II® to reproduce individual human behavior. Using machine learning techniques, we were able to create reactive, tactical, or strategic behaviors [BTS03, TBS04]. In later work, Weber et al. [WM09] learned predictive opponent models by analyzing expert gameplay from game logs. Drachen et al. [DCY09] recorded player behavior in TOMB RAIDER:UNDERWORLD® and determined four general types of players. More recently, data mining has been applied to analyze processes within *Massively Multiplayer Online Role Playing Game* (MMORPGs). Ducheneaut et al. [DY08] investigated the structure of social networks in WORLD OF WARCRAFT® based on data collected from more than 300,000 characters. They conclude that social networks in MMORPGs are typically sparse and that players experience a form of "collective solitude". While these contributions are closely related to our research, we are not aware of any previous work that studied formation patterns of social groups in web-based games multiplayer games.

Next, we briefly summarize the in-game mechanics of WORLD OF WARCRAFT® . Then, we introduce and explain the data and experimental setting of our study. A discussion of our findings and a conclusion will end this contribution.

2 WORLD OF WARCRAFT®

WORLD OF WARCRAFT® is an MMORPG that takes place in a medieval fantasy world. It is being played by millions of people worldwide and is arguably the most successful and most popular multiplayer game in video game history[1]. The technical architecture of the game consists of hundreds of separate servers each hosting virtual worlds, the so called *realms*, that are populated by several thousand players. Every player controls a single virtual character whose *strength* increases with successful completion of in-game tasks or *quests*. Strength or experience is measured in levels reaching from 1 (a newly created character) to 80 (currently the highest possible level which grants most powerful abilities such as special moves or mighty spells). Note that the initial release of WORLD OF WARCRAFT® capped the maximal experience level at 60. The first expansion package "Burning Crusade" (Jan 2007) extended it to level 70 and the second expansion "Wrath of the Lich King" (Dec 2008) further extended it to level 80.

Due to its open and flexible nature, the game lacks an explicit goal. However, most players strive to advance their character's experience towards level 80. Other examples of commonly pursued goals are the acquisition of virtual treasures or better equipment.

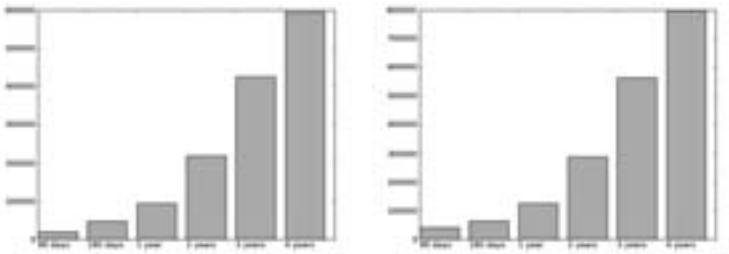
Social interactions are an important characteristic of the game. While players could play on their own, certain quests are designed such that they can only be accomplished by a group of players. In fact, the most valuable in-game items can only be obtained by groups comprised of several level 80 characters. Therefore, joining a team, also called a *guild*, is vital to successful play. Each character can only join a single guild. Players may leave or join guilds at their liking, however, for a joyful game experience, they typically try to find and stick with a guild that matches their own playing style.

3 Data and Features

For our investigation of group formation patterns in WORLD OF WARCRAFT® , we collected a large corpus of player logs from `http://www.warcraftrealms.com`. Our corpus covers the period from 2005 (when WORLD OF WARCRAFT® was released) to early 2009. It contains records of online appearances of players from European and United States WORLD OF WARCRAFT® realms. Each record contains the player's name, level, class, as well as information on guild membership. In total, we gathered 192,000,000 recordings of 18,000,000 characters belonging to 1,400,000 guilds. Since we can infer from this data when characters joined or left a guild, how many characters were with a guild at what time, and how experience was distributed among the members of a guild, the data roughly summarizes social in-game activities.

Since a characters's experience level provides a measure of skill, a guild formed by more higher level characters is more likely to succeed whereas a guild of only low level characters is literally excluded from certain parts of the game. The distribution of experience

[1] Note that the famous facebook application *FarmVille* is not a multiplayer game.



(a) EU guild numbers (b) US guild numbers

Figure 1: Growth of the total number of guilds in WORLD OF WARCRAFT® . For the US and the EU realms, the increase is surprisingly similar and exponential in nature.

levels within a guild therefore provides a feature to asses the guild in terms of in-game success (see Fig. 2). If we compute an experience histogram for a guild, it will also characterize the evolutionary state of the guild. This is regardless of whether we compute separate histograms for consecutive observations from distinct time slices, or if we compute a single histogram for the the overall monitoring period: consider the example of a guild newly formed by several level 80 characters. There will be no activity of level 10 players in this guild and the corresponding histogram bin will be empty. A guild originally formed by level 10 players, on the other hand, is likely to evolve over time so that in the end, its histogram may also contain observations of players of level 20 to 80.

4 Results

The results we present in the following were obtained using convex-hull non-negative matrix factorization (CH-NMF). This technique has recently been introduced as an efficient approach towards finding meaningful lower dimensional representation of massive amounts of data. In contrast to most known clustering techniques, it looks for the most extreme data points rather than for local means. Using CH-NMF, we can express every data point as a convex combination of the resulting extreme prototypes. This, in turn, allows for intuitive interpretation of the data because human cognition and language, too, often relies on extremes (consider the example of an almost empty glass, or a bottle that is half full). Due to lack of space we refer the reader to [TKB09] for the intricate mathematical details of the method.

We applied CH-NMF to 1,400,000 guild histograms that summarize data covering a period of 4 years. All results reported here were obtained by parameterizing CH-NMF to produce 8 basis vectors. Note that we experimented with different numbers of basis vectors and found that larger numbers hamper visualization whereas smaller numbers fail to capture the data variability. The 8 basis vectors used in this study are shown in Fig. 2. Using

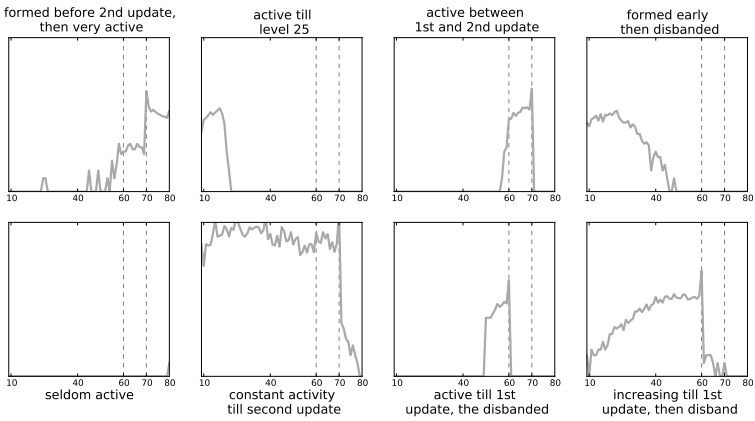


Figure 2: Basis vectors resulting from the application of CH-NMF to the WORLD OF WARCRAFT® guild database. The x-axis denotes experience levels, the y-axis denotes the number of observations per level. These histograms represent *archetypal* guilds and were automatically extracted from 1,400,000 data points.

natural language, these archetypal guild histograms can be paraphrased as “active from 10 to 80”, “active till 2nd update, then disbanded”, “formed early then disbanded”, “seldom active”, “increasing till 1st update, then disband”, “active between 1st and 2nd update”, “formed before 2nd update, then very active”, and “increasing activity, then disbanded”.

Figure 4 shows projections of guilds in European realms into the space spanned by the CH-NMF basis vectors. The projections show the distribution of guilds w.r.t the archetypal guilds after 60, 120, 240, 480, 720, and 1440 days, respectively. Note that each plot shows an 8 dimensional space visualized in a 2D plane. While this introduces distortions, it still preserves the main characteristics of the guild space.

It is noticeable that the number of guilds increases considerably over time (see also Fig. 1). Moreover, with more and more guilds to observe, the guild space becomes more densely covered. Interestingly, most guilds cluster close to the category “seldom active”. Only a rather small number of guilds (still many thousands) do not at least in part belong to this category. One explanation may be found in the exponential growth of the number of guilds. Since the growth rate of the number of players (not shown due to lack of space) is slower, it appears that formation of new guilds and abandoning of existing guilds are rather common occurrences in the game. Comparing European and American realms (see Fig. 3), we find similar growth patterns and no qualitative differences between the distributions.

5 Conclusion

This paper presented first results of a comprehensive study to understand general group formation processes MMORPGs. To the best of our knowledge, this is the first time that a



(a) after 720 days (b) after 1440 days

Figure 3: Distribution of American guilds in an 8 dimensional space spanned by archetypical guilds. The two plots show the situation after 720 and 1440 days, respectively. A qualitative difference to the European realms cannot be noticed.

vast amount of data was used for categorizing types of groups and their distribution over time. Applying convex-hull NMF, we determined 8 archetypal guilds and found there is a strong tendency towards more casual guilds. Our study of the temporal evolution of guilds revealed that from the release of the game onwards, most players joined guilds that did never evolve towards serious or even professional competition. Interestingly, this holds for American and European guilds alike, for we were not able to discover cultural differences. For the design of future games, we thus conclude that it is of major importance to cater to the needs of the casual gamer.

References

[BTS03] C. Bauckhage, C. Thurau, and G. Sagerer. Learning Human-like Opponent Behavior for Interactive Computer Games. In *Proc. Pattern Recognition*, volume 2781 of *LNCS*, 2003.

[DCY09] A. Drachen, A. Canossa, and G.N. Yannakakis. Player Modeling using Self-Organization in Tomb Raider: Underworld. In *Proc. IEEE Int. Conf. on Computational Intelligence and Games*, 2009.

[DY08] N. Ducheneaut and N. Yee. Collective Solitude and Social Networks in World of Warcraft. In C.T. Romm, C. Romm-Livermore, and K. Setzekorn, editors, *Social Networking Communities and E-dating Services: Concepts and Implications*. IdeaGroup, 2008.

[TBS04] C. Thurau, C. Bauckhage, and G. Sagerer. Learning Human-Like Movement Behavior for Computer Games. In *Proc. Int. Conf. on the Simulation of Adaptive Behavior*, 2004.

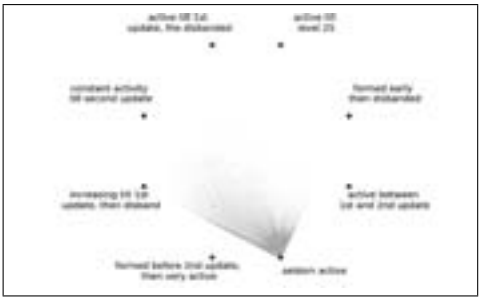
[TKB09] C. Thurau, K. Kersting, and C. Bauckhage. Convex Non-Negative Matrix Factorization in the Wild. In *Proc. IEEE Int. Conf. on Data Mining*, 2009.

[WM09] B. Weber and M. Mateas. A Data Mining Approach to Strategy Prediction. In *Proc. IEEE Int. Conf. on Computational Intelligence and Games*, 2009.

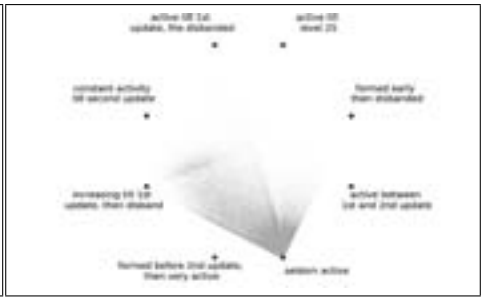


(a) after 60 days

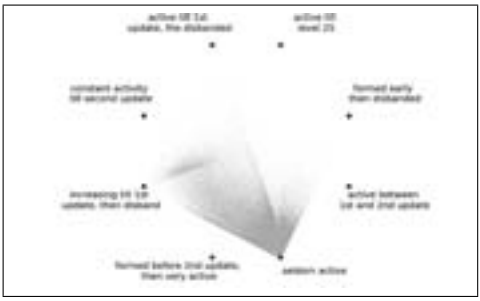
(b) after 120 days



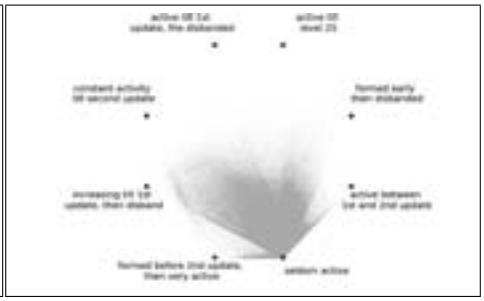
(c) after 240 days



(d) after 480 days



(e) after 720 days



(f) after 1440 days

Figure 4: Temporal evolution of the distribution of European guilds in an 8 dimensional space spanned by archetypical guilds. Shortly after the release of the game, most existing guilds were composed of casual players, i.e. they cluster around the *seldom active* archetype. With increasing time, the guild space becomes more densely covered. In particular the release of the first update is noticeable. However, even after four years, the vast majority of guilds consist of players that apparently play for fun rather than professionally and are therefore not very active.

Estimating the relevance of search results in the Culture-Web: a study of semantic distance measures

Laura Hollink
Delft University of Technology
l.hollink@tudelft.nl

Mark van Assem
VU University Amsterdam
mark@cs.vu.nl

Abstract: More and more cultural heritage institutions publish their collections, vocabularies and metadata on the Web. The resulting Web of linked cultural data opens up exciting new possibilities for exploring these cultural heritage collections. We report on ongoing work in which we investigate the estimation of relevance in this Web of Culture. We study existing measures of semantic distance and how they apply to two use cases. The use cases relate to the structured, multilingual and multimodal nature of the Culture Web. We distinguish between measures using the Web, such as Google distance and PMI, and measures using the Linked Data Web, i.e. the semantic structure of and semantic links between vocabularies. We perform a small study in which we compare these semantic distance measures to human judgements of relevance. Although it is too early to draw any definitive conclusions, the study provides new insights into the applicability of semantic distance measures to the Web of Culture, and clear starting points for further research.

1 Introduction

More and more datasets are released on the Web, open for anyone to use and explore. This is happening in the government sector, the scientific sector, and also in the cultural heritage (CH) sector. Museums, archives and libraries are opening up not only their collections of artworks, but also the associated metadata and thesauri. Many CH institutions facilitate reuse of their data and linking to other collections by adhering to standards such as SKOS and Dublin Core. Several institutions have now published their data as Linked Open Data. The resulting Culture Web[1] opens up exciting new possibilities for searching and browsing through cultural heritage collections. A significant effort towards a Culture Web is made in the Multimedian E-Culture project[2], in which collections and vocabularies of several cultural heritage institutions are linked and made available on the Web.

In this paper we report on ongoing work in which we investigate the estimation of relevance in this Web of Culture. When is a search result relevant? The specific characteristics of the Web of Culture – its level of structure, multilingual and multimodal nature – ask for a re-evaluation of existing techniques. At the same time these characteristics enable new methods for estimating the relevance of search results. The cultural heritage sector has always had a large amount of structured background knowledge available. Collections

[1] Term first coined by Guus Schreiber in his inaugural speech, see `http://www.cs.vu.nl/~guus/talks/oratie/index.html`.

[2] Project: `http://e-culture.multimedian.nl/`. See `http://e-culture.multimedian.nl/resources/datacloud/` for an image of the subset of the Culture Web in this project.

are annotated by professional documentalists, using extensive but weakly structured vocabularies. Cultural heritage collections consist largely of items of a visual or physical nature: videos of television broadcasts, images of paintings, drawings, statues and cultural artifacts such as juwellery and china. Unlike the web of science, where English is the dominating language, the web of culture consists of many languages. Since the cultural objects are usually tied to the culture or history of a nation, annotation and search in the native language is natural and necessary.

One way to estimate relevance is with semantic distance measures, as for example applied by [Bro98, TH07, WSA+09]. Our aim is to contribute to the knowledge about how these measures perform on the Web of Culture. We study existing measures of semantic distance and how they apply to two use cases. The use cases relate to the structured, multilingual and multimodal nature of the Culture Web. We distinguish between measures using the Web, such as Google distance, and measures using the Linked Data Web, i.e. the semantic structure of and semantic links between vocabularies. To this end, we ask two human subjects to judge the relevance of search results to a query and compare these judgements to relevance estimates based on semantic distance. We will not be able to present any hard conclusions but our goal is rather to provide insights into the notion of relevance and to bootstrap future work in this area.

2 Related Work

Semantic distance measures can be classified into: (1) co-occurrence based measures which rely on a large corpus of text and (2) structural measures, which rely on a predefined semantic structure. Both types of measures show good results when compared to human assessments of semantic distance. It appears that the results can even be improved when the two types of measures are combined [MMR09]. Examples of the first type are Latent Semantic Analysis [LD97], Pointwise Mutual Information [Tur01] and Google Distance [CV07]. The latter is based on the ratio between the number of hits returned by Google for keywords x and y and the number of hits for a combined search for x AND y, normalized by the total amount of webpages indexed by Google. An example of the second type of measure is Leacock and Chodorow's [LC98] simLC, which measures similarity of concepts in a hierarchy as the shortest path between x and y, normalized by twice the maximum depth of the hierarchy. Other measures make use of, for example, the lowest common subsumer of two concepts or the information content of the concepts [BH01].

Several researchers have exploited the availability of linked data on the web, and the application of different variants of structural semantic distance measures to retrieval or recommendation of relevant items is getting renewed attention. SemRank [AMS05] uses a mix of semantics to rank results for conventional or discovery oriented searches. Ruotsalo et al. [TH07] combine the semantic distance between concepts over several relations. Wang et al. [WSA+09] analyze which relationships between artworks deliver the best recommendations. The task we study differs from Ruotsalo and Wang in that we target search instead of recommendation, i.e. we start with a query instead of an annotated (and possibly rated) work of art. We set out to explore how semantic distance measures can be applied to estimate relevance in the Culture Web. As a first step, we implemented two established measures and tested them in two use cases: Leacock and Chodorow [LC98] as an

example of a structural measure and Normalized Google Distance (NGD) as a Web-based co-occurrence measure.

3 A comparative study

We have selected two highly specific but typical situations to study the applicability of semantic distance measures in the Web of Culture. Semantic distance is generally measured between concepts or words. However, in the Web of Culture a user is often interested in the relevance of an image that is annotated with these concepts or words. This potentially alters the way semantic distance measures can be relied on. We have tested this situation using Dutch and English language queries and annotations, taken from domain specific and general purpose semantic resources. Both use cases are performed with vocabularies and collections of images of art objects from the Multimedian E-culture project.

Use Case 1: Ethnological artifacts and SVCN The first use case is search in the collection of artifacts from the Dutch National Museum of Ethnology. The artifacts are annotated with a facetted thesaurus, the SVCN. The annotations are of high quality, made by experts. They include the culture, region and period from which the artifact originates, and the type of object it concerns. For this use case, 21 queries were drawn randomly from SVCN's Object facet, such as "tools for spinning and plying", "flutes", "membranophones: struck". These examples show that SVCN concepts can be highly domain specific.

Use Case 2: Rijksmusem, AAT and WordNet In the second use case we study the situation in which queries are from a different vocabulary than the annotations. We search a collection of images of artifacts in the Dutch Rijksmuseum, which is annotated with object types from the Art and Architecture Thesaurus (AAT). Each AAT concept has one or more English labels. The annotations are of lower quality; the artifacts were originally annotated with an in-house vocabulary and later automatically translated to the AAT in the Multimedian E-Culture project [SAvA+06]. We selected queries from WordNet. It's widespread use and general purpose concepts makes it an interesting test case. WordNet is a freely available thesaurus of the English language. Each concept has multiple synonymous labels. We drew 24 query concepts from the Physical object branch of the hierarchy, with labels such as "pot", "feather boa", "ring armor" and "artillery".

Evaluation of two semantic distance measures We implemented two established semantic distance measures and tested them on the two use cases: simLC by Leacock and Chodorow as an example of a structural measure and Normalized Google Distance as a Web-based co-occurrence measure. SimLC actually measures semantic proximity of concepts: the higher the score, the more similar two concepts are. NGD measures the distance between words or, in our case, labels. We slightly adapted NGD in the sense that we limited our Google searches to Dutch language pages in the first use case, and English language pages in the second.

To evaluate the two measures, we created a test set. Two human raters were presented with the same series of queries and resulting images of artifacts. They were asked to evaluate the

relevance of the images according to the following question: "How relevant is this work of art as a search result for the given query?". Ratings were on a 6-point scale ranging from "highly irrelevant" to "highly relevant". Rating was done with a web-based application displaying (1) a query including a definition and synonymous terms if available and (2) a search result in the form of an image of an artifact with a description. The two raters were not familiar with the art domain, but were habitual users of the Web. Although the raters sometimes disagreed on the judgements, all conclusions presented below are valid for both raters (with one small exception, see section 4). We tested inter-rater agreement with a weighted Cohen's κ [Coh68], which was reasonably high: 0.70 (squared weights).

We determined the correlation between the computed semantic distance and the human judgements with Spearman's ρ [Spe04]. This measure ranges from -1 to 1; values close to 0 mean a poor correlation and thus a poor estimation of relevance, values closer to -1 or +1 denote a stronger correlation and thus a good estimation of relevance.

The human raters judged relevance of query-artifact pairs. However, one artifact can be annotated with multiple concepts. As a result, multiple semantic distance scores between a query-concept and an annotation-concept need to be combined to come to an estimation of the relevance of an artifact for a query. This effect is even stronger for NGD since this measures the distance between words (or labels), not concepts. Query-concepts and annotation-concepts can both have multiple labels, resulting in a number of NGD measures for each query-artifact pair. We studied three methods to combine distance scores: (1) the mean of the scores, (2) the maximum of the scores, and (3) selecting only the score(s) of the annotation-concept that denotes the type of object depicted in the image, leaving out other annotations such as creator, location or culture. This part of the annotation comes closest to the queries, since all queries are for object types (e.g. "teapot", "firearm").

4 Results

Use Case 1: Ethnological artifacts and SVCN Table 1(a) shows the correlation between the human relevance judgements and estimations of relevance based on simLC scores for query–annotation pairs. The different ways to combine the semantic distance scores have a clear impact on the quality of the estimations. Taking the maximal score works better than taking the mean of the scores. A possible explanation is that each artifact is annotated with a few concepts that have no similarity at all with the query, such as the creator or the country of origin. Taking the mean score unjustly lowers the estimated relevance. The only method that leads to reasonable correlation is to select only object type annotations. This shows that it is beneficial to take into account what kind of query is posed (a person, location, object type, etc.) and to use knowledge of the metadata schema to pre-select suitable parts of the annotation.

We applied Normalized Google Distance to the same use case; results are shown in table 1(b). Contrary to simLC, the mean of the scores leads to a better estimate of relevance than the maximum of the scores. We assume this is because NGD measures co-occurrence rather than similarity. Using only the object type annotations again brings the best performance. Overall the relevance estimations of NGD are lower than those of simLC and the observed correlations are not statistically significant. Our hypothesis is that this is due to the domain specific, rare nature of our concepts and their labels, which is aggravated by

the fact that they are in Dutch. This leas to a low number of hits in Google and unreliable distance scores. NGD estimates are reasonable for relatively common query and annotation terms such as svcn:footgear by form — svcn:sandals and svcn:kit bag — svcn:sling (as in baby sling). They are completely off for terms like holders for chewing tobacco. We checked if the low performance was caused by labels consisting of multiple words, but we did not find such an effect.

Use Case 2: Rijksmusem, AAT and WordNet Table 1(c) show the results of use case 2. We see a clear rise in the performance of NGD as compared to use case 1. A plausible reason is that this use case contains queries in English from the general purpose resource WordNet. In line with this hypothesis we again observe that NGD works better for relatively common query words (e.g. "pot") than for rare words (e.g. "caldron") or ambiguous words (e..g "shovel", meaning a type of artillery). In the Rijksmuseum collection that was used for this use case, the only structured metadata that we had access to were object type annotations. Each annotation concept and each query had multiple labels. Table 1(c) therefore shows two different ways to combine the NGD scores for the labels of the object type annotations: mean and max. Again the mean of all NGD scores gives a better estimate of relevance than the maximum of all NGD scores of a query-result pair. As this use case consists of two separate vocabularies, standard structural similarity measures are not applicable.

(a)

Use case 1	**simLC**
mean	0.19*
max	0.25*
object type	0.45*

(b)

Use case 1	**NGD**
mean	-0.08
max	-0.02
object type	-0.19

(c)

Use case 2	**NGD**
object type mean	-0.25*
object type max	-0.16†

Table 1: Correlation between human relevance judgements and estimations of relevance the measures simLC and NGD. A * denotes significance at the 0.05 α-level; † denotes a significant correlation with the judgements of only one of the two raters.

5 Discussion and future work

In this paper we have applied two basic measures of semantic distance to two use cases in the Web of Culture. This exercise has provided us with some valuable insights into the relation between semantic distance and the notion of relevance on the Web of Culture. The applicability of purely web-based measures such as Normalized Google Distance seems to be limited in this context. The terms and their interpretation are so domain specific that word sense disambiguation errors are common. The rareness of the terms and the fact that the Web of Culture consists of many smaller languages leads to very low numbers of hits for google queries, which in turn leads to unreliable distance scores. One interesting direction for future research is to try other, more domain specific corpora for word co-occurrence based measures. The general Web might not be the best choice in this case.

Structural measures seem to work, but only when implemented with knowledge of the collection, metadata and queries. This means that semantic distance measures cannot be used

off the shelve. We plan future research into how to incorporate this type of knowledge in semantic distance measures, and moreover, how to (semi) automatically accumulate this knowledge. There is currently only little evidence that structural semantic distance measures work also across vocabularies. With the growth of open linked data this is likely to become more and more demand for measures that work across vocabularies and datasets.

These insights into the behavior of semantic distance measures on the Culture Web can be used to take the next step and design meaningful measures that combine the strengths of measures based on the Web or a domain specific corpus with structural measures, perhaps taking into account knowledge about the collection.

References

[AMS05] K. Anyanwu, A. Maduko, and A. Sheth. SemRank: Ranking Complex Relationship Search Results on the Semantic Web. In *Proc. of the 14th Intl. Conf. on World Wide Web*. ACM Press, 2005.

[BH01] A. Budanitsky and G. Hirst. Semantic distance in WordNet: an experimental application oriented evaluation of five measures. In *Proc. of the NACCL Workshop: on WordNet and other lexical resources: Applications, extensions, and customizations*, pages 29–34, 2001.

[Bro98] T. A. Brooks. The Semantic Distance Model of Relevance Assessment. In *Proc. of the 61st Annual Meeting of ASIS*, pages 33–44, 1998.

[Coh68] J. Cohen. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin*, 20:213–220, 1968.

[CV07] R. Cilibrasi and P. Vitanyi. The google similarity distance. *IEEE Transactions on knowledge and data*, 19(3):370–383, 2007.

[LC98] C. Leacock and M. Chodorow. *WordNet:An Electronic Lexical Database*, chapter Combining Local Context and WordNet Similarity for Word Sense Identification, pages 265 – 285. MIT Press, 1998.

[LD97] T. Landauer and S. Dumais. A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psych. Review*, 104(2):211–240, 1997.

[MMR09] Y. Marton, S. Mohammad, and P. Resnik. Estimating Semantic Distance Using Soft Semantic Constraints in Knowledge-Source-Corpus Hybrid Models. In *Proc. of the Conf. on Empirical Methods in Natural Language Processing*, pages 775–783, 2009.

[SAvA+06] G. Schreiber, A. Amin, M. van Assem, V. de Boer, L. Hardman, M. Hildebrand, L. Hollink, Z. Huang, J. van Kersen, M. de Niet, B. Omelayenko, J. van Ossenbruggen, R. Siebes, J. Taekema, J. Wielemaker, and B. Wielinga. MultimediaN E-Culture demonstrator. In *Proc. of the 5th Int'l Semantic Web Conf.*, number 4273 in Lect. Notes in Computer Science, pages 951–958, 2006.

[Spe04] C. Spearman. The proof and measurement of association between two things. *Amer. J. Psychol*, 15:72–101, 1904.

[TH07] T.Ruotsalo and E. Hyvönen. A method for determining ontology-based semantic relevance. In *Proc. of the Intl. Conf. on Database and Expert Systems Applications*, pages 680–688, 2007.

[Tur01] P. Turney. Mining the Web for Synonyms: PMI-IR versus LSA on TOEFL. In *Proc. of the 12th European Conference on Machine Learning*, pages 491–502. Springer, 2001.

[WSA+09] Y. Wang, N. Stash, L. Aroyo, L.Hollink, and G. Schreiber. Using Semantic Relations for Content-based Recommender Systems in CH. In *Proc. of the Workshop on Ontology Patterns at ISWC*, 2009.

Towards Standardized Vectorial Resource Descriptors on the Web

Wolfgang Orthuber[1]

[1]Department of Orthodontics
University Clinic Schleswig-Holstein
Arnold Heller Str. .3, House 26
24105 Kiel / Germany
orthuber@kfo-zmk.uni-kiel.de

Stefan Dietze[2]

[2]Knowledge Media Institute
The Open University
MK7 6AA
Milton Keynes / United Kingdom
s.dietze@open.ac.uk

Abstract: Resources with quantitative properties, e.g. measurable resources or sources for feature extraction (e.g. fingerprints), play an important role, particularly in scientific areas such as Life Sciences, the medical domain and nature sciences. In this paper we propose similarity-based representation of resources using so called Vectorial Resource Descriptors (VRDs) on the Web. The VRDs are standardized data structures which build the basis of Vectorial Web Search. Every VRD contains a feature vector and a Vector Space Identifier (VSI), and further data. In contrast to conventional keyword search, which requires matching of free text, Vectorial Web Search is well defined similarity search of numeric data. Users provide a VRD, or only the searched numeric data (i.e. the feature vector, as sequence of numbers) together with the VSI. The VSI is a HTTP URI which identifies the vector space of the feature vector, and which points to a standardized Vector Space Descriptor (VSD). So the valid distance function and the meaning of every dimension (number) of the feature vector is known by the system. For quantification of similarity the (in the VSD specified) distance function of the chosen vector space is used. The smaller the distance, the greater is the similarity of a VRD, and the higher is its rank in the search result.

1 Introduction

Due to the enormous amount of data on the Web there is an increasing need for information integration. This lead to the Semantic Web [Be01] and the Linked Data approach [Be06][Bi09] which aim at meshing together meaningful machine readable data on the Web. The mesh defines neighborhood: data which are directly linked together are neighboring in the mesh. While that approach requires to establish links explicitly, there is a well known mathematical concept for definition of neighborhood which allows implicit inference of neighborhood: similarity in a vector space.

While similarity computation usually exploits linguistic or structural similarities, we propose vectorial representation of resources to facilitate computation of similarities by means of distance functions (metrics) in shared vector spaces according to individual requirements. A vector space is much finer than any mesh of links. In our view, vectorial search complements but does not substitute traditional search approaches. In applications

in which it is usable, e.g. representation of measurements or other Quantifiable Resources (QR, see 2.1), it is a very efficient extension. A compact data structure called "Vectorial Resource Descriptor" (VRD) can serve as connector between the mesh of linked information and vector spaces, it is introduced below (chapter 3).

The usage of vector spaces for data integration is already topic of research, e.g. the Conceptual Space approach. Conceptual Spaces (CS) [Ga00][Pr97] follow a theory of describing entities at the conceptual level in terms of their natural characteristics similar to natural human cognition in order to avoid the symbol grounding issue [Di09a]. CS enable representation of resources as vectors within a geometrical space which is defined through a set of quality dimensions. For instance, a particular color may be defined as vector with the dimensions hue, saturation, and brightness. This is finer and more precise than a symbolic representation. Describing instances as vectors furthermore enables the automatic calculation of their semantic similarity, in terms of their distance, in contrast to the costly representation of such knowledge through symbolic representations. Even complex data which describe e.g. faces, speech and various types of fingerprints can be processed by feature extraction to vectorial form for similarity comparison and recognition. Generally, feature extraction is an essential pre-processing step to pattern recognition and machine learning problems. The resulting feature vectors open a large spectrum of applications for vector spaces [Gu06]. Vectors, embedded in VRDs, can describe all quantitative (and with this also all measurable) properties. This is the basis of *Vectorial Web Search* which means web based similarity search for quantitative data in standardized vectorial representation, using the well known metric space approach [Ze05]. In contrast to conventional keyword search [Br98], which requires matching of words, Vectorial Web Search is similarity search which exploits the ordered nature of quantitative (numeric) data. This is even necessary: quantitative descriptions are often very precise, having numeric representations with many digits and/or many dimensions, that not even one 1:1 match can be found worldwide, but many similar matches. These can be sorted in well defined way. In this paper we describe the application of vector spaces for representation of quantitative properties and data integration in general and propose a data structure and framework for efficient implementation of vectorial similarity search in the current Web infrastructure.

2 Approach

Precondition for vectorial (numeric) representation is reproducible quantification. To be suitable for a vectorial representation, a web resource must have one or several quantitative properties, i.e. one or several attributes which each have an inherent directed order (from "little" to "great"). We will call such a resource "Quantifiable Resource" (QR). One can say that “all possible” resources are quantifiable, because also quantification of the property “non existing” or “existing” can be done by associating both possibilities with numbers (e.g. with “0” or “1”). In many applications, however, fine distinction is required. When defining a vectorial (numeric) representation, aim is to find most important (i.e. decision relevant) ordered properties of a resource, so that most important variants of the resource can be represented by as few as possible numbers and small changes of the resource are mapped to small changes of the numbers.

Examples for QR: Human diagnostic parameters and measurements are attractive QR, because their standardized quantification allows world wide grouping of medical records of patients with similar parameters, to find the most efficient treatment for this group [Or06][Or10]. Some further examples for QRs are: Descriptors of personal profiles, digital representations of various items of daily life and their direct recognition (e.g. melodies, faces etc.), measurement and classification schemes of products and services and with this more individual and reproducible adaptation of products and services to the customers needs, semantic web services, concepts or situations, defined as elements (vectors) in conceptual spaces [Di09], numeric quantitative data, e.g. GPS coordinates, higher level sensor data, e.g. measurements of environmental or climate parameters, results of feature extraction, measurements and classifications in all areas of daily life.

3 Vectorial Resource Descriptors (VRDs)

Our proposed standardized data structure for representation of a QR on the Web is called *Vectorial Resource Descriptor* (VRD). It contains:

1. The *Identifier of the QR* (QRI). It is a HTTP URI which points to the resource.
2. The *Vector Space Identifier* (VSI). It is a HTTP URI which points to the *Vector Space Descriptor* (VSD).
3. The *feature vector* (usually a sequence of numbers). It represents the quantitative properties of the resource.

Additionally it can contain:

4. Auxiliary data, e.g. date, keywords.

The VSI is a HTTP URI [Bi09] which uniquely identifies the vector space and with this the meaning of every dimension (number) of the feature vector. The feature vectors of all VRDs with the same VSI are elements of the same vector space and with this comparable. Similarity search is done within this space. Being a HTTP URI, the VSI not only identifies the content of the feature vector, it simultaneously points to the Vector Space Descriptor (VSD).

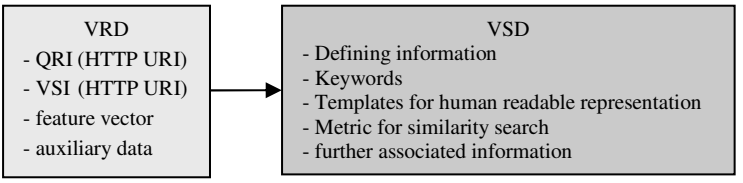


Figure 1. Contents of the VRD and VSD. The VRD's *Vector Space Identifier* (VSI) points to the VSD which provides important information about all VRDs with this VSI.

The VSD provides all necessary information about the vector space, particularly about the metric (distance function) for comparison, definitions of dimensions, templates for human readable representation of instances and links to further related Web content.

VRDs and VSDs are machine readable and can be embedded into the semantic Web as Linked Data [Bi09]. They extend the mesh of linked data by vector spaces, are uniformly

comparable and searchable, and it is possible to share the work for their definition and generation among all domain name owners.

The Resource Description Framework (RDF) [W3C04] can be used for VRD and VSD representation [Or10a]. If wished, a VRD can be also embedded in HTML files via RDFa. The feature vector is usually a sequence of numbers. This sequence can generally represent mathematical objects (e.g. coefficients, matrices of linear operators for conversion of vectors, tensors) whose definition can be done in the VSD. An even more expanded definition is possible: More generally the feature vector is a data structure which is defined in the VSD, usually designed for efficient comparison, so that the result of the comparison is a number. This convention makes the VRDs adaptable to the programmers' needs.

Example: Lets assume that a manufacturer of *round tables* owns the web address http://a.com and wants that his products are searchable according to their main geometric features height and diameter. Therefore he places a VSD on his website on the Web address http://a.com/round-table.rdf. The VSD holds the defining information of the vector space: It is two dimensional and contains feature vectors $V = (v_1, v_2)$ in which v_1 = height of the round table in cm and v_2 = diameter of the round table in cm. Keywords are: "Round Table". Template for human readable representation is: "round table with (D1) cm height and (D2) cm diameter" in which (D1) is placeholder for dimension 1 and (D2) is placeholder for dimension 2 of the feature vector. The metric of similarity search is the (by reciprocals of the standard deviations) weighted Manhattan distance *dws*: Let $V = (v_1, v_2)$ and $U = (u_1, u_2)$ denote two feature vectors, then $dws(V,U) = \left(\left| \frac{v_1 - u_1}{sd_1} \right| + \left| \frac{v_2 - u_2}{sd_2} \right| \right)$, in which sd_1 is the standard deviation of dimension 1 (of the height in cm) and sd_2 is the standard deviation of dimension 2 (of the diameter in cm). Reciprocals of standard deviations can be used as default weighting factors if further information is missing. This compensates inter alia the influence of the chosen unit.

After this the manufacturer (and other people) can add on the Web to every product description of a round table a VRD with QRI = Web address of the product description, VSI = Web the address of the VSD = "http://a.com/round-table.rdf", feature vector $V = (v_1, v_2)$ in which v_1 = height of the round table in cm and v_2 = diameter of the round table in cm. This makes the product descriptions of round tables searchable by height and diameter using vectorial web search (chapter 4).

4 Vectorial Web Search

The VRDs are the fundament of Web-scale vectorial similarity search (Vectorial Web Search). Due to the standardized structure of the VRDs one and the same search engine can be used for all search queries. Vectorial Web Search consists of the following steps:

- User provides a VRD or only its feature vector *V* with *VSI*.
- User confines search by a regular expression and/or by a conventional word based search string *S* (optional)
- Search engine selects all VRDs
  - with the chosen *VSI*
  - optionally with string *S* at the associated resource (identified by *QRI*)
- If a regular expression is given in step two, the collection is confined so that it fulfills this expression.
- Using the metric provided in the VSD (figure 1) the search engine calculates distances between the feature vector *V* and the feature vectors of the collected VRDs and sorts them according to distance using e.g. introsort [MU97] or another algorithm which allows fast parallel processing.
- In the search result the rank of collected VRDs and associated resources is the higher, the smaller the distance is.

As it is apparent in the above list, Vectorial Web Search starts with word based search for the VSI: After the user has provided a VSI and feature vector *V*, among all VRDs those with this *VSI* are selected (an index can accelerate this) and after optional further confinement used for comparison, i.e. the distances between their feature vectors and the searched feature vector V are calculated. In the search result the rank of VRDs and (via QRI, as described in chapter 3) associated resources is the higher, the less the distance is, i.e. most similar resources are listed first, using the proved and tested metric space approach [Ze05]. Concerning technical limitations of the approach we want to mention the curse of dimensionality. Low dimensional vector spaces are preferable because they can be handled much more efficiently than high dimensional spaces. Therefore it can become necessary that the search engine introduces an upper limit of dimensionality. Practical limitations can result from inefficient or redundant definitions of vector spaces. Therefore the VSD contains keywords, so that is possible to search for existing definitions to a topic before making a new definition.

5 Conclusions

Conventional language-based Web search does not facilitate similarity search of resources with certain quantitative properties, e.g. measurements. With this article we introduced a framework for standardized Vectorial Resource Descriptors (VRDs) on the Web, so that Vectorial Web Search, and with this similarity search of resources with quantitative properties can be realized in efficient way.

If there is enough support, further advanced evaluation of the approach will be part of our future work. The next important step is an online prototype. The concrete RDF format of VRDs and VSDs is currently topic of discussion [Or10a], comments are welcome. Advanced implementations could support flexible queries using e.g. SPARQL query language and/or realize a first practical medical or commercial application.

Bibliography

[Be01] Berners-Lee, T., Hendler, J., and Lassila, O. The Semantic Web. Scientific American 284(5) 2001, pp. 34-43.

[Be06] Berners-Lee, T. Linked Data. W3C Design Issues, 2006.

[Bi09] Bizer, C., Cyganiak, R., Heath, T. How to Publish Linked Data on the Web. http://www4.wiwiss.fu-berlin.de/bizer/pub/LinkedDataTutorial/, viewed 2009-05-27.

[Br98] Brin, S., Page, L. The Anatomy of a Large-Scale Hypertextual Web Search Engine. Proc. of 7th WWW Conference 1998, pp. 107-117.

[Di09] Dietze, S., Gugliotta, A., and Domingue, J.: Exploiting Metrics for Similarity-based Semantic Web Service Discovery, IEEE 7th International Conference on Web Services (ICWS), 2009, Los Angeles, CA, USA.

[Di09a] Dietze, S., Orthuber, W., and Domingue, J. Blending the Physical and the Digital through Conceptual Spaces, Workshop: OneSpace 2009 at Future Internet Symposium (FIS) 2009, Berlin, Germany.

[Fo02] Fodor I.K. A survey of dimension reduction techniques, US DOE Office of Scientific and Technical Information, 2002.

[Ga00] Gärdenfors, P. Conceptual Spaces - The Geometry of Thought. MIT Press, 2000.

[Gu06] Guyon, I., Gunn, S., Nikravesh, M., Zadeh, L., ed. Feature Extraction, Foundations and Applications. Studies in Fuzziness and Soft Computing, Springer, Berlin, Heidelberg, 2006.

[MU97] Musser, D., Introspective Sorting and Selection Algorithms. Software: Practice and Experience (Wiley) 27(8), 1997, pp. 983-993.

[Or06] Orthuber, W. Vectorial Web Search. http://www.orthuber.com/wpa.htm, since 2006-06-02.

[Or08] Orthuber, W., Fiedler, G., Kattan, M., Sommer, T., Fischer-Brandies, H.: Design of a global medical database which is searchable by human diagnostic patterns. The Open Medical Informatics Journal 2, 2008, pp. 21-32.

[Or10] Orthuber, W., Papavramidis E., Standardized Vectorial Representation of Medical Data in Patient Records, Medical and Care Compunetics 6, pp. 153-166.

[Or10a] Orthuber, W. Proposal for VRD and VSD representation. http://www.orthuber.com/wvstandard.pdf, since 2010-06-29.

[Pr97] Pratt I., Lemon O.: Logical and Diagrammatic Reasoning: the Complexity of Conceptual Space. Proceedings of the Nineteenth Annual Conference of the Cognitive Science 1997, Stanford, pp. 430-435.

[W3C04] W3C: RDF/XML Syntax Specification (Revised); Recommendation 10 February 2004. http://www.w3.org/TR/2004/REC-rdf-syntax-grammar-20040210/, viewed 2008-08-12.

[W3C07] W3C: XQuery 1.0: An XML Query Language; W3C Recommendation 23 January 2007. http://www.w3.org/TR/xquery/, viewed 2009-06-03.

[W3C08] W3C: SPARQL Query Language for RDF; W3C Recommendation 15 January 2008. http://www.w3.org/TR/rdf-sparql-query/, viewed 2009-06-03.

[Ze05] Zezula, P., Amato, G., Dohnal, V., Batko, M.: Similarity Search. The Metric Space Approach. Series: Advances in Database Systems, Vol. 32., Springer, Berlin, Heidelberg, 2005.

Workshop

Neue Forschungen zu Konrad Zuse

Hans Dieter Hellige

Neue Forschungen zu Konrad Zuse

Vorwort

Die „Fachgruppe Informatik- und Computergeschichte“ nimmt in Absprache mit der "Konrad-Zuse-Gesellschaft" den 100. Geburtstag Konrad Zuses zum Anlass für einen ganztägigen Workshop über diesen bedeutendsten deutschen Computerpionier. In Abgrenzung zu den zahlreichen Ehrungen und Gedenkveranstaltungen des "Zuse-Jahres" widmet sich der Workshop gezielt neuen Forschungsergebnissen, die auf der kritischen Auswertung des bisherigen Forschungsstandes, der Erschließung des Nachlasses und neuer Archivquellen sowie auf neuen Fragestellungen und Forschungsansätzen beruhen.

Es soll dabei nicht erneut der Prioritätenstreit über den „Vater des Computers“ ins Zen trum gerückt werden, der lange Zeit die Geschichtsschreibung dieses Technikbereiches von den interessanteren technikhistorischen und techniksoziologischen Fragestellungen abgehalten hat. Vielmehr soll eine Brücke von der engeren Computer-Historie zur Geschichte der Informatik und Informationsverarbeitung geschlagen werden und damit die viel umfassenderen Leistungen Zuses in den Blick geraten. Denn die Frage nach *dem* Erfinder *des* modernen Computers hat wesentliche Aspekte von Zuses Inventions- und Innovationstätigkeit in den Hintergrund treten lassen, insbesondere seine frühen Visionen und Konzepte der Nutzung von Rechenanlagen, die weit über das hinaus gingen, was in dieser Zeit in den USA und Westeuropa an Einsatzmöglichkeiten gesehen wurde. Zuses Perspektive war hierbei nicht mehr die eines Rechenmaschinen-Erfinders, der nach neuen Anwendungen Ausschau hielt, sondern bereits die eines ‚Informatikers‘, der auf der Basis des Logikkalküls zu einer „allgemeinen Theorie des Rechnens“ vordrang. Auf der anderen Seite sind die frühen Computing-Konzepte Zuses stark vom historisch-gesellschaftlichen Kontext geprägt, so dass sie stärker als bisher im zeitgeschichtlichen Zusammenhang betrachtet werden sollten.

Ziel des Workshops ist auf der Basis einer kritischen Bestandsaufnahme der Zuse- Forschung die Vorstellung neuester Forschungsresultate und die Erschließung neuer Forschungsfragen und -aspekte. Thematische Schwerpunkte sind dabei:

- Neue Forschungsperspektiven durch den Zuse-Nachlass
- Konrad Zuse als Erfinderunternehmer im NS-Staat
- Konrad Zuses Anwendungs-Erfindungen in ihrem Zeitkontext
- Konrad Zuses Rolle als Unternehmer und Technologievermittler
- Konrad Zuses Beiträge zur Informatik-Forschung und -Theorie

Bremen, Juni 2010 Hans Dieter Hellige

Programmkomitee:

Dr. Gerard Alberts (Universität Amsterdam)
Dr. Paul E. Ceruzzi (Smithsonian Institution, Washington, D.C.)
Dr. Ulf Hashagen (Deutsches Museum, München)
Prof. Dr. Hans Dieter Hellige (Universität Bremen)
Prof. Dr. Stephan Lindner (Universität der Bundeswehr München)
Prof. Dr. Friedrich Naumann (Technische Universität Chemnitz)
PD Dr. Rudolf Seising (European Centre for Soft Computing, Mieres)
Prof. Dr. Roland Vollmar (Universität Karlsruhe)
Prof. Dr. Horst Zuse (Technische Universität Berlin)

Einführung: Die Zuse-Forschung und die Probleme der Pionier-Geschichten

Rudolf Seising

European Centre for Soft Computing,
Mieres, Asturias (Spanien)
rudolf.seising@softcomputing.es

Abstract

Biographische und monographische Darstellungen zu Konrad Zuse konzentrierten sich in der Vergangenheit sehr stark auf den Nachweis der Priorität der Zuse-Maschinen vor den amerikanischen und englischen Computer-Entwicklungen. Der einführende Beitrag gibt einen Überblick über die Problemstellungen, Ergebnisse und offenen Fragen der bisherigen Forschung über den wichtigsten deutschen Computerpionier. Er formuliert vor dem Hintergrund der gewandelten Erkenntnisinteressen der neueren Computer- und Informatikgeschichte Desiderate einer kritischen Zuse-Forschung.

Der wissenschaftlich-technische Nachlass Konrad Zuses als Grundlage für neue Forschungsfragen und eine biografische Neubewertung

Wilhelm Füßl
Deutsches Museum München
w.fuessl@deutsches-museum.de

Abstract

Erst seit wenigen Jahren ist der Nachlass Zuses in einem öffentlichen Archiv frei zugänglich. Seither besteht erstmals die Möglichkeit, die schriftliche Hinterlassenschaft des Computerpioniers vollständig im Original durchzusehen und kritisch zu erforschen. Eine erste Bilanz wurde in der Sonderausstellung des Deutschen Museums „100 Jahre Konrad Zuse – Einblick in den Nachlass“ und im begleitenden Ausstellungskatalog gezogen.

Der wissenschaftlich-technische Nachlass Zuses im Archiv des Deutschen Museums umfasst heute insgesamt 26 Regalmeter mit schätzungsweise 88000 Seiten, 2500 Schaltplänen bzw. Zeichnungen und mehrere Tausend Fotografien. Enthalten sind darin biografische Unterlagen, Notizen, Typoskripte und Manuskripte Zuses, Veröffentlichungen von und zu ihm, Pläne und technische Zeichnungen, Fotomaterial sowie Briefwechsel.

Die Geschichte des Nachlasses von Konrad Zuse ist eng mit der Biografie und der Historie der von ihm gebauten frühen Rechner verknüpft. Bedingt durch mehrfache Umzüge und Verluste durch Luftangriffe auf Berlin wurden nicht nur die Rechner Z1 bis Z3 vernichtet, sondern auch umfangreiche Teile der Dokumente aus der Pionierzeit. Trotzdem scheint Zuse vor der Verlagerung seiner Firma nach Göttingen und dann nach Hinterstein eine Reihe wichtiger Schriften und Zeichnungen gerettet zu haben. Zusammen mit seit 1945 neu angefallenem Material lagerte der Nachlass bis 2005 in seinem Haus in Hünfeld.

In den Jahren 1977 bis 1979 förderte die Gesellschaft für Mathematik und Datenverarbeitung (GMD) ein Projekt, das die Sicherung eines wesentlichen Teils der Unterlagen von Konrad Zuse vorsah. Die Idee ging dahin, gemeinsam mit dem Computerpionier zentrale Texte auszuwählen, zu verzeichnen und dann auf Mikrofilm zu sichern. Diese Mikrofilmaufnahmen wurden im folgenden Vierteljahrhundert die Grundlage für nahezu alle Forschungen zu Zuse. Auch das von Raúl Rojas online gestellte „Konrad Zuse Internet Archiv" basierte auf diesen Verfilmungen. Bei dem GDM-Projekt wurden jedoch lediglich 569 Dokumente aus dem Gesamtnachlass erschlossen. Dabei erfolgte die Auswahl in enger Zusammenarbeit mit Zuse, der zu dieser Zeit „sehr häufig die Datumsangabe beziehungsweise das Entstehungsjahr, Herkunftsangaben und ähnliches" ergänzte. Gleichzeitig ordnete Zuse offensichtlich im Zuge der GMD-Erschließung einen Großteil seiner schriftlichen Dokumente neu und stellte sie in Zusammenhänge, in denen sie bis dahin nicht eingebettet waren. Die historisch gewachsene Ordnung wurde so aufgelöst. Es fällt auf, dass sich heute in zahlreichen Archivmappen Dokumente aus unterschiedlichen Zeiträumen finden. Vermutlich war es 1977-1979 die Intention Zuses, bestimmte Sachzusammenhänge herzustellen.

In Verbindung mit späteren Datierungen vermittelten die Mikrofilme eine scheinbare Authentizität. Um den Nachlass für biografie- und computerhistorische Forschungen zu Zuse adäquat zu nutzen, ist es heute notwendig, anhand der Originaldokumente die ursprünglichen Zusammenhänge zu erkennen und die durch Zuse und die GDM-Filme intendierten Sachverhalte zu dekonstruieren. Gleichzeitig müssen die Datierungen der Dokumente kritisch hinterfragt werden.

Eine zusätzliche Schwierigkeit für die Zuse-Forschung bilden Tausende von Stenogrammen, die sich im Nachlass befinden. Zuse hat zeit seines Lebens wichtige Notizen in Stenografie verfasst. Teilweise wurden sie im Auftrag der GMD in Maschinenschrift transkribiert und von Zuse gegengelesen. Aber auch hier gibt es Abweichungen der Transkriptionen von den Stenogrammen. Im Vortrag werden diejenigen Stenogramme, die primär als tagebuchartige Notizen zu interpretieren sind, kritisch auf ihre Zusammenhänge und Datierungen hin untersucht. Insgesamt werden 160 Stenogramme bearbeitet.

Die Kontrollnetze und 'Rechnenden Räume' des Konrad Zuse im „Dritten Reich"

Hans Dieter Hellige
Universität Bremen
hellige@artec.uni-bremen.de

Abstract

Die über den Nachlass jetzt zugänglich gewordenen Korrespondenzen und zusätzlichen Dokumente und Texte vermitteln ungeachtet aller Nachlasslücken ein schärferes Bild der technisch-wissenschaftlichen, politischen und weltanschaulichen Entwicklung Konrad Zuses von den späten 20er bis in die 40er Jahre. Sie lassen deutlicher den Wandel von einem künstlerisch ambitionierten, technisch verspielten, mit neuen Medien und Automaten experimentierenden Erfindertalent zu einem durch die Weltwirtschaftskrise verunsicherten Ingenieurstudenten erkennen, der nach beruflichem, gesellschaftlichen und weltanschaulichem Halt suchte. Er ordnete sein Ingenieurweltbild mit Henry Fords Heilsbotschaft, die sozialen Konflikte durch ein neues gesellschaftliches Regulationsmodell auf der Basis von Automatisierung und Massenproduktion zu lösen. Er fand Halt in Spenglers radikalkonservativem „preußischen Sozialismus", der mit den Mitteln der „faustischen Technik und Erfindung" die Gesellschaftskrise in Deutschland überwinden und mit einer „weißen Weltrevolution" den drohenden Untergang aufhalten wollte. Zuse begriff sich mit seinem „faustischen Projekt" eines 'universalen' Rechenautomaten sogar als Vollstrecker der Spenglerschen Prophezeihung des kommenden „Zeitalters, in dem alles berechnet werden kann". Er formte seit den frühen 30er Jahren ein rechtskonservatives Welt- und Gesellschaftsbild, in dem sich technokratische Aspirationen zur Berechnung, Kontrolle, Steuerung und Ordnung der gesellschaftlichen Kräfte und die politisch-soziale Konstellation des nationalsozialistischen Deutschland mehr und mehr durchdrangen.

Die Integration wurde verstärkt durch den erfolgreichen beruflichen Einstieg in ein Unternehmen, das unmittelbar von Hitlers Aufrüstung profitierte. Mit seinem Bestreben, die Erfindung unter allen Umständen zu realisieren und, nach Kriegsbeginn, der Einberufung an die Front zu entgehen, geriet er immer mehr in die Mühlen von Militärtechnik und Kriegswirtschaft. Zuses ambitioniertes Projekt einer programmgesteuerten Rechenmaschine kam dadurch trotz des zivilen Ursprungs Ende der 1930er Jahre zunehmend in militärisches Fahrwasser. Seine Versuchsanlagen wurden ab 1940 entgegen späterer Selbsteinschätzung mit insgesamt ca. 300 Tsd. RM relativ großzügig von (halb-) staatlichen und militärischen Stellen gefördert. Zudem erhielt er seit Anfang 1943 und verstärkt seit der Einbeziehung in das „Jägerprogramms" ab November 1944 immer weitergehende kriegswirtschaftliche Privilegien. Die Zuse-Firma wurde so am Ende zu einem Anhängsel der militärischen Flugzeug- und Flugbomben-Entwicklung.

Im Sog der „Endfertigung“ richtete er sich nun bereits auf eine Serienfertigung seiner Maschinen ein und berechnete die bei einer Verteilung vieler Rechengeräte über das Reich erzielbare Arbeitseinsparung auf eine Million Arbeitstunden, wobei er ständig neue Einsatzmöglichkeiten anvisierte. Durch den zeitgeschichtlichen Kontext verschoben sich auch Zuses Nutzungsvorstellungen und Anwendungskonzepte für seine Geräte, so dass das anfängliche Ziel einer „Rechenmaschine des Ingenieurs“ immer mehr in den Hintergrund trat. Die Entdeckung der Potentiale des „Allgemeinen Rechnens“ ab 1938/39 fiel nämlich zeitlich zusammen mit rüstungs- bzw. kriegwirtschaftlichen Organisationsproblemen sowie kriegsgesellschaftlichen Kontrollbedürfnissen des NS-Staates. Die Einbindung in die Kriegswirtschaft des “Dritten Reiches“ bewog Zuse, konzeptionell die Bestandsführungs- und Kontrolltechniken der Lochkartenwelt mit den technisch-wissenschaftlichen Berechnungen des frei programmierbaren Computers zu verknüpfen. Daraus entstanden bei ihm umfassende Ideen der „automatischen Behandlung schematisch-kombinatorischer Aufgaben“ und der Erfassung des „ganzen schematischen Wissens“ in ständig aktualisierbaren organisatorisch vernetzten „Speicherwerken“. Zuse entwarf in Tagebuchnotizen wie in Nutzungsvorschlägen für staatliche Instanzen perfekte Systeme betrieblicher und verwaltungsmäßiger Erfassung und Kontrolle, Staat und Wirtschaft wurden so für ihn permanent ‘rechnende Räume’. Seine Entwurfsskizzen zur „Gefolgschaftskontrolle“, zur Warenbestandsstatistik und „Warenkreislaufkontrolle“ sowie zur allgemeinen Personendaten-Speicherung unterschieden sich dabei in der Konzeption kaum von den Lochkarten-basierten Erfassungs-, Sortier- und Kontrollsystemen des „Dritten Reiches“. Allerdings hatten Zuses Personen- und Sachen-Erfassungskonzepte wegen des noch unfertigen Zustandes seiner Versuchsanlagen ausschließlich Projektcharakter, sie blieben durch das Kriegsende im Mai 1945 lediglich Ausdruck ausufernder technokratischer Kontrollphantasien.

In seinen Visionen einer mithilfe verkoppelter Rechenmaschinen und Erfassungsstationen durchrationalisierten Staatserwaltung und Wirtschaft durchdrangen sich politisch-weltanschauliche Orientierungen, professionelle Einstellungen und zeitgeschichtliche Rahmenbedingungen. Die durch das „Dritte Reich“ bewirkte Prägung ging dabei so tief, dass Zuse auch noch in den Nachkriegsjahren die Realisierung von umfassenden computerbasierten Erfassungs- und Kontrollsystemen weiterverfolgte und zeitlebens an dem Ideal eines informationstechnisch perfekt organisierten Staates festhielt, wobei er die daraus erwachsenden vermehrten staatlichen Kontrollmöglichkeiten als unvermeidlich hinnahm. Trotz seiner Erfahrungen während des „Dritten Reiches“ wurde er sich der Gefahren des Zusammenspiels von technischen Kontrollpotentialen und politischem Interesse an umfassenden Massenbeobachtungs- und Kontrollsystemen offenbar niemals bewusst.

Von Berlin ins Allgäu: Der Erfinderunternehmer Konrad Zuse 1945-1948

Ulf Hashagen
Deutsches Museum München
u.hashagen@deutsches-museum.de

Abstract

Konrad Zuse hat in seiner Autobiographie eine Darstellung seiner Tätigkeit als Erfinderunternehmer im NS-Staat und in den Nachkriegsjahren gegeben, in der er intendiert, dass er durch die politischen Randbedingungen des „Dritten Reiches“, durch den Krieg und durch die politisch-wirtschaftlichen Rahmenbedingungen der Nachkriegsjahre als einer der „Erfinder der Computers“ stark behindert wurde. Diese Selbstdeutung wird auch durch seine autobiographische Darstellung der abenteuerlichen Umstände der Evakuierung seiner Firma (sowie seines Rechners Z4) aus Berlin im Februar 1945 und seiner Lebensumstände in den Jahren 1945 bis 1948 im Allgäu gestützt: Hiernach wurde die Evakuierung aus Berlin nur möglich, weil es gelang, die deutschen Kriegsbehörden zu täuschen und dadurch die notwendigen Bescheinigungen für die Verlegung seines Rechners und seiner Mitarbeiter nach Göttingen und schließlich ins Allgäu zu erlangen; in Hinterstein bzw. Hopferau im Allgäu habe er sich von 1945 bis 1948 in weitgehender wissenschaftlicher Isolation sowie ohne die Möglichkeit befunden, seinen Arbeiten am Rechner Z4 fortzusetzen, und sich daher vor allem seinen theoretischen Arbeiten zum „Plankalkül“ und zum „Rechnenden Raum“ gewidmet.[1] Die Zuse-Historiographie hat zwar die Beziehungen Zuses zum militärisch-wissenschaftlichen Komplex des NS-Staates nachzuzeichnen versucht, ist aber in der Bewertung dem von Zuse gezeichneten Bild dann in weiten Teilen gefolgt.[2] Gerade in neueren Arbeiten wurde diese Interpretation kanonisiert: so konstatierten mehrere Autoren in einem 2004 erschienenen Sammelband, dass Zuses „Rückhalt in den staatlichen Institutionen und der Industrie bis zum Kriegsende gering geblieben“ sei, und dass die deutsche Kriegsführung den Wert der Rechentechnik verkannt habe;[3] und im renommierten *Dictionary of Scientific Biography* hieß es unlängst – ganz im Sinne dieser Deutung – dazu: “Zuse’s own career was cut short by the war.”[4]

[1] K. Zuse: Der Computer, mein Lebenswerk. Verlag Moderne Industrie, 1970; K. Zuse: Der Computer: Mein Lebenswerk. Springer, 1984.
[2] H. Petzold: Rechnende Maschinen. Eine historische Untersuchung ihrer Herstellung und Anwendung vom Kaiserreich bis zur Bundesrepublik. VDI-Verlag, 1985; R. Zellmer: Die Entstehung der deutschen Computerindustrie: Von den Pionierleistungen Konrad Zuses und Gerhard Dirks' bis zu den ersten Serienprodukten der 50er und 60er Jahre. Universität Köln, 1990.
[3] H. D. Hellige (Hrsg.): Geschichten der Informatik. Visionen, Paradigmen, Leitmotive. Springer, 2004.
[4] R. Rojas: Zuse, Konrad. In: New Dictionary of Scientific Biography. Scribner, 2008, 25, S. 408-414.

Die Biographie Konrad Zuses wurde damit zu einem Symbol für die durch externe Faktoren in den 1940er und 1950er Jahren behinderte Entwicklung der deutschen Computerindustrie und passte sich ideal in die von Naturwissenschaftlern und Ingenieuren nach 1945 zu ihrer Entlastung geschaffene „Erzählung“ vom Niedergang des deutschen Wissenschafts- und Innovationssystems im „Dritten Reich“ ein: Der Niedergang des deutschen Wissenschaftssystems sei aufgrund der systemimmanenten Wissenschaftsfeindlichkeit des NS-Staates und seiner verfehlten, überbürokratisierten und zentralisierten Planungswut sowie der zu stark betonten Abschottung von Forschungsbereichen eingeleitet worden. Die in Deutschland verbliebenen deutschen Ingenieure hatten vor allem ihre scheinbar zweckfreien technikutopischen Visionen verwirklicht, indem sie eine „Kriegswichtigkeit“ ihrer Projekte vortäuschten. Die neuere Forschung zur Wissenschafts- und Technikentwicklung im NS-Staat hat dieses Bild nachdrücklich in Frage gestellt und sowohl für die wissenschaftsorganisatorischen Institutionen des NS-Staates wie für eine Reihe von Forschungsbereichen eine weitgehend effektive Planung wie eine praktisch problemlose Anpassung und Integration der natur- und ingenieurwissenschaftlichen Forschung in die rüstungspolitischen Ziele des NS-Staates konstatiert. Ebenso hat die historische Forschung inzwischen ein wesentlich differenzierteres Bild über die Beziehungen zwischen den deutschen Wissenschaftlern und Ingenieuren zu den alliierten Besatzungsmächten nach 1945 gezeichnet.

Der Vortrag nimmt diese neueren Ergebnisse zur Wissenschafts- und Technikentwicklung im NS-Staat und zur Forschungspolitik der Alliierten in Deutschland sowie eine vom Vortragenden durchgeführte Studie zur Entwicklung des „Scientific Computing“ im deutschen Wissenschaftssystem[5] zum Ausgangspunkt für eine Analyse der politischen, wirtschaftlichen und wissenschaftspolitischen Randbedingungen, unter denen der Erfinderunternehmer Konrad Zuse in den letzten Monaten vor Kriegsende in Berlin und Göttingen sowie nach Kriegsende im Allgäu agierte. Neue Quellenfunde ergeben ein in wesentlichen Punkten von Zuses autobiographischer Darstellung abweichendes Bild über die Entwicklungen von 1945 bis 1948. Insbesondere wird analysiert, wie sich die Beziehungen zwischen Zuse und einer Reihe von alliierten Wissenschaftlern gestalteten und wie Zuses Rechnerentwicklung 1945 bis 1948 von deutschen und alliierten Wissenschaftlern bewertet wurden.

[5] http://www.histsem.uni-freiburg.de/DFG-Geschichte/Hashagen.htm.

Unternehmenskultur, Entscheidungsstrukturen und Führungsprobleme in der Zuse KG

Horst Zuse

Technische Universität Berlin /
Hochschule Lausitz in Senftenberg
horst.zuse@T-Online.de

Abstract

Die Zuse KG war zugleich eine Erfolgs- und eine Misserfolgsgeschichte. Sie war ein Erfolg, weil es einem Außenseiter gelang, auf der Basis seiner Computererfindungen ein mittelständisches Unternehmen aufzubauen und zwischen 1949 und 1969 insgesamt mehr als 251 Rechenanlagen zu bauen. Sie wurde jedoch ein Mißerfolg, weil es Konrad Zuse nicht schaffte, den hohen Kapitalbedarf der Firma für den Umstieg auf die Mikroelektronik zu sichern und das Management an die Anforderungen eines entstehenden Großunternehmens anzupassen. Der Beitrag untersucht die noch stark von dem patriarchalischen Führungsstil und den informellen Kommunikationsstrukturen eines Ingenieurbüros geprägte Unternehmenskultur der Zuse KG. Er zeigt an ausgewählten Beispielen, wie Führungsprobleme und Entscheidungsstrukturen zum Scheitern strategisch wichtiger Entwicklungsprojekte führten.

Konrad Zuse und N. Joachim Lehmann

Hartmut Petzold
Berlin
h.petzold@deutsches-museum.de

Abstract

Lehmann traf Zuse zum ersten Mal Ende 1949 bei einem Besuch in der neugegründeten Zuse KG im hessischen Neukirchen. Zuse baute gerade an der Z4, die er nach den Wünschen von Eduard Stiefel von der ETH Zürich ergänzte. Da Zuse sich damals, wie auch noch in den folgenden Jahren, entschieden für die elektromechanischen Relais und gegen die Elektronenröhren als Elementarbausteine für die Rechenautomaten stark machte, während Lehmann entschlossen auf einen elektronischen Rechner setzte, blieb es bei einem eher zurückhaltenden gegenseitigen Abtasten. Immerhin nahm Lehmann die originelle Idee Zuses wahr, dass als Lochstreifen ein Kino-Normalfilm verwendet werden konnte, da dafür eine ausgereifte Transporttechnik zur Verfügung stand. Als Jahre später die vorgesehenen Fernschreiber nicht zur Verfügung standen, versah auch er den Dresdener D1 mit dieser Kinofilmeinrichtung, so dass die Auswirkungen dieses ersten Zusammentreffens in Dresden bis in die 1960er Jahre hinein wahrgenommen werden konnten.

Sehr viel später sollte sich Lehmann erinnern, dass das wichtigste, was dieser Besuch für das Dresdener Rechnerprojekt ergeben habe, zwei Argumente gewesen seien: dass ein programmgesteuerter Rechenautomat auch mit - im Vergleich zu den damals bekannten Rechnern in den USA und in England - bescheidenen Mitteln gebaut werden konnte, und dass durch die Ersetzung der parallelen Ziffernverarbeitung in der Z4 durch die serielle, jedoch vielfach schnellere elektronische Ausführung noch weitere Einsparungen zu erzielen seien. So habe der Bezug auf die Z4 wesentlich zur Bewilligung des Dresdener Projekts beigetragen.

Es folgten weitere Zusammentreffen bei den Jahrestagungen der GAMM, wobei sich beide im gegenseitigen Meinungsaustausch besser kennen und auch schätzen lernten. Während der gesamten 1950er Jahre - in beiden deutschen Staaten war der Aufbau einer Computerindustrie mit staatlicher Hilfe das große Thema - stand Zuse als einzige Unternehmerpersönlichkeit immer auch als weitsichtiger Berater zur Verfügung. Wenn in der DDR entsprechend dem Programm der SED die damalige Bundesrepublik auch auf dem Gebiet der Computertechnik überholt werden sollte, dann waren es die Rechner der Zuse KG, die bis in die beginnenden 1960er Jahre den Maßstab setzten. Bei der Beurteilung dieser Situation hatte die Stimme von Lehmann immer Gewicht.

Das Jahr 1968, in dem Zuse aus seinem Unternehmen ausscheiden musste und Lehmanns so erfolgreich betriebenes „Institut für Maschinelle Rechentechnik“ der Hochschulreform geopfert wurde, dürfte die beiden Persönlichkeiten einander noch näher gebracht haben. Für beide waren damit aber offenbar auch Hindernisse beiseite geräumt, die bisher einer engagierten Zuwendung zur Programmier- und Softwaretechnik entgegengestanden hatten. Während der 11 Jahre ältere Zuse nun endlich seinen Plankalkül vollständig publizierte, engagierte sich Lehmann für die Ausarbeitung und Verbreitung des Fachsprachensystems DEPOT. Er sollte sich darum bis zum Ende der DDR bemühen.

Konrad Zuse und die parallele Rechnerarchitektur. Einige Gedanken und Beispiele

Klaus Waldschmidt
Johann Wofgang Goethe Universität Frankfurt
Technische Informatik
waldsch@ti.informatik.uni-frankfurt.de

Abstract

Konrad Zuse ist heute anerkannt als Erfinder und Konstrukteur der ersten vollautomatischen, programmgesteuerten und frei programmierbaren Rechenanlage der Welt.

Seine ersten Maschinen, wie auch die später in der Zuse-Fabrik entstandenen Computergenerationen, zählen überwiegend zu der Klasse der sequentiellen Rechenmaschinen.

Konrad Zuses Überlegungen und Aktivitäten im Bereich der parallelen Rechnerarchitektur sind ebenfalls sehr vielfältig, aber leider nur sehr wenig bekannt.

Im Vortrag soll versucht werden, anhand einiger Beispiele die große Genialität Konrad Zuses als Rechnerarchitekt auch im Bereich der parallelen Rechnerarchitektur zu beleuchten.

Als Beispiele werden ausgewählt:

- die Feldrechenmaschine
- der Assoziativspeicher und
- der parallele Addierer

Es sind drei Beispiele aus drei verschiedenen Bereichen der parallelen Rechnerarchitektur, die ebenfalls originär auf Konrad Zuse zurückgehen. Sie stammen aus den 40er und 50er Jahren, also noch aus der Frühzeit der modernen Rechnerarchitektur.

Konrad Zuses informatisches Universum

Roland Vollmar
Universität Karlsruhe
vollmar@ira.uka.de

Abstract

Konrad Zuse wurde des öfteren als erster (deutscher) Informatiker bezeichnet. In meinem Vortrag will ich mich mit dieser Charakterisierung befassen. Dazu wird ein Teil seines Werkes näher betrachtet werden, die Wissenschaft "Informatik" wird kursorisch in ihrer Entwicklung dargestellt und nicht zuletzt wird Konrad Zuses Haltung (in seinen verschiedenen Lebensabschnitten) zu ihr und zur Mathematik und Physik beleuchtet werden.

Workshop

8. Workshop Automotive Software Engineering

Prof. Dr. Jörn Schneider und Dr. Dirk Ziegenbein

8. Workshop Automotive Software Engineering

Prof. Dr. Jörn Schneider[1] und Dr. Dirk Ziegenbein[2]

[1]Fachhochschule Trier
Fachbereich Informatik
54293 Trier
joern.schneider@fh-trier.de

[2]ETAS GmbH
70469 Stuttgart
dirk.ziegenbein@etas.com

Vorwort

Mobilität von Personen, Waren und Gütern ist in unserer vernetzten Welt nach wie vor Grundlage oder zumindest die als selbstverständlich betrachtete Voraussetzung nahezu jeglicher Dienstleistung. Kraftfahrzeuge sind somit unverzichtbare Basis und oftmals auch Mittel erheblicher Anteile der Dienstleistungswirtschaft. So wie die Informatik neue Perspektiven im Bereich „Service Science“ eröffnet und gleichsam daraus bezieht, so schafft Softwaretechnik in der Automobil- und Nutzfahrzeugindustrie, nicht erst seit heute, neue Horizonte und bezieht als Disziplin ebenso Möglichkeiten und Herausforderungen zu weiterer Innovation aus dem Komplex Fahrzeug - Mensch - Umwelt.

Am und im Kraftfahrzeug trifft die hohe Dynamik von originär automobilfremden Anwendungsbereichen auf die klassischen, häufig sicherheitsrelevanten Funktionen des Fahrzeugs. An der Fähigkeit die „Schnittstelle Kraftfahrzeug“ mittels Automotive Software Engineering so zu gestalten, dass Dienstleistungsmehrwert mit Zuverlässigkeit und Sicherheit koexistieren kann, entscheidet sich wie groß die Möglichkeiten und Kosten unserer zukünftigen Mobilität sind. Thema dieses Workshops ist **„Automotive Software Engineering“** im Zentrum der oben geschilderten Herausforderungen.

Die ausgewählten Papiere behandeln ein breites Spektrum aktueller Themen rund um Technologien, Architekturen, Konzepte, Prozesse, Methoden und Werkzeuge zur Entwicklung und Integration softwarebasierter Fahrzeugfunktionen sowie Darstellungen von praxisbezogenen Lösungsansätzen. Das gleichermaßen aus Industrie- und Hochschulvertretern besetzte Programmkomitee hatte die schwierige Aufgabe unter den Einreichungen das beste Drittel an Beiträgen auszuwählen. Aufgrunddessen mussten auch vielversprechende und gut geschriebene Papiere abgelehnt werden.

Die Organisatoren des Workshops möchten ihren besonderen Dank gegenüber den Autoren der eingereichten Beiträge und den Mitgliedern des Programmkomitees für Ihren persönlichen Beitrag zum Gelingen dieses Workshops ausdrücken.

Juli 2010
Trier

Programmkomitee

Dr. Christian Allmann, Audi Electronics Venture GmbH
Hans-Jürgen Belz, Hella KGaA Hueck & Co
Prof. Jürgen Bortolazzi, Dr. Ing. h.c. F. Porsche AG
Prof. Manfred Broy, TU München
Dr. Mirko Conrad, The MathWorks, Inc.
Dr. Michael Daginnus, Volkswagen AG
Prof. Werner Damm, Universität Oldenburg
Bernd Frielingsdorf, Ford Werke GmbH
Dr. Bernhard Hohlfeld, ICS AG
Prof. Stefan Jähnichen, TU Berlin, Fraunhofer FIRST
Ralf Kalmar, Fraunhofer IESE
Dr. Hubert Keller, FZ Karlsruhe GmbH
Prof. Stefan Kowalewski, RWTH Aachen
Dr. Thomas Kropf, Robert Bosch GmbH
Dr. Ulrich Lefarth, ETAS GmbH
Peter Manhart, Daimler AG
Dr. Stefan Ortmann, Carmeq GmbH
Prof. Klaus Pohl, Universität Duisburg-Essen
Prof. Wolfgang Pree, Universität Salzburg
Dr. Alexandre Saad, BMW Group
Prof. Jörn Schneider, FH Trier
Prof. Michael Uelschen, FH Osnabrück
Gerhard Wirrer, Continental AG
Dr. Fabian Wolf, Volkswagen AG
Dr. Dirk Ziegenbein, ETAS GmbH

Geschäftsprozessautomatisierung im Nutzfahrzeuge-Support-Bereich

Andreas Loos
pro et con
Innovative Informatikanwendungen GmbH
Dittesstraße 15
D-09126 Chemnitz
andreas.loos@proetcon.de

Richard Sum
ESC - Engineering services and consultation
MAN Nutzfahrzeuge AG
Dachauer Straße 667
D-80995 München
Richard.Sum@man.eu

Abstract: Der Artikel beleuchtet Softwaretechnik in der Automobil- und Nutzfahrzeugbranche aus der Sicht eines „Third-Level-Support"- Bereiches der MAN Nutzfahrzeuge AG. Dieser beschäftigt sich vorrangig mit der Bereitstellung von Dienstleistungen für MAN-Nutzfahrzeuge (LKW und Bus) und MAN-Einbaumotore. Durch einen fest definierten Workflow werden Modifikationen in der individuellen Fahrzeugsoftware zeitnah und zuverlässig bereitgestellt und können so weltweit per Softwareupdates direkt auf die Steuergeräte jedes MAN-Nutzfahrzeugs/-Motors übertragen werden. Gemeinsam mit der Firma pro et con Innovative Informatikanwendungen GmbH wurde eine flexible und effiziente Lösung entwickelt, um der stark steigenden Kundennachfrage gerecht zu werden. Die dabei realisierten Strategien sind Inhalt dieses Artikels.

1 Einleitung

Die Abteilung ESC (**E**ngineering **S**ervices and **C**onsultation) der MAN Nutzfahrzeuge AG stellt MAN-Kunden verbindliche, technische Informationen wie CAD-Fahrgestellzeichnungen, Herstellerbestätigungen oder Aufbaurichtlinien zur Verfügung und bearbeitet als technischer Dienstleister („Third-Level-Support") jegliche Anfragen (im Fortlauf Kundenanfragen genannt) von Aufbaufirmen, Aufbauherstellern, TÜV/DEKRA, Zulieferern, Werkstätten, Polizei, Sachverständigen, Kunden u.v.m.

Weiterhin erstellt MAN-ESC Parametrierbescheide und maßgeschneiderte Software als Ersatzteil, über die Werkstätten in die Lage versetzt werden, Änderungen an der Auslegung der Funktionalität der Fahrzeugelektronik, Anpassungen von Hard- und Software (Komfortfunktionen, Anzeigen, Systeme) oder sicherheitskritische und genehmigungspflichtige Änderungen an der Fahrzeugelektronik vorzunehmen. Die damit einhergehende Fahrzeugparametrierung beinhaltet Softwareänderungen und Updates, die durch über 5.500 Servicecomputer in ca. 105 Ländern von Werkstätten direkt im Fahrzeug vorgenommen werden können. Stillstandszeiten der Nutzfahrzeuge, die durch Verzögerungen in diesem Parametrierprozess entstehen würden, wären aus wirtschaftlicher Sicht schwer vertretbar.

Für MAN-ESC ist daher eine weltweit zeitnahe Zustellung der Dokumente und Software bei konstant hoher Qualität unabdingbar. Das umso mehr, da bei Softwareupdates in einem Nutzfahrzeug nahezu jede sicherheitsrelevante Funktion wie z.B. Bremsenmanagement, Steuerung der Luftfederung und Getriebeschaltstrategie beeinflußt werden kann.

Ein typischer MAN-ESC-Arbeitsprozess (Vorgang) besteht aus einer kundenseitig gestellten Anfrage, einer Bearbeitung dieser Anfrage, der MAN-seitigen Generierung neuer Fahrzeugsoftware und dem Versand eines Bescheides an den Kunden. Jeder Vorgang bei MAN-ESC ist mit der Erstellung eines Bescheides, bestehend aus einer beliebigen Anzahl von Dokumenten, verbunden. Trotz nahezu gleichbleibender, personeller Ressourcen bei MAN-ESC stieg die Nachfrage (Anzahl Vorgänge) und zugleich die Menge der zusätzlichen Anforderungen in den letzten Jahren deutlich[1].

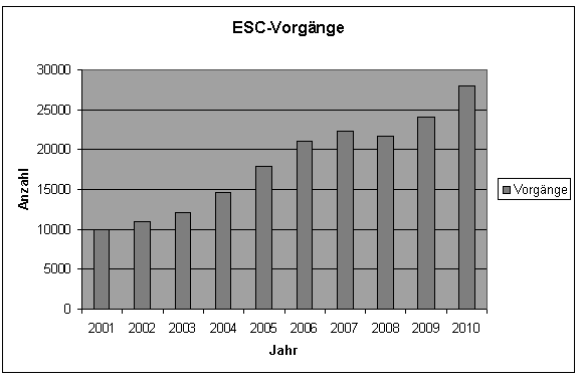


Abbildung 1: Zahl der MAN-ESC-Vorgänge

Die Erfüllung dieser Anforderungen war nur durch einen grundsätzlichen Wechsel der Technologie in der Bearbeitung von Kundenanfragen bei MAN-ESC möglich. Dieser wurde durch ein Softwareprojekt „**H**armonisierung und **I**ntegration von **T**echnik-Datenbanken“ („**HIT**“) vollzogen.

2 Ausgangssituation und Zielstellung

Bis ca. 2003 existierten dezentrale IT-Lösungen, die essentiell den bei MAN-ESC existierenden Arbeitsgruppen zugeordnet waren. I.d.R. kamen Microsoft-Office-Produkte zum Einsatz. Bedingt durch die Dynamisierung der Märkte, durch das kontinuierliche Ansteigen der Vorgangszahlen und die steigende Variantenvielfalt der MAN-Nutzfahrzeuge erwies sich dieser eingeschlagene Weg als nicht mehr tragfähig.

Weiterhin wurde ein typisches Dilemma aller Einzelplatzlösungen offensichtlich: Datenmodelle und Datenstände, z.B. eine Kundenverwaltung, welche allen MAN-ESC-Anwendungen inhärent sind und damit eine redundanzfreie, einheitliche Darstellung erfordern, drifteten im Laufe der Zeit zwischen den einzelnen Anwendungen kontinuierlich auseinander.

MAN-Kunden weltweit hatten bis zu diesem Zeitpunkt nur die Möglichkeit, Anträge an MAN-ESC per Fax, Post oder E-Mail zu senden. Dokumente, die während der Vorgangs-

[1] Die Zahlen für 2010 wurden auf Basis bislang ermittelter Werte hochgerechnet.

bearbeitung bei MAN-ESC entstanden, wurden gedruckt und aus Gründen der Revisionssicherheit als Microfiche langzeitarchiviert. Dieser Prozess war kosteninstensiv und langwierig.

Ausgehend von diesen Aspekten wurden unterschiedliche Strategien diskutiert, mit denen dieser Misere begegnet werden sollte. Eine Polarisierung erfolgte einerseits in Richtung einer etablierten Standardsoftware und andererseits in Richtung abteilungsspezifischer Webapplikationen auf Basis einer Abteilungssoftware, mit welcher der gesamte MAN-ESC-Workflow modelliert werden kann. Wesentlicher Grund für die Entscheidung zugunsten der zweiten Variante war, dass die Anwendung damit hochgradig flexibel und skalierbar wird. Eine schnelle Reaktionsmöglichkeit auf modifizierte Situationen ist gegeben. Ebenso ist ein feingranulares Design möglich, mit welchem auch detaillierte Optimierungen der Abteilung umgesetzt sowie neue zusätzliche Anforderungen und Aufgaben integriert werden können. Diesen Ansprüchen werden Standardprodukte nicht in jedem Fall gerecht.

Kundenseitig wurde eine Webapplikation konzipiert, welche die Möglichkeit bietet, Anträge online an MAN-ESC zu stellen.

Ziel ist damit die Schaffung eines integrierten Workflowsystems, welches die Welten „Kunde“ und „MAN-Sachbearbeiter“ koppelt. Beide haben einen entsprechend autorisierten Zugang zum Gesamtsystem. Der für den Kunden entscheidende Synergieeffekt besteht in der deutlichen Verkürzung der Bearbeitungs- und dem damit verbundenen Absenken ineffizienter Stillstandszeiten der Nutzfahrzeuge. Gleichzeitig erfolgt eine Verbesserung der Qualität der Kundenanfragen, vor allem in Bezug auf Lesbarkeit, Identifizierbarkeit und der Sprachproblematik.

3 Strategische Designentscheidungen

3.1 Architektur

Die Architektur des Gesamtsystems ist das einer typischen Webapplikation. HIT beinhaltet dabei eine Reihe von Besonderheiten wie integrierte Intranet- und Internetfunktionen oder Interfaces zu angrenzenden MAN-Systemen. Über einen Onlinezugang arbeiten MAN-ESC-Sachbearbeiter mit HIT im MAN-Intranet, über das Internetportal https://www.manted.de erhalten Kunden weltweit Zugang zu allen Dienstleistungen von MAN-ESC. Das System wurde softwareseitig vollständig neu entwickelt. Existierende Datenbestände hingegen mussten in die neue Umgebung migriert werden. Die Geschäftslogik repräsentiert alle bei MAN-ESC ablaufenden Geschäftsprozesse, ohne dabei Arbeitsgruppenspezifika zu vernachlässigen. Integriert in die Architektur sind Versandprozesse, mit denen Dokumente automatisiert zwischen MAN-ESC und Kunden ausgetauscht werden können.

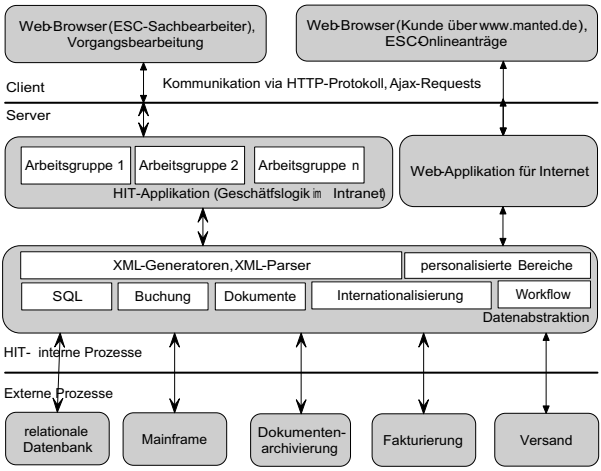


Abbildung 2: Schematische Darstellung der HIT-Applikation

3.2 Daten- und Dokumentenmigration

Inhalt der Datenmigration war eine Schema- und Datentransformation der antiquierten ACCESS-Datenhaltung in Oracle-Datenbanken. Der Bestandskonsolidierung an Kunden-, Fahrzeug- und Vorgangsdaten wurde in diesem Zusammenhang besondere Aufmerksamkeit gewidmet. Vorteilhaft war, dass pro et con als Technologieführer für Software-Migration und Reverse Engineering über jahrelange Erfahrungen auf diesen Gebieten verfügt.

Bedingt durch Differenzen zwischen Ausgangs- und Zielschema musste ein Schemamapping [2] durchgeführt werden, welches u.a. das Mapping korrespondierender Attribute und die Überführung des Schemas in die 3. Normalform beinhaltete. Der Migrationsprozess erfolgte vollautomatisch unter Einsatz eines speziell für diese Aufgabe entwickelten Tools. Auch mussten die im Laufe der Zeit entstandenen Redundanzen und Widersprüche zwischen den dezentralen Lösungen behoben werden. Hier wurden Algorithmen wie die „Levenshtein-Distanz“ (Editierabstand) angewendet. Der Editierabstand [1] beschreibt die Zahl Editieroperationen, die notwendig ist, um eine Zeichenkette S_1 in eine andere Zeichenkette S_2 zu überführen. Ist er zwischen S_1 und S_2 gering, kann in gewissen Situationen auf die Identität von S_1 und S_2 geschlossen werden. Differenzen dieser Art treten u.a. durch Schreibfehler auf. Auf diese Art und Weise konnten ca. 80% der Inkonsistenzen im Datenbestand ermittelt und beseitigt werden. Ca. 15% mussten manuell nachbearbeitet werden und 5% waren redundant und wurden nicht migriert.

3.3 MAN-ESC-Onlineanträge

Ein strategisches Ziel der HIT-Architektur besteht in der Integration eines Webportals zur kundenseitigen Antragserstellung. Dieses ist über *https://www.manted.de* erreichbar. Anträge, die bislang per Fax, Post oder E-Mail vom Kunden an MAN-ESC versendet wurden, können jetzt online gestellt werden. Damit arbeiten jetzt Kunden und MAN-ESC als Dienstleister direkt mit einem Gesamtsystem, welches alle Informations- und Dokumentenflüsse digital verwaltet. Informations-/Qualitätsverluste und zeitlicher Versatz, der aus manuellen Prozessen resultiert, wird damit eliminiert.

3.4 Kernmodule

Die hohen Qualitätsanforderungen an alle von MAN-ESC erstellten Bescheide setzen voraus, dass der mit einem Vorgang betraute Sachbearbeiter von Routinearbeiten befreit ist und sich seinem Kerngeschäft widmen kann. Aus diesem Grund wurden in HIT Basismodule für unterschiedliche Anwendungsfälle und Szenarien integriert, von denen hier nur ausgewählte erwähnt werden. So erzeugt HIT aus heterogenen Dokumentenvorlagen (Templates), Textbausteinen und Datenbankinhalten automatisiert Dokumente und nimmt somit die Aufgaben eines konventionellen Dokumentenmanagementsystems (DMS) wahr. Die Templates selbst liegen in einem XML-Format vor und werden durch XSL-Transformation in verschiedene Formate (.doc, .pdf) konvertiert. MAN-ESC bietet seine Dienstleistungen in bis zu 21 Sprachen weltweit an. Für HIT ergibt sich daraus die Aufgabe, Dokumente mehrsprachig zu generieren. Für den Sachbearbeiter ist eine manuelle Erzeugung oder Kontrolle der textuellen Inhalte fremdsprachiger Dokumente nicht mehr möglich. Der gesamte Übersetzungsprozess erfolgt vollautomatisch unter HIT-Kontrolle. HIT besitzt Schnittstellen zu MAN-internen Systemen wie Fakturierung oder Archivierungstools. Die Analyse dieser externen Prozesse zeigte, dass Daten teilweise nur in proprietären Formaten vorliegen. Ein allgemeingültiges Schnittstellenformat existiert nicht. Daher wurde eine Datenabstraktionsschicht in HIT integriert, die diverse Datenformate in XML-Formate transformiert. Den Funktionalitäten der HIT-Geschäftslogik bietet sich damit ein syntaktisch homogener Datenstrom, der durch gängige XML-Parser analysiert wird. HIT arbeitet hochgradig mit Ajax- (Asynchron Javascript and XML) Methoden und bietet damit eine performante Schnittstelle, in der XML-Datenströme zwischen Server und Client (Sachbearbeiter) ausgetauscht werden.

3.5 Jobcontroller

Die einzelnen HIT-Komponenten arbeiten auf unterschiedlichen Plattformen. Weiterhin ist HIT als Onlinesystem ausgelegt. Diese beiden Aspekte sind ein Grund dafür, dass plattformübergreifende Prozesse zur Steuerung des Gesamtsystems notwendig sind. Die über hausinterne Fakturierungssysteme erstellten Rechnungen beispielsweise werden zeit-

verzögert an HIT geliefert, so dass hier ein Versand im direkten Onlinebetrieb nicht möglich ist. Deshalb verfügt HIT über eine Jobsteuerung. Sie synchronisiert u.a. die Interaktion zwischen HIT und allen angrenzenden Systemen.

4 Produktentwicklung und Wartung

Seit der Inbetriebnahme 2005 wurde HIT kontinuierlich weiterentwickelt und gewartet. Von Vorteil war, dass ESC-Sachbearbeiter von Beginn an mit Layoutvorschlägen und fachlichen Tests involviert waren. Dadurch wurde sofort eine hohe Akzeptanz erreicht. HIT wurde mit einer vertikalen Konzeption entwickelt, d.h., es wurde ein Paket für eine Arbeitsgruppe komplett fertiggestellt und in Produktion genommen und danach die Entwicklung für die nächste Gruppe begonnen. Der Vorteil dieser Vorgehensweise besteht in einer kontinuierlichen Umstellung der gewohnten Arbeitsprozesse der Sachbearbeiter in kleinen Schritten, was die mit Migrationsprozessen generell verbundenen, erhöhten individuellen Fehlerquoten reduziert. Weiterhin wird Wert auf kurze Updatezyklen gelegt. Neue fachliche Anforderungen treten in der Regel kurzfristig auf und müssen umgehend mit einer produktiven Lösung untersetzt werden. HIT verfügt über ein integriertes Ticketsystem, mit dem Fehlermeldungen direkt an die Entwickler geleitet werden. Die Reaktionszeit bei kritischen Situationen liegt im Bereich von ca. einer Stunde.

5 Fazit

Die MAN-ESC-Applikation „HIT“, gekoppelt mit einer Internetschnittstelle, welche weltweit über https://www.manted.de erreichbar ist, sichert die zeitnahe Belieferung der MAN-Kunden mit Dokumenten und Fahrzeugsoftware, um der hohen Dynamik im Nutzfahrzeugsektor gerecht zu werden. Sie trägt damit wesentlich zur Senkung der Ausfall- und Stillstandszeiten der Fahrzeuge bei und stellt das Basistool für die Realisierung zukünftiger, neuer Anforderungen dar.

Literatur

[1] Vladimir I. Levenshtein: Binary codes capable of correcting deletions, insertions, and reversals. In: Doklady Akademii Nauk SSSR, 163(4) S. 845-848, 1965 (Russisch). Englische Übersetzung in: Soviet Physics Doklady, 10(8) S. 707-710, 1966.

[2] Erhard Rahm, Philip Bernstein: A survey of approaches to automatic schema matching. VLDB Journal 10(4), 2001.

[3] R.Sum, A.Loos: HIT - Aspekte einer Migration von Einzelplatzlösungen in webbasierte Client-Server-Systeme. In: Lecture Notes in Informatics, 10th Workshop Software Reengineering (WSR 2008). S. 166-170, 2008, ISBN 3-88579-220-8

Konzepte zur Entwicklung, Integration und Vernetzung von Car-to-X Funktionen

Torsten Ronneberger, Daniela Thum, Claudia Kratzsch

I/AEV-2, I/EB-11, I/EK-543
AUDI AG
85045 Ingolstadt
{Torsten.Ronneberger, extern.Daniela.Thum, Claudia.Kratzsch}@audi.de

Abstract: Innerhalb des Forschungsprojekts sim^{TD} werden domänenübergreifend Car-to-X Funktionen entwickelt, die miteinander vernetzt in einem Gesamtsystem integriert und letztendlich evaluiert werden. In diesem Beitrag soll anhand des Architekturkonzeptes einer sim^{TD} Hauptfunktion auf die domänenübergreifenden Herausforderungen und die projektspezifischen Lösungen bei der Entwicklung von kooperativen hoch vernetzten Car-to-X Funktionen eingegangen werden. Diese Arbeit wurde im Rahmen des Projektes sim^{TD} durch die Bundesministerien für Wirtschaft und Technologie (BMWi) sowie Bildung und Forschung (BMBF) gefördert und durch das Bundesministerium für Verkehr, Bau und Stadtentwicklung (BMVBS) unterstützt.

1 Einführung in die Car-to-X Kommunikation

Innovative Fahrzeugkonzepte sind darin motiviert, dass sie die Sicherheit erhöhen, den Fahrkomfort verbessern oder die Effizienz im ökologischen und ökonomischen Sinne steigern sollen. Um diese Ziele zu erreichen, muss ein Fahrzeug nicht mehr nur als autarke Einheit gesehen werden, sondern es muss in Interaktion mit seiner Umwelt stehen. Interaktion bedeutet dabei, dass ein Fahrzeug mit der Umwelt kommuniziert.

Die Car-to-X Kommunikation [LA08] definiert die Kommunikation des Fahrzeuges über einen drahtlosen Kommunikationsstandard [JD08] mit seinem Umfeld. Das Umfeld kann dabei innerhalb einer Domäne durch Fahrzeug-zu-Fahrzeug Kommunikation bzw. domänenübergreifend z.B. durch Fahrzeug-zu-Infrastruktur Kommunikation beschrieben werden. Weiterhin gibt es eine Vielfalt von weiteren Kommunikationspartnern wie z.B. Werkstätten, Tankstellen, Parkhäusern oder auch dem privaten Wohnbereich mit dem das Fahrzeug relevante Informationen austauscht und diese entweder direkt oder in einer verarbeiteten Form als Hinweise oder Warnungen dem Fahrer zur Verfügung stellt. Es lässt sich somit ableiten, dass es für Funktionen der Car-to-X Kommunikation zwingend notwendig ist, domänenübergreifend Konzepte zur Vernetzung der einzelnen Komponenten zu entwickeln.

2 Vorstellung des Entwicklungskontextes

Das Projekt sim^{TD} [Sim10] verfolgt die Ziele, die Praxistauglichkeit und Wirksamkeit von Car-to-X Kommunikation innerhalb eines Großversuchs unter realen Bedingungen mit Blick auf Verkehrssicherheit, Verkehrseffizienz und Fahrkomfort nachzuweisen.

Dieser Nachweis stellt eine hoch komplexe Entwicklungs- und Integrationsaufgabe dar. Dazu ist eine starke Vernetzung der Projektabläufe sowie der beteiligten Partner erforderlich. Um dies zu erreichen, haben sich Unternehmen der Automobil-, Zuliefer-, sowie Telekommunikationsbranche, der kommunalen Infrastrukturbetreiber ebenso wie Bildungs- und Forschungseinrichtungen in einem Konsortium zusammengeschlossen (siehe Abbildung 2.1).



Abbildung 2.1: Partner und Förderer im nationalen Forschungsprojekt sim^{TD}

In Anbetracht der Interessen der Partner und deren Erfahrungen aus vorangegangenen Forschungsprojekten wurden insgesamt 21 Funktionen unterschiedlicher Art zur Umsetzung im Projekt und zur Evaluierung im realen Feld identifiziert. Diese Menge an Funktionen wurde in Funktionsbereiche (sog. Hauptfunktionen) eingegliedert, um eine inhaltliche Abgrenzung und somit logische Clusterung zu erreichen. Die Cluster Verkehrsinformation, Verkehrssteuerung, lokale Gefahrenwarnung, Fahrerassistenz sowie Informationsdienste adressieren die unterschiedlichen Domänen im verkehrlichen Umfeld. Ein weiterer Vorteil dieser logischen Clusterung liegt in der Auflösung von möglichen Inkonsistenzen sowie Überlappungen und der Entkopplung unterschiedlicher Funktionsbereiche im Gesamtsystem, wodurch eine unabhängige Umsetzung dieser in der Entwicklung gewährleistet ist. In diesem Rahmen gilt es die 21 Funktionen innerhalb einer vierjährigen Projektdauer zu entwickeln, zu integrieren, zu testen und zu bewerten.

Im Idealfall sollen die einzelnen Entwicklungsphasen inkrementell ablaufen. Da die Projekteigenschaften die Projektabläufe stark beeinflussen, ist es erforderlich eine Optimierung der Entwicklungsdauer durch Parallelisierung und damit verbunden einer höheren Vernetzung der Entwicklungsphasen zu erreichen.

3 Konzepte einer domänenübergreifenden Funktionsentwicklung

Ausgehend von den Projektzielen, der Projektstruktur, dem Projektumfang und dem Durchführungszeitraum ist es notwendig, in den einzelnen Entwicklungsphasen einen adaptiven Entwicklungsansatz und pragmatische Entwicklungslösungen zu finden. Dabei muss berücksichtigt werden, dass schon eine Vielzahl von funktionalen und hardwaretechnischen Teillösungen vorliegt, die nicht neu entwickelt werden muss. Das Ziel der sim^{TD}-Entwicklung ist es, ein einheitliches Gesamtsystem bereit zu stellen.

3.1 Wirkkettenorientierte Funktionsspezifikation

Die Projekteigenschaften bestimmen den Entwicklungsprozess maßgeblich. Somit ist in sim^{TD} ein Spezifikationsprozess aufgestellt worden, der folgende wesentliche Eigenschaften berücksichtigt:

- die Wiederverwendung des vorhandenen Wissens
- der enge Spezifikationszeitraum im Verhältnis zum Entwicklungsumfang
- das Ziel der Überprüfung der Funktionswirkung im realen Umfeld
- die Heterogenität und die örtliche Entfernung der einzelnen Entwicklungsteams

Um diese wesentlichen Eigenschaften im projektspezifischen Spezifikationsprozess unterzubringen, wurde ein deduktiver Top-down-Beschreibungsansatz erarbeitet, der es ermöglicht, Car-to-X Funktionen einheitlich und parallel zu beschreiben. Dabei ist eine Funktionsbeschreibung so aufgebaut, dass sie in Form eines Leitfadens die notwendigen Spezifikationsinhalte Top-down strukturiert. Der Ausgangspunkt des Leitfadens und somit auch der Funktionsspezifikation sind die Projektziele. Davon abgeleitet werden entsprechende funktionsspezifische Ziele aufgestellt, welche für den Funktionskontext gelten. Im Funktionskontext ist die Ist-Situation mit den darin enthaltenen Defiziten beschrieben, auf die die Funktionsidee wirken soll. Durch die Beschreibung des zu erzielenden Soll-Zustandes lässt sich eine Wirkhypothese ableiten, die es durch die Test- und Versuchsauswertung und mit Hilfe der Mess- und Kenngrößen zu bestätigen gilt [BD06]. Nach der Beschreibung der Wirkung wird die Funktionsidee, die Ausgangspunkt der Funktionsbeschreibung ist, systematisch weiter konkretisiert bzw. spezifiziert. Dabei erhalten die einzelnen Spezifikationselemente im gesamten Leitfaden einen eindeutigen Bezeichner, der in einer verfeinerten Beschreibungsstufe weiter referenziert wird.

Durch die Top-down Herleitung und die Referenzierung der Spezifikationselemente ist es den verteilt agierenden Entwicklern eines Entwicklungsteams möglich, die parallele Arbeit an einem Leitfaden nachzuvollziehen. Weiterhin ermöglicht die Inhaltsstuktur des Leitfadens, Spezifikationsmeilensteine zu definieren, anhand dieser die parallel entwickelten Spezifikationsinhalte innerhalb einer Hauptfunktion synchronisiert und über Hauptfunktionen hinweg harmonisiert werden. Neben der funktionalen Beschreibung der Car-to-X Funktionen, konnte die Wirkung beschrieben werden, auf dessen Grundlage die letztendliche Evaluation durchgeführt werden kann.

3.2 Architekturkonzept eines kooperativen Car-to-X Systems

Abbildung 3.1 zeigt die Architektur des Gesamtsystems im sim^{TD} Projekt. Grundsätzlich lässt sich das Gesamtsystem in zwei Subsysteme unterteilen:

- Fahrzeugseitiges Subsystem (ITS Vehicle Station (IVS) mit Hardware- und Softwarekomponenten im Fahrzeug)
- Infrastrukturseitiges Subsystem (ITS Roadside Station (IRS) und Versuchszentrale ITS Central Station (ICS) mit Anbindung zu den Verkehrszentralen der Stadt Frankfurt und des Landes Hessen)

Die sim^{TD} Gesamtarchitektur sieht für die Kommunikation zwischen und innerhalb der verschiedenen Domänen drei Kanäle vor:

- Direkte Kommunikation zwischen Fahrzeugen und zwischen Fahrzeugen und der Infrastruktur über ITS-G5A (IEEE 802.11p)
- Direkte Kommunikation weiterhin über Consumer WLAN (IEEE 802.11b/g)
- IP-basierte Kommunikation über GPRS, EDGE, UMTS oder HSPA

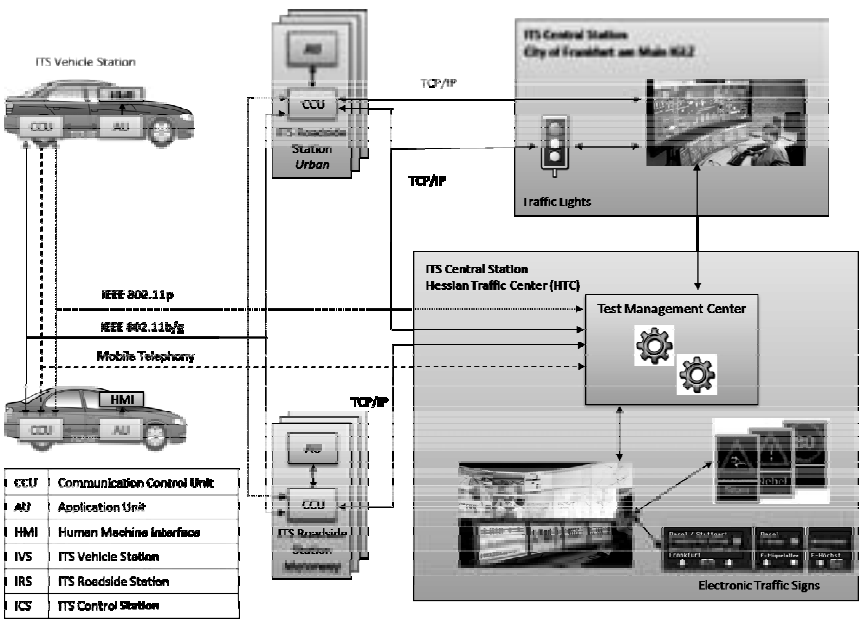


Abbildung 3.1: Gesamtarchitektur in sim^{TD}

Sowohl IVS als auch IRS sind in zwei wesentliche Teilsysteme aufgeteilt:

- Communication Control Unit (CCU) übernimmt die Aufgaben der Kommunikation (ITS-G5A, WLAN, UMTS, CAN).
- Application Unit (AU) beinhaltet Software der Applikationen (Hauptfunktion und Funktionen), Systemkomponenten und OSGi Komponenten.

Abbildung 3.2 zeigt die Aufteilung in Systemkomponenten, die den Teilsystemen CCU und AU auf der IVS zugeordnet sind. Es ist zu erkennen, dass die einzelnen Funktionen, die ein zentrales Element auf der AU darstellen, ebenfalls als eine Systemkomponente beschrieben sind. Die Funktionen sind in Hauptfunktionen geclustert und im sim^{TD} Architekturkonzept voneinander unabhängig. Sie können nicht direkt aufeinander zugreifen. Funktionalitäten, auf die alle Hauptfunktionen zugreifen, sind als separate Systemkomponenten extrahiert (z.B. Umfeld-Tabelle oder Communication Client).

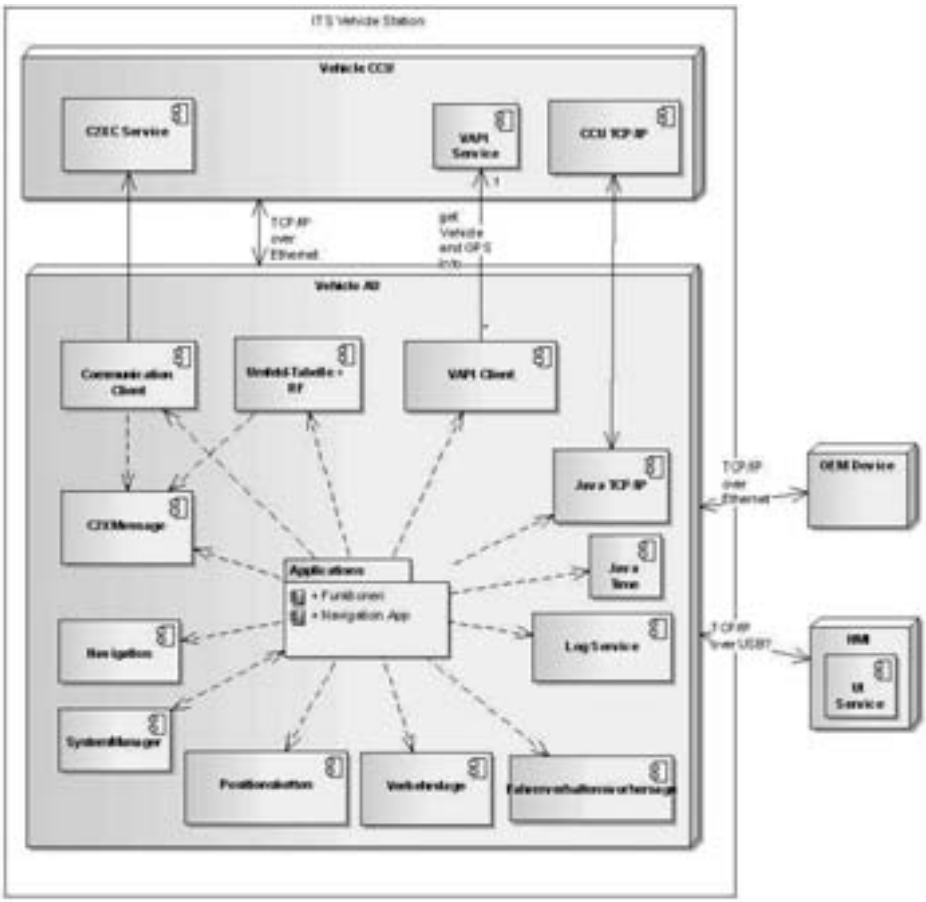


Abbildung 3.2: sim^{TD} Teilsystem IVS

Abbildung 3.3 stellt die Hauptfunktion 2.2 (HF2.2) Fahrerassistenz in einer detailierten Funktionsebene dar. In der Abbildung sind beispielhaft einzelne Funktionen und die Zugriffe auf Systemkomponenten ausgehend von der HF2.2 ersichtlich. Weiterhin sind gemeinsame Funktionalitäten innerhalb der HF2.2 identifiziert und als Basisdienste (z.B. HMI-Koordinator) ausgegliedert. Die Funktionen innerhalb einer Hauptfunktion sind somit voneinander unabhängig. Zugriffe innerhalb einer Hauptfunktion erfolgen ausschließlich über Basisdienste oder über externe Systemkomponenten.

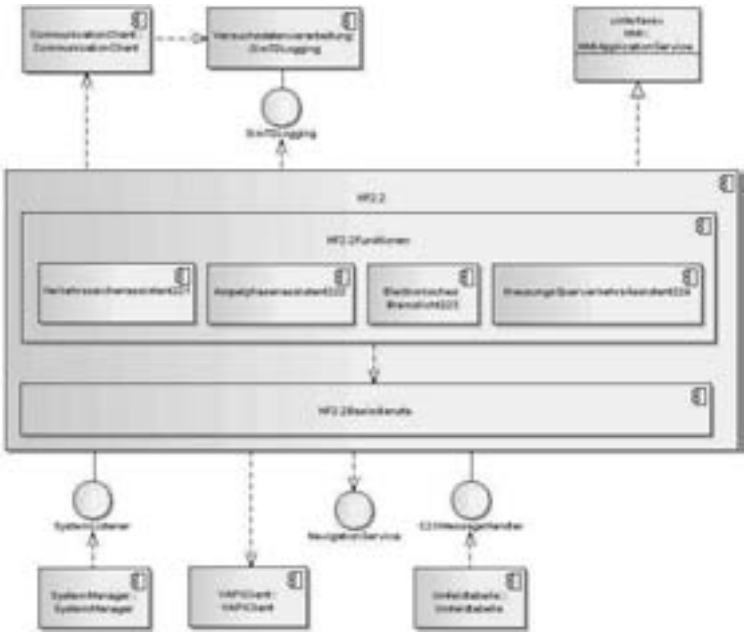


Abbildung 3.3: sim^{TD} Systemkomponente Hauptfunktion 2.2

4 Fazit und Ausblick

Im Kontext der Fahrzeugfunktionsentwicklung ist Car-to-X ein prägnantes Beispiel für eine zunehmend domänenübergreifende Vernetzung. Deutlich wird diese Vernetzung u.a. im Forschungsprojekt sim^{TD}, in dem viele unterschiedliche Wirtschaftsbereiche gemeinsam an der Entwicklung eines hoch vernetzten Car-to-X Systems arbeiten. Dabei ist neben der inhaltlichen Vernetzung der zu entwickelten Funktionen, die organisatorische Vernetzung der heterogenen Entwicklungsteams eine Herausforderung. Ziel muss es sein, über die Identifizierung von inhaltlichen und organisatorischen Gemeinsamkeiten eine geringe Kopplung und hohe Kohäsion zu erreichen.

Literaturverzeichnis

[Sim10] http://www.simtd.de

[LA08] Lübke, Andreas: Der aktuelle Stand der Car-to-X Kommunikation, VDE Kongress, München, 2008

[BD06] Bortz, J.; Döring, N.: Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler. Berlin: Springer, 2004.

[JD08] Jiang, D. Delgrossi, L.: IEEE 802.11p: Towards an International Standard for Wireless Access in Vehicular Environments, Vehicular Technology Conference, 2008

Test von verteilten C2X-Applikationen

Jürgen Grossmann, Carsten Neumann, Andreas Hinnerichs,
Horst Rechner, Thomas Hecker, Ilja Radusch

{juergen.grossmann | carsten.neumann | andreas.hinnerichs |
horst.rechner | thomas.hecker | ilja.radusch}@fokus.fraunhofer.de

Abstract: Zukünftige Fahrzeuggenerationen werden, stärker noch als heute, in eine umfassende Kommunikationsinfrastruktur eingebunden sein, die den Datenaustausch sowohl zwischen Fahrzeugen wie auch mit der Verkehrsinfrastruktur ermöglicht. Auf Basis einer solchen Car to X (C2X) Infrastruktur lassen sich intelligente Telematik- und Fahrassistenzsysteme realisieren, die den Straßenverkehr insgesamt effizienter, sicherer und komfortabler machen werden. Voraussetzung dafür ist eine hohe Qualität und Zuverlässigkeit dieser Systeme.

Die systematische Qualitätssicherung solcher Systeme bringt eine Reihe neuer Herausforderungen mit sich, die in ihrer Kombination bisher einzigartig sind und besondere Anforderungen an die Beschreibung der Testfälle sowie an die für die Prüfung verwendeten Prüfsysteme stellen. In diesem Papier stellen wir eine flexible Referenzarchitektur für einen C2X-Prüfstand vor, mit dem sich C2X-Systeme effektiv und abgestimmt auf ihre spezifischen Kommunikationsanforderungen testen lassen. Ausgehend von den speziellen Prüfanforderungen der C2X-Systeme zeigen wir weiterhin, wie sich Spracherweiterungen für die standardisierte Testbeschreibungssprache TTCN-3 motivieren lassen und zeigen ihre Integration mit der zuvor definierten Prüfstandsarchitektur. Die in diesem Papier vorgestellten Ergebnisse sind im Rahmen der Projekte sim^{TD} und TEMEA entwickelt worden.

1 Motivation

C2X-Systeme kommunizieren mittels komplexer Nachrichten. Sowohl die Nachrichtenübermittlung wie auch der Nachrichteninhalt hängen vom Ort und Zustand des Fahrzeuges ab und weisen einen Komplexitätsgrad auf, der bisher im Fahrzeugbereich unüblich gewesen ist. Die Nachrichtenformate und Nachrichtenprotokolle für den europäischen Raum werden derzeit von der ETSI [ETS09a] standardisiert. Grundsätzlich wird zwischen sogenannten Cooperative Awareness Messages (CAM) und sog. Decentralized Environmental Notification Messages (DENM) unterschieden. Während CAMs von den Fahrzeugen regelmäßig versandt werden und den aktuellen Fahrzeugstatus an die Umgebung propagieren, sind DENMs ereignisbezogene Nachrichten, die andere Verkehrsteilnehmer über konkrete Ereignisse (z.B. Hindernisse, Baustellen, etc.) informieren sollen.

Der komplexe Zusammenhang zwischen dem dynamischen Fahrzeugstatus, der Nachrichtenverteilung und dem Nachrichteninhalt führt dazu, dass C2X-Applikationen i.d.R. schwer zu prüfen sind. Neben der Stimulation des Systems mit Nachrichten muss zusätzlich

und kontinuierlich der Fahrzeugstatus (d.h. Geschwindigkeit, Beschleunigung, Position etc.) gesetzt werden.

2 Lösungsansatz

Um der Komplexität der Anwendungen gerecht zu werden, wurden in der Vergangenheit für die Prüfung von C2X-Systemen häufig Verkehrssimulatoren [SMR08, EOSK05, GM09] eingesetzt. Diese wurden benutzt, um die notwendigen Positions- und Geschwindigkeitsdaten für die Fahrzeuge zu liefern. Während sich die überwiegende Anzahl dieser Ansätze auf die Prüfung der C2X-Software beschränken, adressieren wir in diesem Artikel analog zu [SGR+09] den Test vollständig integrierter C2X-Systeme, bestehend aus der Applikationssoftware, der Basissoftware sowie der Zielhardware. Um zusätzlich bei der Auswertung der Tests einen hohen Grad an Automation gewährleisten zu können und die Nachvollziehbarkeit, Dokumentation und Wiederholbarkeit der Tests sicherzustellen, setzen wir für die Spezifikation der Tests auf die formale Sprache TTCN-3 [ETS09b] bzw. TTCN-3 embedded [SG08, GSW08, TEM09a]. TTCN-3 embedded wurde im Rahmen des Projekts TEMEA [TEM09b] entwickelt und bietet den vollen Umfang von TTCN-3, d.h. die Unterstützung für die Prüfung kommunikationsbasierter, ereignisgesteuerter Systeme. Darüber hinaus stellt es Erweiterungen für den Test kontinuierlicher, d.h. auf Sensordaten basierender Steuer- und Regelsysteme, zur Verfügung.

Im Folgenden werden wir die Definition einer flexiblen Referenzarchitektur für einen C2X-Prüfstand vorstellen und zeigen, wie sich C2X-Systeme auf einem solchen Prüfstand durch die Verwendung von TTCN-3 embedded systematisch und effektiv testen lassen.

3 Referenzarchitektur für einen C2X-Prüfstand

Am Fraunhofer Institut FOKUS wird eine Prüfstandsarchitektur entwickelt, die es erlaubt, sowohl einzelne C2X-Applikationen wie auch die Kommunikation zwischen verschiedenen C2X-Systemen zu testen, bevor die Systeme in reale Fahrzeuge verbaut werden. Abbildung 1 zeigt eine solche Architektur, die als Grundlage für den Prüfstandsaufbau in den Projekten sim^{TD} [sim] und PreDrive [PRE] verwendet wurde.

Grundsätzich besteht ein solcher Prüfstand aus einem Prüfstandsrechner und dem zu testenden System inklusive zusätzlicher Instrumentierung[1]. Das zu testende System kann, abhängig vom Testziel, aus einer oder mehreren On Board Units (OBUs) und Infrastrukturkomponenten, wie z.B. Road Side Units (RSUs), bestehen. Die *Testing API* auf dem Test-PC bietet eine gemeinsame Schnittstelle zum Einspielen aller Stimulationsdaten in die angeschlossenen OBUs und RSUs sowie zur Analyse des Systemverhaltens. Über diese Schnittstellen können verschiedene Programme, Testwerkzeuge wie auch Simulatoren in das Testsystem integriert werden. Die Gegenstellen der Testing API auf den OBUs und

[1] Die zum Prüfstand gehörenden Komponenten sind in der Abbildung fett hervorgehoben.

RSUs sind die *C2XAPI* und die *Vehicle API (VAPI)* [EK04]. Die C2XAPI gehört zum Testsystem und wird ausschließlich zu Testzwecken auf die Zielsysteme installiert. Sie dient dazu, C2X-Nachrichten, die über die Testing API in das Testsystem eingespielt werden, an die internen Schnittstellen des C2X-Systems zu übermitteln. Die Übermittlung einer Nachricht erfolgt entweder an den *COM Client*, der die Nachricht über die Kommunikationsschnittstelle versendet oder an die *Local Data Map (LDM)*, die eine Nachricht für die lokale Verwendung innerhalb einer OBU verfügbar macht.

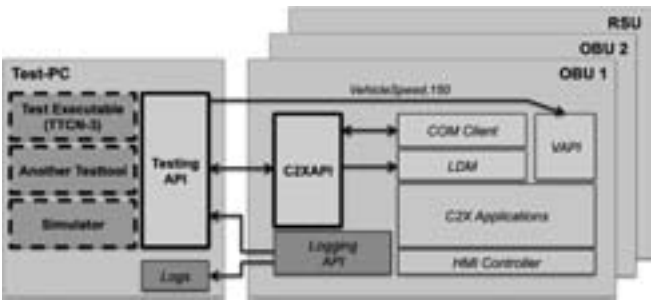


Abbildung 1: Architektur des Prüfstands

Während die C2XAPI für die Verarbeitung der C2X-Nachrichten zuständig ist, wird die VAPI verwendet, um den aktuellen Fahrzeugstatus zu setzen. Im Gegensatz zu den C2X-Nachrichten, die i.d.R. sporadisch gesendet werden, setzt sich der Fahrzeugstatus aus sich kontinuierlich ändernden Größen (z.B. Geschwindikeit, Längsbeschleunigung, Querbeschleunigung etc.) zusammen. Um ein interaktives Testen zu ermöglichen, ist zusätzlich ein Rückkanal vorgesehen, der über den COM Client eingehende und ausgehende Nachrichten sowie Präsentationsaufträge vom *HMI Controller* (letzteres über Verwendung der *Logging API*) an die Testing API zurückmeldet.

Der Prüfstandsrechner ist grundsätzlich in der Lage, mehrere OBUs bzw. RSUs über Ethernet mit synthetischen Eingabedaten zu stimulieren. Die Testing API bietet in diesem Zusammenhang eine werkzeugunabhängige Schnittstelle, die eine Integration mehrerer Testwerkzeuge ermöglicht und die Verteilung der Daten an das System unter Test übernimmt. Die über die Testing API ins System eingebrachten Daten werden von den Fahrzeugfunktionen (*C2X Applications* in der Abbildung) wie reale Daten behandelt und erlauben so einen systematischen Test der integrierten C2X-Funktionen auf dem Prüfstand.

Als Ergebnis des Tests werden sämtliche Aktionen und Daten der Fahrzeugfunktionen, LDM, COM Client, HMI Controllerund gegebenenfalls VAPI geloggt. Die automatische Auswertung dieser Logs kann auf dem Test-PC über TTCN-3 geschehen.

4 TTCN-3 embedded

TTCN-3 embedded ist eine Weiterentwicklung des bereits in der Telekommunikationsbranche etablierten Standards TTCN-3. Die neuen TTCN-3 embedded Konzepte erlau-

ben die Spezifikation von Tests, die kontinuierlich physikalische Größen wie Geschwindigkeit, Beschleunigung, Temperatur etc. in ihrer Abhängigkeit von der Zeit adressieren können. Zu diesem Zweck werden indizierte Datenstromports zur Verfügung gestellt, die einen wahlfreien Zugriff auf die aktuellen Werte sowie die Historie eines Signalverlaufs ermöglichen. Zusätzlich erlauben angepasste Kontrollstrukturen die Spezifikation getakteter, zyklischer Zuweisungen. Listing 1 zeigt einen TTCN-3 embedded Programmabschnitt, der die Geschwindigkeit eines Fahrzeugs in Abhängigkeit von der Ausführungszeit kontinuierlich erhöht (Zeile 2). Zur selben Zeit wird geprüft, ob die Umdrehungszahl des Motors unter einer gegebenen Schwelle (hier 5000.0 U/min, Zeile 3) verbleibt. Stimulation und Prüfung erfolgen in einer zeitgesteuerten Schleife (Zeilen 1,4), deren Schrittweite zuvor separat festgelegt werden kann. Der Programmabschnitt wird verlassen, wenn die Geschwindigkeit den Wert von 180.0 Km/h erreicht.

Listing 1: Konzepte zur Beschreibung und Auswertung kontinuierlicher Signale

```
cont{                                                              1
 VelocityStream.value:= 10.0 * duration;                           2
  assert(EngineSpeed.value <= 5000.0)}                             3
until (VelocityStream.value >= 180.0)                              4
setverdict(pass);                                                  5
```

Neben physikalischen Größen lassen sich mit TTCN-3 embedded nach wie vor die klassischen Eigenschaften kommunikationsbasierter Systeme prüfen. Zu diesem Zweck steht der volle Funktionsumfang des aktuellen TTCN-3 Standards zur Verfügung, u.a. ein mächtiges Typsystem zur Spezifikation von Nachrichtenformaten sowie die Möglichkeit Musterausdrücke (Templates) zu definieren, mit denen sich eingehende Nachrichten flexibel auf ihren Inhalt prüfen lassen. Listing 2 zeigt die Spezifikation eines Musterausdrucks, der eingehende Nachrichten filtert und nur Nachrichten durchlässt, deren Feld `_messagekind` mit dem Wert `'CAM'` belegt ist. Das heißt DENM Nachrichten werden unterdrückt.

Listing 2: Definition der Nachrichtenformate

```
template C2XMessage CAMMessage :={ _messageKind:="CAM",            1
                                   _timestamp:=?, _pload:=?}       2
```

5 Der Test von C2X-Systemen mit TTCN-3 embedded

Betrachten wir ein vereinfachtes Szenario aus der Fahrzeug-zu-Fahrzeug Kommunikation. Ein mit Kommunikationstechnologie ausgerüstetes Fahrzeug sendet bei einer gegebenen Toleranz von 10% zyklisch alle 5 Sekunden seinen Status per CAM an andere Fahrzeuge. Ändert sich hingegen die Geschwindigkeit des Fahrzeugs mit einem Wert größer $2m/s^2$, wird sofort eine Statusmeldung abgesetzt. Wir prüfen, ob bei sich ändernden Geschwindigkeitswerten die Nachricht ad-hoc verschickt wird.

CAMs werden in TTCN-3 durch den Datentyp C2XMessage modelliert und haben eine Nachrichtenkennung (`_messageKind`), die sie als CAM Nachricht (siehe Listing 2) identifizieren. Wir wählen einen Prüfaufbau, der der Referenzarchitektur in Abbildung 1 ent-

spricht und aus zwei OBUs besteht. Eine für den Prüfling (OBU1) und eine als Empfänger (OBU2), sprich als Bestandteil des Testsystems.

Listing 3: CAM Test

```
testcase PostionMessageTest_AdHocCAM runs on TTCN3Tester{                    1
                                                                             2
  cont{OBU1_VAPI_LongAccel.value:= 0.4;}                                     3
  until{                                                                     4
   [duration> 4.5 and duration < 5.5]                                        5
    OBU2_COM_IN.receive(posMessage){repeat}                                  6
   [] OBU2_COM_IN.receive(CAMMessage){setverdict(fail); stop}                7
   [] OBU2_COM_IN.receive(?){continue}                                       8
   [] [now >= 30.0]{}}                                                       9
                                                                             10
 cont{OBU1_VAPI_LongAccel.value:= 2.5;}                                      11
 until{                                                                      12
   [] OBU2_COM_IN.receive(CAMMessage){setverdict(pass)}                      13
   [] OBU2_COM_IN.receive(?){continue}                                       14
   [duration > 0.5] {setverdict (fail)}}                                     15
}                                                                            16
```

Das Setzen der Größen für den Fahrzeugstatus im Prüfling erfolgt über die VAPI und wird im Gegensatz zur CAM Nachricht als kontinuierlicher Stimulus modelliert. Das Testszenario in Listing 3 ist zweistufig.

Im ersten Abschnitt des Tests (Zeilen 3 - 9) wird die Normalsituation, d.h. das Verhalten bei einer nahezu gleichförmigen Geschwindigkeit, geprüft. Das Fahrzeug wird durch Beschleunigungsdaten stimuliert, die gemäß einer zuvor gesetzten Samplingrate zyklisch gesetzt wird (Zeile 3). Am COM Client eingehende Nachrichten werden entsprechend ihres Nachrichteninhalts gefiltert. Zeitkorrekte zyklische CAM-Nachrichten sowie beliebige DENM-Nachrichten werden geschluckt (Zeile 5,6 sowie Zeile 8). Werden azyklische CAM-Nachrichten empfangen, so schlägt der Testfall fehl (Zeile 7). Nach einer Dauer von 30 Sekunden (Zeile 9) wechselt der Testfall in den zweiten Abschnitt (Zeile 11-15), in dem für eine halbe Sekunde auf die Ankunft einer Ad-hoc CAM-Nachricht gewartet wird. Wird keine CAM-Nachricht empfangen, schlägt auch hier der Testfall fehl.

6 Zusammenfassung

In diesem Kurzartikel haben wir die grundlegenden Herausforderungen für das Testen interaktiver Fahrzeugsysteme beschrieben, den Aufbau eines Prüfstandes zum Testen von C2X-Systemen und den Einsatz von TTCN-3 embedded als Spezifikationssprache für die formale Definition ausführbarer Tests skizziert. Die dargestellte Prüfstandsarchitektur erlaubt den systematischen Test von C2X-Systemen und ist flexibel auf verschiedene Prüfszenarien anpassbar (Test einzelner C2X-Applikationen auf einer OBU, Test der C2X-Kommunikation, Test eines C2X-Systems über mehrere OBUs). Die Verwendung von TTCN-3 erlaubt die Automatisierung der Tests und sorgt zusätzlich für die bei sicherheitskritischen Systemen notwendige Formalisierung und Wiederholbarkeit der Tests. Diese Arbeit wurde im Rahmen des Projektes sim^{TD} durch die Bundesministerien für Wirtschaft

und Technologie (BMWi) und Bildung und Forschung (BMBF) gefördert und durch das Bundesministerium für Verkehr, Bau und Stadtentwicklung (BMVBS) unterstützt sowie im Projekt TEMEA durch Mittel finanziert, die aus dem Europäischen Fonds für Regionale Entwicklung (EFRE) stammen.

Literatur

[EK04] Oliver Eggers and Jens Krüger. Das Fahrzeug als standardisiertes, generisches Objektmodell, in:. In *ÏMA 2004 - Informationssysteme für mobile Anwendungen", Beiträge zum gleichnamigen 2. Braunschweiger Symposium am 20. und 21. Oktober 2004*, pages 205 –218, Braunschweig, 2004.

[EOSK05] Stephan Eichler, Benedikt Ostermaier, Christoph Schroth, and Timo Kosch. Simulation of Car-to-Car Messaging: Analyzing the Impact on Road Traffic. In *MASCOTS '05: Proceedings of the 13th IEEE International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems*, pages 507–510, Washington, DC, USA, 2005. IEEE Computer Society.

[ETS09a] ETSI. ITS Activity Report 2009; http://portal.etsi.org/its/ActivityReport2009.asp, 2009.

[ETS09b] ETSI: ES 201 873-1 V4.1.1. Methods for Testing and Specification (MTS). The Testing and Test Control Notation Version 3, Part 1: TTCN-3 Core Language, Febr. 2009.

[GM09] Markus Glaab and Alois Mauthofer. NEW TEST AND EVALUATION METHODS FOR FUTURE CAR2X COMMUNICATION BASED DRIVER ASSISTANCE. In *21st Int. Technical Conference on the Enhanced Safety of Vehicles (ESV)*, Stuttgart, 2009.

[GSW08] Jürgen Großmann, Ina Schieferdecker, and Hans-Werner Wiesbrock. Modeling Property Based Stream Templates with TTCN-3. In Kenji Suzuki, Teruo Higashino, Andreas Ulrich, and Toru Hasegawa, editors, *TestCom/FATES*, volume 5047 of *Lecture Notes in Computer Science*, pages 70–85. Springer, 2008.

[PRE] PRE-DRIVE C2X. http://www.pre-drive-c2x.eu/index.dhtml/394bd0148054d449805e/-/deDE/-/CS/-/. Die Website wurde zuletzt am 22.04.2010 besucht.

[SG08] Ina Schieferdecker and Jürgen Großmann. Testing hybrid control systems with TTCN-3: an overview on continuous TTCN-3. *STTT*, 10(4):383–400, 2008.

[SGR+09] Oliver Sander, Benjamin Glas, Christoph Roth, Jurgen Becker, and K. D. Muller-Glaser. Testing of an FPGA Based C2X-Communication Prototype with a Model Based Traffic Generation. In *RSP '09: Proc. of the 2009 IEEE/IFIP Int. Symposium on Rapid System Prototyping*, pages 68–71, Washington, DC, USA, 2009. IEEE Computer Society.

[sim] sim^{TD} : Sichere Intelligente Mobilität - Testfeld Deutschland. http://www.simtd.de. Die Website wurde zuletzt am 20.04.2010 besucht.

[SMR08] Björn Schünemann, Kay Massow, and Ilja Radusch. A Novel Approach for Realistic Emulation of Vehicle-2-X Communication Applications. In *VTC Spring*, pages 2709–2713. IEEE, 2008.

[TEM09a] TEMEA. Draft Proposal for Continuous TTCN-3 Concepts; http://t-ort.etsi.org/file_download.php?file_id=2282&type=bug, 2009.

[TEM09b] TEMEA. TEMEA Project (Testing Methods for Embedded Systems of the Automotive Industry), founded by the European Community (EFRE), http://www.temea.org, 2009.

Transfer von Prozessen des Software-Produktlinien Engineering in die Elektrik/Elektronik-Architekturentwicklung von Fahrzeugen

Martin Jaensch, Bernd Hedenetz, Markus Conrath, Klaus D. Müller-Glaser*

E/E-Architektur (GR/PSA)
Daimler AG
HPC: G007-BB
71059 Sindelfingen
{martin.jaensch, bernd.hedenetz,
markus.m.conrath}@daimler.com

Institut für Technik der
Informationsverarbeitung (ITIV)*
Karlsruher Institut für Technologie
Vincenz-Prießnitz-Straße 1
76131 Karlsruhe
klaus.mueller-glaser@kit.edu

Abstract: In der Entwicklung von Elektrik/Elektronik-Architekturen[1] in Fahrzeugen gibt es derzeit den starken Trend hin zu dem Einsatz von Modulkomponenten, die innerhalb eines Automobilherstellers modellübergreifend, teilweise sogar herstellerübergreifend eingesetzt werden. Ziel dieses Ansatzes ist die Reduzierung von Komponentenvarianten durch Wiederverwendung bei gleichzeitiger vereinfachter Ableitung von Varianten einer Baureihe[2].

In der SW-Entwicklung stellten sich ähnliche Fragestellungen bei der Wiederverwendung und Variantenbildung von SW-Komponenten, so dass daraus die Prozesse des SW-Produktlinien Engineering entstanden. In diesem Artikel wird analysiert, welche Prozesse für die Wiederverwendung und Variantenbildung in der E/E-Architekturentwicklung übernommen und adaptiert werden können.

1 Stand der Technik – SW-Produktlinien Engineering

In der Produktentwicklung verspricht die Idee der Entwicklung einer Basis von gemeinsamen Artefakten[3] und deren Wiederverwendung in davon ableitbaren Varianten eine höhere Effizienz im Entwicklungsaufwand [HP02]. Der SW-Produktlinienansatz ist genau ein solches Konzept, welches eine systematische Wiederverwendung von gemeinsamen SW-Komponenten in mehreren Produktenvarianten ermöglicht.

In der Abb.1 ist der Prozess des SW-Produktlinien Engineering gem. [PBL05] dargestellt. In der oberen Hälfte der Abbildung befindet sich der Prozess des Domain Engineering, in der unteren Hälfte der Prozess des Application Engineering.

[1] Elektrik/Elektronik wird nachfolgend E/E abgekürzt
[2] Fahrzeugmodell eines Automobilherstellers, z.B. Mercedes-Benz E-Klasse.
[3] Ergebnis eines Prozesses im Domain oder Application Engineering; umfasst Anforderungen, Architektur, Komponenten oder Tests [PBL05].

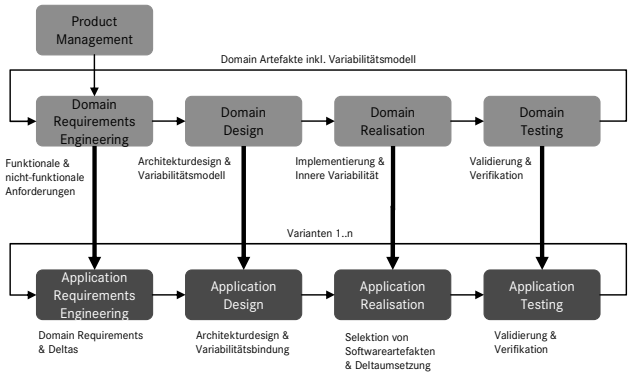


Abbildung 1: Prozess des SW-Produktlinie Engineering gem. [PBL05]

Im Prozess des Domain Engineering werden die allgemeingültigen Produktlinien-Aspekte und Variationsmöglichkeiten, d.h. die gemeinsamen und variablen Artefakte der Produktlinie, definiert und entwickelt.

Das Application Engineering befasst sich mit der Erstellung konkreter Varianten der Produktlinie unter Verwendung der vom Domain Engineering gelieferten Artefakte, z.B. der Architektur. Ziel ist ein hoher Grad der Wiederverwendung dieser Artefakte bei der Variantenentwicklung und damit die Ausnutzung von gemeinsamen Anforderungen zur Bildung der Produktvariante.

2 Anforderungen der E/E-Architekturentwicklung

Die E/E-Architekturentwicklung in der Automobilindustrie ist durch spezifische Rahmenbedingungen geprägt. Diese werden in diesem Kapitel nachfolgend betrachtet.

Variantenvielfalt der E/E-Architekturen: Ein Automobilhersteller besitzt unterschiedliche E/E-Plattformen[4] bedingt durch unterschiedliche Anforderungsprofile der Fahrzeuge. Es werden von diesen E/E-Plattformen die E/E-Architekturen der unterschiedliche Baureihen (in Abb.2: BR-übergreifende Varianz) sowie Derivate[5] (in Abb.2: BR-interne Varianz) für verschiedene Aufbauvarianten, Ausstattungsgrade oder Angebotsmärkte abgeleitet. Der Modulbaukasten (siehe Kap. 2 nachfolgend) enthält unterschiedliche Ausprägungen der Modulkomponenten (in Abb.2: Modul-interne Varianz), welche jeweils in verschiedenen E/E-Plattformen integriert werden.

Die zeitliche Varianz der E/E-Architekturen ist bedingt durch die Weiterentwicklung der Modulkomponenten des Modulkastens, durch die Evolution der E/E-Plattformen bei neuen Innovationen oder weiterentwickelten Modulkomponenten sowie durch den Lebenszyklus der Baureihen (in Abb.2: zeitl. Varianz).

[4] Eine E/E-Plattform ist eine Basis von Komponenten, aus denen eine wiederverwendbare E/E-Architektur für mehrere Baureihen gebildet wird (entspricht einer Kernarchitektur gem. [PBL05]).
[5] Derivat ist eine Karosserievariante innerhalb einer Baureihe, z.B. Limousine/ Cabrio/ Coupe/ etc.

Aufgrund der Variantenvielfalt wird zur Aufwandsreduzierung in der zukünftigen E/E-Architekturentwicklung durch die Einbindung eines Modulbaukastens ein hoher Grad der Wiederverwendung von Modulkomponenten angestrebt.

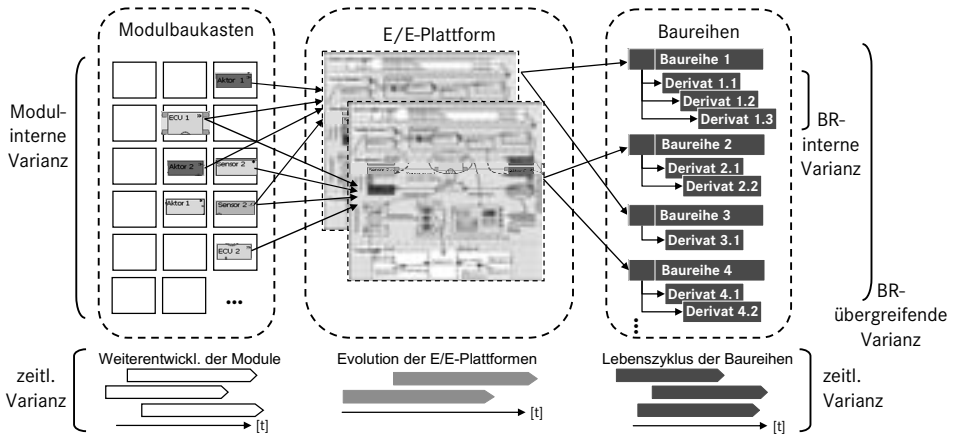


Abbildung 2: Varianz und Abhängigkeiten in der E/E-Architekturentwicklung

Modularisierung von Komponenten der E/E-Architekturen: Die Modularisierung ist für die Automobilindustrie ein wichtiger Ansatz, um mit Skaleneffekten durch die erhöhte Wiederverwendung von Komponenten den Entwicklungsaufwand zu begrenzen.

Das Fahrzeug wurde in Modulkomponenten unterteilt und in einem Modulbaukasten gruppiert. Dabei zählten als Kriterium der Zerlegung die montageorientierten Aspekte der Produktion, z.B. dass eine Modulkomponente durch eine definierte Arbeitseinheit oder einen definierten Arbeitsschritt montiert wird, oder dass eine Modulkomponente in der Logistik als Einheit gesehen wird. Somit bilden die Modulkomponenten fahrzeugbauliche Einheiten und bezeichnen nicht wie in der SW-Technik von einer Plattform abgeleitete Einheiten.

Durch die Einbindung des Modulbaukastens in die E/E-Architekturentwicklung wird eine Wiederverwendung der Modulkomponenten erleichtert. Zurzeit sind die frühe Phase der E/E-Architekturentwicklung und die Modularisierung strukturell entkoppelt.

Entwicklungsphasen der E/E-Architekturen: Die E/E-Architekturentwicklung teilt sich in die Konzeptphase sowie in die Serienentwicklung auf. In der Konzeptphase wird der E/E-Architekturentwurf inklusive der Untersuchungen von alternativen Realisierungen der E/E-Architektur und E/E-relevanten Innovationen durchgeführt. Dabei muss die E/E-Plattform über den gesamten E/E-Architekturlebenszyklus für neue Innovationen erweiterbar und gegenüber technologischen Veränderungen belastbar ausgelegt sein. Diese „Robustheit“ der E/E-Plattform muss in der Konzeptphase virtuell abgesichert werden, da sie vor der Erstellung von Spezifikationen (Lastenheften) der abgesicherten Innovationen in der Serienentwicklung stattfindet.

3 Konzept des E/E-Produktlinien Engineering

In diesem Kapitel wird das SW-Produktlinien Engineering gem. [PBL05] für eine E/E-Produktlinie[6] gem. den folgenden Anforderungen aus Kap.2 angepasst:
- Einbindung der Modulkomponenten in die E/E-Architekturentwicklung,
- Absicherung der E/E-Plattform in einer frühen Phase des Plattformentwurfs.

Die Abb.3 zeigt die Anpassung des SW-Produktlinien Engineering an die E/E-Architekturentwicklung. Die grau eingefärbten Prozesse entsprechen dem SW-Produktlinien Engineering in Abb.1. Die schwarz bzw. weiß eingefärbten Prozesse sind Erweiterungen resultierend aus den Anforderungen in Kap.2. Die horizontalen Gruppierungen (gestrichelte Linien) verdeutlichen die organisatorische Zugehörigkeit der Prozesse. Das gesamte E/E-Produktlinien Engineering ist mit den Entwicklungsphasen gem. Kap.2 hinterlegt.

3.1 Management der E/E-Produktlinie

Die Kundenanforderungen werden nicht zentral für die gesamte E/E-Produktlinie definiert (gem. *Product Management* in Abb.1), sondern in den Prozessen {*Plattform*, *Baureihen*, *Modul*} *Management* jeweils für die E/E-Plattform, die Derivate und die Modulkomponenten spezifiziert. Dieses hat verschiedene Vorteile, wie z.B. die E/E-Plattform kann strategisch für mehrere Baureihen definiert, neue Anforderungen noch in der Serienentwicklung für die Baureihe bzw. das Derivat abgeleitet sowie die Entkopplung des Entwicklungszyklus der Modulkomponenten unterstützt werden.

3.2 E/E-Plattform

Die E/E-Plattform wird in der Konzeptphase modellbasiert entwickelt [RSB07] und virtuell abgesichert [GMRM08]. Dazu wird im Prozess *Plattform Requirements Engineering* die Umsetzung der Funktionen vom *Plattform Management* festgelegt, d.h. welche Komponente für die E/E-Plattform modelliert wird. Im Prozess *Plattform Design* werden diese Komponenten modelliert, wobei der Modulbaukasten über das *Plattform Requirements Engineering* die zur Auswahl stehenden Modulkomponenten vorgibt.

Im Prozess *Virtuelle Absicherung Plattform* wird die Absicherung der E/E-Plattform durchgeführt. Dazu wird iterativ eine E/E-Plattform im Prozess *Plattform Design* entworfen, auf Konzepttauglichkeit im Prozess *Virtuelle Absicherung Plattform* bewertet und ggf. Optimierungen durchgeführt. Das Ergebnis ist ein Architekturkonzept, welches die technische Machbarkeit einer E/E-Plattform nachweist. Die iterative Architekturoptimierung ist notwendig, um die E/E-Plattform bei Änderungen durch die Weiterentwicklung der Modulkomponenten oder neuen Innovationen anzupassen. Die abgesicherte E/E-Plattform wird als Architekturmodell zur Wiederverwendung in verschiedenen Derivaten dem Prozess *Architektur Design* bereitgestellt.

---

[6] E/E-Produktlinie bezeichnet nachfolgend eine Produktlinie in der E/E-Architekturentwicklung

Die Prozesse *Domain Realisation* und *Domain Testing* aus Abb.1 entfallen hier, da die Umsetzung der Komponenten in die Prozesse des Modulbaukastens verlagert wird.

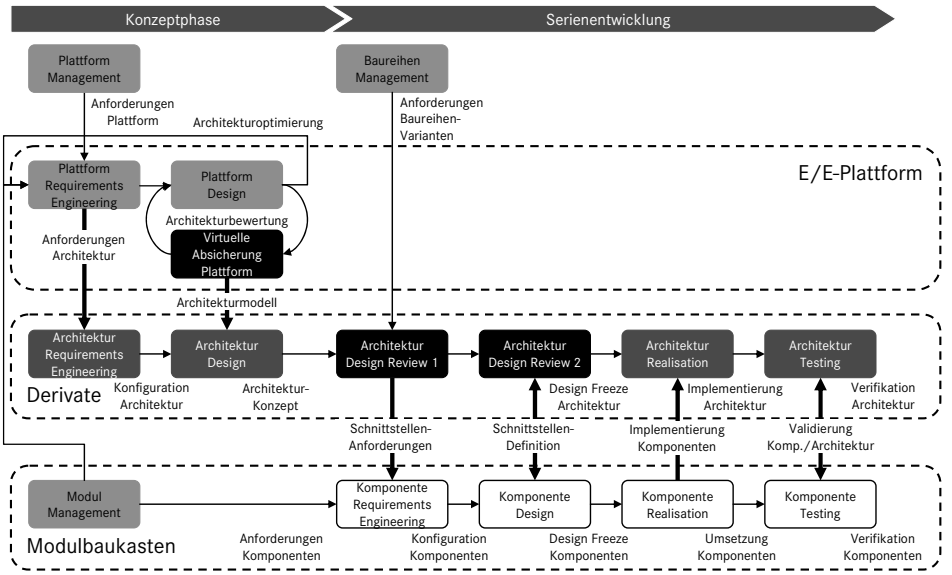


Abbildung 3: Produktlinienansatz für die E/E-Architekturentwicklung

3.3 Derivate

Die Derivate in Abb.3 werden als E/E-Architekturvariante gebildet. Dabei stellt die E/E-Plattform eine abgesicherte E/E-Architektur (in Form eines Modells) und der Modulbaukasten stellt die notwendigen Modulkomponenten in der Serienentwicklung.

Im Prozess *Architektur Requirements Engineering* wird die Konfiguration für die E/E-Architekturvariante festgelegt, d.h. welche Modulkomponenten wiederverwendet werden. Im Prozess *Architektur Design* wird das abgesicherte Architekturmodell aus der E/E-Plattform übernommen und gem. der Konfiguration eine modellbasierte E/E-Architekturvariante für das Architekturkonzept erstellt. Dieses Architekturkonzept wird als Nachweis der Konzepttauglichkeit in die Serienentwicklung übernommen und dort als Vorgabe für den weiteren Spezifikations- und Umsetzungsprozess beachtet.

In der Serienentwicklung sind die beiden eingeführten *Design Review*-Prozesse zum Schnittstellenabgleich mit den Modulkomponenten bei Übernahme in die E/E-Architektur notwendig. Die elektrischen (z.B. Buskommunikation) und ausgewählten physikalischen Schnittstellen (z.B. Bauraum) sind in der *Virtuellen Absicherung Plattform* überprüft worden, so dass nur für weitere Schnittstellen oder Änderungen durch das *Baureihen Management* eine neue Schnittstellendefinition durchgeführt wird. Die Prozesse *Architektur Realisation* sowie *Architektur Testing* entsprechen dem Vorgehen gem. des SW-Produktlinien Engineering.

3.4 Modularisierung

Der Modulbaukasten ist in die E/E-Architekturentwicklung mit dem Ziel der Wiederverwendung von Modulkomponenten gem. Kap.2 eingebunden. Im Prozess *Modul Management* werden die Modulkomponenten verwaltet sowie zur Absicherung der E/E-Plattform in der Konzeptphase bereitgestellt.

In der Serienentwicklung der Derivate werden Schnittstellenanforderungen der E/E-Architektur (z.B. Typ der Busanbindung) an die Modulkomponenten über dem Prozess *Komponente Requirements Engineering* vorgegeben. Bei Änderungen der Anforderungen durch den Prozess *Baureihen Management* wird im Prozess *Komponente Design* die neue Schnittstellen definiert, was gleichzeitig zu einer Weiterentwicklung der Modulkomponenten führt (siehe Abb.2). Die Prozesse *Komponente Realisation* und *Komponente Testing* verwenden die Modulkomponenten gem. dem SW-Produktlinien Engineering aus Kap.1.

4 Zusammenfassung und Ausblick

In der E/E-Architekturentwicklung wird die Reduzierung der Variantenvielfalt von Komponenten mittels deren modulorientierter Wiederverwendung angestrebt. Der eingeführte Modulbaukasten gibt nach dem Baukastenprinzip die Vorgaben für die Wiederverwendung von Modulkomponenten, allerdings ist kein Prozess zur Einbindung in die E/E-Architekturen vorgegeben. Der Ansatz des SW-Produktlinien Engineering hat als Ziel die Wiederverwendung von SW-Komponenten, weshalb dessen Prozesse auf die Eignung für die E/E-Architekturentwicklung untersucht wurden.

Im Kap.3 werden die Prozesse eines E/E-Produktlinien Engineering dargestellt. Die notwendigen Anpassungen und Erweiterung des SW-Produktlinien Engineering wurden gem. den Anforderungen aus Kap.2 sowie der Einführung des Modulbaukasten definiert.

Als nächster Schritt wird die Einbindung des Modulbaukastens in der Konzeptphase umgesetzt und die Eignung der Wiederverwendung von Modulkomponenten in der modellbasierten E/E-Architekturentwicklung der Konzeptphase bewertet.

Literaturverzeichnis

[GMRM08] Gebauer, D.; Matheis, J.; Reichmann, C.; Müller-Glaser, K. D.: Ebenenübergreifende, variantengerechte Beschreibung von Elektrik/Elektronik-Architekturen, Haus der Technik, Juli 2008, S. 142-153.

[HP02] Halmans, G.; Pohl, K.: Modellierung der Variabilität einer Software-Produktfamilie. LNI Vol. 12 - Modellierung in der Praxis, 2002; S. 63-74.

[PBL05] Pohl, K.; Böckle, G.; van der Linden, F.: Software Product Line Engineering - Foundations, Principles, and Techniques. Springer-Verlag, Berlin Heidelberg, 2005.

[RSB07] Ringler, Th.; Simons, M.; Beck, R.; Reifegradsteigerung durch methodischen Architekturentwurf mit dem E/E-Konzeptwerkzeug, VDI-Berichte 2007, Baden-Baden, Okt. 2007.

Hardware-independent Software Development with AUTOSAR

Stefan Bunzel, Khosrau Heidary (Continental); Simon Fürst, Andre Lajtkep (BMW Group); Jürgen Mössinger, Jürgen Cordes (Bosch); Stefan Schmerler, Christian Kühn, (Daimler); Frank Kirschke-Biller, Bernd Frielingsdorf (Ford); Robert Rimkus, Rick Kacel (GM); Alain Gilberg, Bertrand Delord (PSA Peugeot Citroën); Kenji Nishikawa, Hiroyuki Hirano (Toyota); Andreas Titze, Bernd Kunkel (Volkswagen)

AUTOSAR cooperation
Bernhard-Wicki-Str. 3
80636 Munich, Germany
admin@autosar.org

Abstract: This paper describes the development of application software with AUTOSAR. The AUTOSAR architecture and the development methodology enable a hardware independent development of application software, so that higher reuse and increased flexibility and scalability are possible.

1 Introduction

AUTOSAR (AUTomotive Open System ARchitecture) has become a de-facto standard for embedded software in the automotive industry. Since 2003 a cooperation of car manufacturers, suppliers, and other companies from the electronics, semiconductor, and software industry have been working on the development and introduction of this standard [AUTO]. By now many AUTOSAR members apply the standard in their product development and there are even first series products on the road [KG08], [Fü09].

AUTOSAR is a key enabling technology to manage the growing electrics/electronics complexity. It aims to increase the reuse of software components, in particular between different vehicle platforms, and between OEMs and suppliers. It enables the scalability of embedded automotive software to different vehicle and platform variants, the transferability of functions throughout the vehicle network, and the integration of functional modules from multiple suppliers. Therefore the AUTOSAR standard defines an architecture that separates application software from infrastructure related basic software, as depicted in Figure 1. The functional contents of the application software are different and related to the brand identity and the desired characteristics of the car manufacturer, or its system suppliers, whereas the functionality of the basic software is not visible to the customer and thus could been standardized by AUTOSAR in detail.

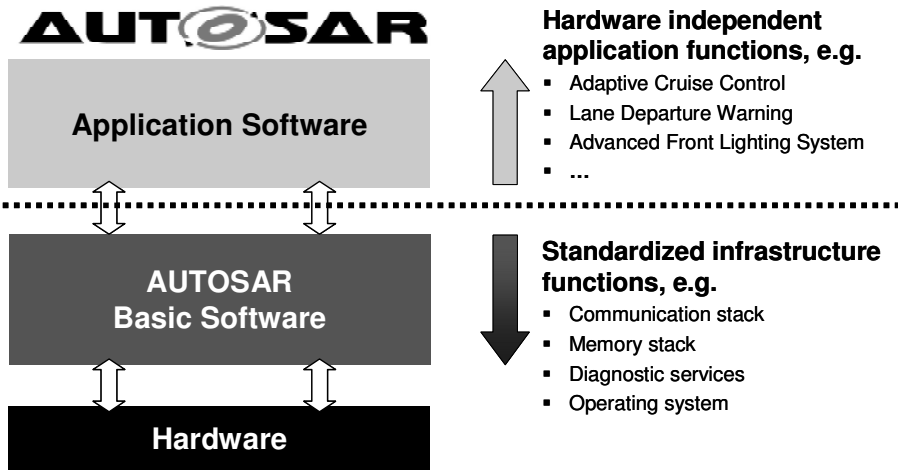


Figure 1: Application software separated from infrastructure functions

Beneath the architecture AUTOSAR has worked out a methodology how to develop software according to this architecture. A third topic of standardized interfaces for typical automotive applications completes the technical approach of the standard.

This paper emphasizes the benefits for application software development with AUTOSAR. Thus it highlights the methodology, explains the hardware independency of application development with the concept of the virtual functional bus (VFB), the assignment of application software to ECUs and the integration, and last but not least it describes the usage of standardized application interfaces.

2 AUTOSAR Methodology

The AUTOSAR methodology describes the major development steps of an overall AUTOSAR system, which means the entire software for a network of interconnected ECUs in a vehicle. In the release 4.0 the methodology is specified as a model yielding a set of HTML-documentation [AM40]. It addresses the wide range of software development from the system-level configuration to the generation of an ECU executable, and it supports a widely decoupled development and implementation of application functionality, as well as a seamless integration and configuration of both, the overall system and its individual ECUs. So the methodology is a guiding framework of how to use the AUTOSAR architecture [BF08].

Figure 2 depicts an abstract top-level view on the methodology. Firstly on the level of functional architecture there is the development of a so called virtual functional bus (VFB) system description. This description is independent of any network topology or deployment of features across multiple ECUs. It contains a component model of all application functions. This functional architecture means a partitioning of functions into components, which can even be a hierarchical model where components are clustered into compositions. The development on this level also yields a data model for the interaction of the components. In addition it optionally deals with timing constraints on VFB level. Subsection 2.1 describes further details on the VFB.

The next level considers the physical architecture of the entire system, i.e. the activity 'design system' leads to a system description that defines the system topology of ECUs, the network, and the mapping of components to ECUs. Before the software for each ECU can be built, the information regarding to this ECU have to be extracted from the system description. This allows building and integrating the software for each ECU separately from the other ECUs. Of course, building the ECU software requires appropriate basic software for the ECU and all application software components mapped to the ECU. The delivery of basic software is out of scope in this paper. By now there is a broad variety of basic software implementations from different vendors for many hardware platforms on the market, so that the basic software and corresponding configuration tools can be regarded as off-the-shelf products. The development of the application software components with the definition of the internal behavior, coding, and implementation are independent from hardware and can be done separately for each component. Subsection 2.2 explains the component development in more detail.

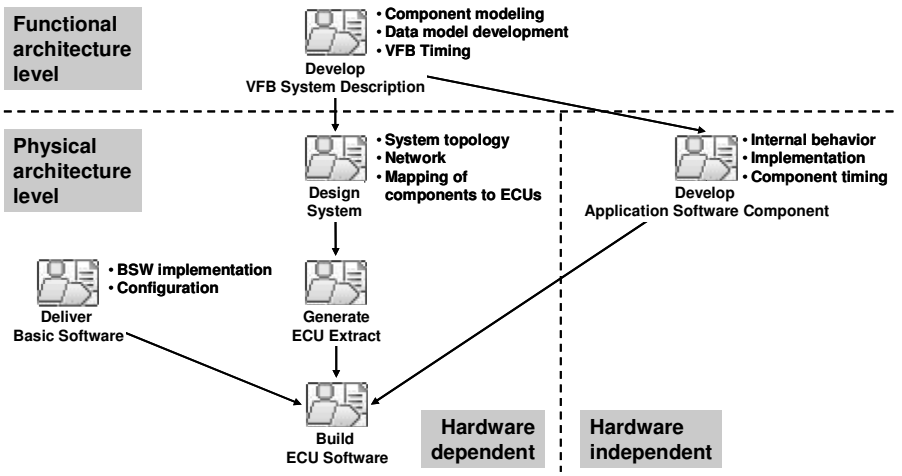


Figure 2: Methodology overview

2.1 Virtual Functional Bus

The Virtual Functional Bus (VFB) is a technical concept that enables the development of the functional architecture of the entire system independent from the actual hardware topology of ECUs and network. The functional architecture is a partitioning or clustering of the application functions into components, i.e. the so called AUTOSAR software components (SWC). In order to formally handle and model SWCs and their interaction each SWC needs a formal description: the SWC description. For the development on VFB level, the SWC description does not have to be complete. At least the data model, i.e. the used interfaces, data types, and services from the basic software must be defined, and in addition a component model with the top-level components is necessary. As mentioned before, components can be grouped hierarchically in arbitrary structure, but this kind of break-down can be derived iteratively. The development of the functional architecture on VFB level means a virtual integration of the applications – and this in a very early phase of the development process.

Figure 3 shows an example of the VFB level architecture. The components interact via ports. Ports can have different characteristics. For instance, they can receive or provide information, or they can implement different communication paradigms like sender-receiver-, or client-server-communication, which is indicated by a dedicated notation. The ports use interfaces and data types that are defined in the data model.

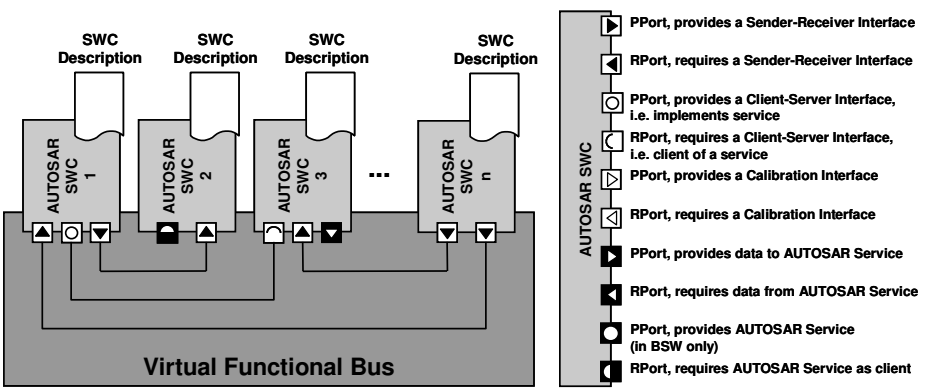


Figure 3: The "virtual functional bus" (VFB) supports a virtual integration

In practice the design of the VFB level architecture is done model-based with a tool. By now several members of AUTOSAR offer VFB level design capabilities in their tools. The design process with such tools is rather component oriented, which means that the data model implicitly will be extended, whenever a component will be extended with a new port, interface, or data type. Such tools furthermore can export the software component description of single components so that the further component development and implementation can run separately.

2.2 Component Development and ECU Integration

An AUTOSAR software component encapsulates an application which runs on the AUTOSAR infrastructure. After the VFB level design yielded the initial software component description, the further component development is independent from other components or from system design steps.

The major task of the component development is the implementation, i.e. the coding of the functionality or algorithms. This could be done either as direct coding in a programming language which is in the AUTOSAR context typically C, or more comfortable by means of a model-based design tool and automatic code generation.

In addition to the implementation, the component developer has to complete the software component description with the definition of runnable entities, events and interrunnable variables. Runnable entities are the smallest code parts schedulable to tasks of the operating system. And a component can consist of multiple runnable entities. With this information in the software component description the interface of the component towards the highest layer in the basic software – towards the Runtime Environment (RTE) – is defined. From the viewpoint of the component, the RTE implements the VFB functionality on a specific ECU.

In case of model-based design the software component description will be completed within that tool, and exported e.g. together with the code generation. The generated code includes even the header towards the RTE so that the component can be compiled.

In case of conventional development, the software component description has to be completed with a specific template editor, which in practice often is part of basic software configuration tools. Such tools commonly can generate the RTE header as well.

As mentioned before, building the ECU software requires appropriate basic software for the ECU and all application software components mapped to the ECU. Although the basic software can be regarded as an off-the-shelf product, it needs a configuration which of course depends on the application. Therefore the basic software configuration tool has to read all software component descriptions of the concerned applications. For a proper configuration it furthermore has to read the ECU extract of the system description and the ECU description. Then all the ECU software, i.e. the configured basic software and all application software components can be compiled and linked.

2.3 Standardized Application Interfaces

AUTOSAR specifies the interfaces of typical automotive applications from all domains regarding syntax and semantics [AI40]. This specification is aggregated in a common table, which contains by now nearly 2500 different ports and more than 500 interfaces. These are clustered into about 40 different compositions, and they use 770 data types and 26 units. The following example shall illustrate the context of these elements.

A composition, e.g. “Mirror Adjustment” contains different ports, e.g. “MirrorPosition” or “MirrorMoveStatus”. Both ports apply the interface “MirrPosnSts1” for describing the actual mirror position, either for status inquiries or for requests. The “MirrPosnSts1” consist of only one data element, i.e. here the data record “MirrAxisPosn1”. The data record has three elements: mirror type, axis type and position. Each element can be either a continuous value or an enumeration. In this example the position is a continuous value of type “perc7” – an unsigned integer with recommended length of 14 bit, unit percent, a resolution of 0,1, a physical range of 0,0 to 1001,0. In contrast the enumeration axis type means 0 no axis, 1 horizontal, and 2 vertical axis.

The entire specification of application interfaces serves as a standard for application software. The usage of these interfaces limits the development effort to adopt application software components to for example, different vehicle platforms. Since AUTOSAR focus is set to mature application interfaces innovation in software functionality is not limited. New functions shall stay proprietary. This enables the OEM and software component developers to keep competition alive. In consequence the number of standardized application interfaces will grow over time.

3 Conclusion

Applying AUTOSAR to embedded automotive software development yields several important advantages. Due to the architecture that provides standardized infrastructure functions for application software, the application software itself is independent from the hardware. The VFB concept allows for developing the functional architecture of the application functions without any need of considering the underlying hardware with ECUs and network. This virtual integration means a frontloading of development work, so that integration issues could be identified and resolved early in the design process. The AUTOSAR approach provides very high flexibility in the software development. Roles and responsibilities can be easily split and therefore software can be handled as a product.

References

[AUTO] AUTOSAR: Official Website of the partnership, http://www.autosar.org.

[AI40] AUTOSAR R4.0 Specification: Table of Application Interfaces

[AM40] AUTOSAR R4.0 Specification: Methodology Model

[BF08] Bunzel, S.; Fennel, H.: The AUTOSAR Methodology. In (VDI): FISITA World Automotive Congress, Springer Automotive Media, September 2008; FISITA 2008/F2008-10-023.

[Fü09] Fürst, S; et al.: AUTOSAR – A Worldwide Standard is on the Road. 14th International VDI Congress Electronic Systems for Vehicles 2009, Baden-Baden

[KG08] Kinkelin, G.; Gilberg, A; et al: AUTOSAR on the Road. In (CTEA): SAE Convergence, October 2008; 2008-21-0019

Ein Qualitätsmodell zur automatisierten Ermittlung der Modellqualität bei eingebetteten Systemen

Jan Scheible, Ingo Kreuz
Daimler AG - Group Research and Advanced Engineering
{jan.scheible | ingo.kreuz}@daimler.com

Abstract: Die Qualitätssicherung der in der Automobilindustrie verwendeten Matlab Simulink-Modelle zur Softwareentwicklung von eingebetteten Systemen wird immer aufwendiger. Der Aufwand steigt, da immer mehr Funktionen im Fahrzeug als Software umgesetzt werden und der Umfang der Modelle stetig zunimmt. Diese Arbeit stellt einen Ansatz vor, der die Modellqualität mit Hilfe eines Qualitätsmodells und Metriken automatisiert ermittelt. Dadurch wird der Umgang mit sehr großen und komplexen Modellen möglich.

1 Motivation

Durch die modellbasierte Softwareentwicklung eingebetteter Systeme kann man im Vergleich zur Entwicklung mit textuellen Programmiersprachen unter anderem eine höhere Abstraktion und letztendlich eine höhere Softwarequalität erreichen [SVEH07]. Die höhere Abstraktion kann die Software-Entwicklung vor allem bezüglich nicht-funktionaler Qualitätsmerkmale unterstützen, weil Strukturiertheit, einheitliche Architektur, Wiederverwendbarkeit, Lesbarkeit und Übersichtlichkeit durch die grafische Notation unterstützt werden. Leider gibt die grafische Notation allein aber noch keine Garantie dafür, dass wirklich hohe Softwarequalität erreicht wird. So kann z. B. trotz grafischer Repräsentation eine ungünstige Architektur gewählt werden, Subsysteme falsch aufgeteilt werden oder Blöcke gewählt werden, die bei der Codegenerierung zu ineffizientem Code führen.

Modelle stellen bei der modellbasierten Entwicklung das zentrale Artefakt dar. Das bedeutet, dass die Modellqualität direkten Einfluss auf die Softwarequalität hat [FHR08]. Das heißt aber auch, dass durch die Qualitätssicherung von Modellen eine frühzeitige Absicherung erreicht werden kann: Fehler werden bereits im frühen Entwicklungsstadium erkannt und können noch kostengünstig behoben werden [FLS01]. Außerdem ist es im Automobilbau sehr wichtig, die Qualität eines Softwarestandes zu kennen, bevor er als Seriencode freigegeben werden kann, da es sich um eine sicherheitsrelevante Domäne handelt. Große Modelle in unserem Umfeld haben derzeit bis zu 15.000 Blöcke, 700 Subsysteme und 16 Hierarchieebenen.

Um die Qualität dieser Modelle mit vertretbarem Aufwand bewerten zu können, haben wir in [Sch10] ein Framework zur automatisierten Ermittlung der Modellqualität vorgestellt. In der aktuellen Arbeit gehen wir im Detail auf unser *Qualitätsmodell* ein, welches im Vergleich zu den klassischen Softwarequalitätsmodellen erweitert wurde. Dann beschreiben

wir die *Erhebung der Messwerte* durch die Metriken des Qualitätsmodells, welche nicht auf dem Quellcode, sondern auf den Modellen arbeiten. Dazu wird der schematische Ablauf der Erhebung und der Zugriff auf die Modelle beschrieben. Abschließend gehen wir auf die *Auswertung der Messergebnisse* ein.

2 Qualitätsmodell

Unser Ansatz für ein Qualitätsmodell folgt der Idee von Cavano und McCall, die bereits Ende der 70er Jahre das FCM-Qualitätsmodell (*Factors*, *Critera*, *Metrics*) für die Bewertung der Softwarequalität vorstellten [CM78]. Wir passen dieses Qualitätsmodell derzeit so an, dass es für grafische Modelle angewendet werden kann. Unser Fokus liegt dabei auf der Anwendbarkeit für Matlab Simulink-, Stateflow- und Targetlink-Modelle in beliebigen Entwicklungsphasen (im Folgenden *Simulink-Modelle* genannt), die in der Automobilindustrie besonders häufig zur Codegenerierung für eingebettete Systeme zum Einsatz kommen. Abbildung 1 zeigt einen Überblick über den momentanen Stand unseres Qualitätsmodells.

Bei Cavano und McCall setzt sich die Softwarequalität aus *Qualitätsfaktoren* (Factors) zusammen, die zum Teil allgemeingültig, aber zum Teil auch projektspezifisch sind. Wir haben diese Faktoren z. B. um den Qualitätsfaktor *Codegenerierbarkeit* erweitert, der für grafische Modelle hinzukommt. Außerdem wurden einige der ursprünglichen Faktoren zusammengefasst, um unser Qualitätsmodell übersichtlich zu halten. Dies war möglich, da manche Kriterien mehreren Faktoren zugeordnet werden konnten. Die Faktoren wiederum setzen sich aus *Qualitätskriterien* (Criteria) zusammen, welche die Faktoren konkretisieren. So wird z. B. der Qualitätsfaktor *Verständlichkeit* durch die Kriterien *Lesbarkeit*, *sprechende Bezeichner*, *kognitive Erfassbarkeit* u. Ä. konkretisiert. Kriterien können durch eine oder mehrere *Metriken* (Metrics) gemessen werden. Im Beispiel kann das Kriterium *Lesbarkeit* durch eine Metrik zur Messung der Schachtelungstiefe und der Anzahl der Rückkopplungen innerhalb eines Subsystems gemessen werden. Zur Qualitätsbewertung wird für jede Metrik neben einem Intervall mit erlaubten Werten eine Priorität definiert.

Zwar handelt es sich bei dem FCM-Modell um ein relativ altes und einfach strukturiertes Qualitätsmodell, aber unser Interesse gilt zunächst vor allem der Sammlung von Faktoren, Kriterien und Metriken, die spezifisch für die modellbasierte Entwicklung sind. So sind die meisten Metriken in unserem FCM-Modell nicht Simulink-spezifisch und daher auf andere Modellierungswerkzeuge übertragbar.

3 Erhebung der Messwerte

Dieser Abschnitt beschreibt die prototypische Anwendung unseres Qualitätmodells. Im Gegensatz zu Codemetriken messen die Metriken in unserem Ansatz direkt auf den Simulink-Modellen (vergleiche [Rau01]).

Abbildung 2 zeigt den schematischen Ablauf der Erhebung der Messwerte. Das Messobjekt

Abbildung 1: Momentaner Stand des Qualitätsmodells zur Modellqualitätsbewertung

Faktoren	Kriterien	Metriken	Beschreibungen
Codegenerierbarkeit	Verständlichkeit des Quellcodes	#Globaler Variablen	Wie viele globale Variablen gibt es im generierten C-Code?
		#Targetlink-Funktionen	Wie viele Subsysteme werden bei der Codegenerierung in C-Funktionen umgesetzt?
	Erhaltung der Architektur	#Architekturverletzungen	Ist die geplante Architektur im C-Code noch erkennbar?
	Explizite Datentypen	#potentieller Probleme mit Datentypen	Wie viele potentielle Probleme gibt es bei der Wahl der Datentypen und Festkommaskalierungen?
	Performance	Speicherverbrauch	Geht der generierte Code effizient mit dem Speicher um?
	Verwendung des Data Dictionaries	#Festkommaskalierungen mit/ohne DD-Eintrag	Wie viele Festkommaskalierungen sind (nicht) zentral im Targetlink-Data Dictionary hinterlegt?
		#Ports ohne DD-Eintrag	Wie viele Ports haben keinen zentral im Targetlink-Data Dictionary hinterlegten Datentyp?
Korrektheit	Einhaltung der Architektur	#Architekturverletzungen	Gibt es Unterschiede zur geplanten Architektur (z.B. einem SysML-Blockdiagramm)?
	Funktionale Korrektheit	#Anforderungen	Wie viele Anforderungen gibt es für das Modell (lässt Rückschlüsse auf die Problemgröße zu)?
		#Umgesetzter Anforderungen	Wie viele der Anforderungen sind im Modell umgesetzt?
Robustheit	Behandlung von Unter-/ Überläufen	#Sättigungsblöcke	Wie viele Sättigungsblöcke werden zur Behandlung von Über- und Unterläufen verwendet?
Testbarkeit	Ausreichende Tests	#Testfälle	Wie viele Testfälle gibt es für das Modell (ist interessant im Verhältnis zu #Anforderungen)?
		Testabdeckung	Wie gut wird das Modell getestet?
Verständlichkeit	Dokumentation	#Linienanmerkungen	Wie viele der Linien im Modell sind beschriftet?
		#Modellinformationsblöcke	In wie vielen Subsystemen wurden Modellinformationsblöcke zur Dokumentation verwendet?
		#Subsystemanmerkungen	In wie vielen Subsystemen wurden Anmerkungen (zur Erhöhung der Verständlichkeit) gemacht?
	Kognitive Erfassbarkeit	ØVernetzungsgrad der In- und Outports	Auf wie viele In-/Outports wirkt sich durchschnittlich ein In-/Outport aus?
		#Verwendeter Blocktypen	Wie viele der möglichen zur Verfügung stehenden Blocktypen wurden in dem Modell verwendet?
		#Rückkopplungen	Wie viele Rückkopplungen sind in dem Modell enthalten?
		Tiefe der Subsystemhierarchie	Wie tief ist die Subsystemhierarchie (ist interessant im Verhältnis zu #Anforderungen)?
	Lesbarkeit	ØBreite/Höhe der Subsystemdarstellung	Wie breit/hoch (in Pixel) sind die Subsysteme auf dem Monitor im Durchschnitt?
		#Überkreuzender Linien	Wie viele überkreuzende Linien enthält das Modell?
	Umfang	ØBreite der In-/Output-Schnittstelle	Wie viele In-/Outports haben die Subsysteme des Modells im Durchschnitt?
		ØSignale pro Bus	Wie viele Signale enthält ein Bus in dem Modell im Durchschnitt?
		#Blöcke	Wie viele Blöcke enthält das Modell (ist interessant im Verhältnis zu #Anforderungen)?
		#Buserstellungen	Wie oft wird ein neuer Bus erstellt?
		#interne Zustände	Wie viele interne Zustände hat das Modell?
		#Linien	Wie viele Linien gibt es im Modell?
		#Subsysteme	Wie viele Subsysteme enthält das Modell (ist interessant im Verhältnis zu #Anforderungen)?
		Zyklomatische Komplexität	Wie groß ist die zyklomatische Komplexität der einzelnen Subsysteme und des Gesamtsystems?
	Verwendung sprechender Bezeichner	ØNamenswechsel pro Signal	Wie oft werden Signale umbenannt?
		#magischer Konstanten	Wie viele magische Konstanten (d.h. Konstanten ohne Namen) werden verwendet?
		#Konstanten mit Name	Wie viele Konstanten mit Name werden verwendet?
Wartbarkeit	Standardkonformität	#Verwendeter Modellierungsmuster	Wie oft wurden etablierte Modellierungsmuster verwendet?
		#Verletzter Modellierungsrichtlinien	Wie oft werden die gewählten Modellierungsrichtlinien verletzt?
	Vermeidung impliziter Konstrukte	#From- und Goto-Blöcke	Wie viele From- und Goto-Blöcke werden verwendet, die den normalen Signalfluss unterbrechen?
		#Verwendetem eingebettetem C-Code	Wie oft wurde bestehender C-Code in das Modell integriert?
		#Verwendung von Speicherblöcken	Werden viele Speicherblöcke verwendet, die nicht Teil des normalen Signalflusses sind?
	Vermeidung von Redundanz	#(fast) identischer Subsysteme/Blockgruppen	Wurden große bzw. viele Modellteile per Copy&Paste erstellt?
		#Gebrochener Bibliotheksverknüpfungen	Wie oft wurden Bibliotheksblöcke verwendet, die nachträglich verändert wurden?
		#unverwendeter Signale auf tiefster Ebene	Wie viele Signale werden nur durch Subsysteme durchgeschleust und nicht verwendet?
		#Verwendeter Bibliotheken	Wie viele Bibliotheken wurden insgesamt verwendet?
		#Verwendeter Bibliotheksblöcke	Wie oft wurden Blöcke aus Bibliotheken verwendet?

enthält sowohl das Simulink-Modell als auch externe Datenquellen (sog. *Data-Provider*). Die Data-Provider stellen direkt Messwerte oder aufbereitete Informationen über Modelle zur Verfügung. Zwei Data-Provider sind z. B. die Simulink Verification and Validation Toolbox, welche die zyklomatische Komplexität von Modellen misst, und der Modellierungsrichtlinenprüfer MXAM von MES[1], welcher die Konformität von Modellen prüft.

Die Metriken erheben ihre Messwerte sowohl durch Berechnungen direkt auf den Modellen als auch durch Abfrage von Informationen von den Data-Providern. Die Struktur der Simulink-Modelle wird mit einem Metamodell beschrieben.

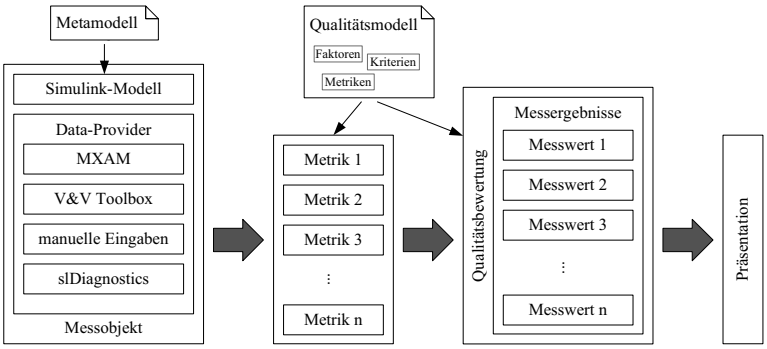


Abbildung 2: Schematischer Ablauf der Erhebung

Dieses Metamodell stellt in unserer Implementierung eine vereinfachte Sicht auf die Simulink-Modelle bereit und ermöglicht die Erhebung der Messwerte in einem Java-Prototypen. Als Grundlage dient der Simulink-Parser der TU-München aus dem ConQAT-Projekt[2]. Da der TUM-Parser weder das Nachladen von referenzierten Blöcken aus Bibliotheken unterstützt noch komfortable Zugriffsmethoden besitzt, wurde sein Metamodell in unserem eigenen Metamodell gekapselt. Dadurch sind die gewünschten Funktionen verfügbar und es ist z. B. möglich, den Wert eines Konstanten-Blocks auszulesen, unabhängig davon, ob er ein Simulink- oder Targetlink-Konstantenblock ist.

Mit Hilfe der Messergebnisse wird schließlich die Modellqualitätsbewertung durchgeführt. Dazu wird für jeden Messwert geprüft, ob er innerhalb der erlaubten Grenzen liegt. Dann wird für jeden Faktor berechnet, wie viele seiner Metriken Messwerte innerhalb ihrer erlaubten Grenzen haben. Das Ergebnis der Modellqualitätsbewertung wird anschließend in der grafischen Oberfläche des Prototypen präsentiert.

4 Auswertung der Messergebnisse

Abbildung 3 zeigt erste beispielhafte Messergebnisse, die mit dem Prototyp ermittelt wurden. Dabei ist zu beachten, dass aus Platzgründen nicht alle bereits umgesetzten Metriken aus

[1] http://www.model-engineers.com/en/our-products/model-examiner.html
[2] http://conqat.cs.tum.edu/index.php/SimulinkLibrary

	#potentieller Probleme mit Datentypen	#Targetlink-Funktionen	#Skalierungen mit DD-Eintrag	#Skalierungen ohne DD-Eintrag	#Ports ohne DD-Eintrag	#Anforderungen	#Testfälle	#Linienanmerkungen	#Subsystemanmerkungen	ØBreite/Höhe der Subsystemdarstellung	#Überkreuzender Linien	#Blöcke	#Buserstellungen	#interne Zustände	#Linien	#Subsysteme	#magischer Konstanten	#From- und Goto-Blöcke	#Verwendung von Speicherblöcken	#Gebrochener Bibliotheksverknüpfungen	#Verwendeter Bibliotheksblöcke
Modell 1	711	6	37	0	63	339	317	5.131	96	161	904	18.779	81	389	19.429	1.924	24	0	70	9.170	10.578
Modell 2	2.984	13	57	96	993	1.651	527	10.280	994	149	2.545	38.427	438	478	41.677	3.732	4	288	175	17.707	20.149
Modell 3	1.236	2	15	20	36	503	192	3.688	145	137	1.116	10.551	250	428	11.757	978	39	150	21	4.305	4.799

Abbildung 3: Messwerte von Simulink-Modellen aus einer sehr frühen Entwicklungsphase

Abbildung 1 enthalten sind und noch keine Modellqualitätsbewertung durchgeführt wurde.

Für eine robuste Modellqualitätsbewertung werden mehr Messwerte benötigt. Erst dann können die erlaubten Grenzen für die einzelnen Metriken festgelegt werden. Deshalb müssen mehr Messwerte gesammelt werden, um die Modellqualitätsbewertung zu validieren. Dabei können z. B. Reviewergebnisse, Experteneinschätzungen oder Fehleranzahlen aus Tests als Referenz verwendet werden. Wenn die Modellqualitätsbewertung nicht zutreffend ist, müssen die Grenzen der Metriken und/oder das Qualitätsmodell angepasst und dann erneut auf die Messwerte angewendet werden. Das Vorgehen ist iterativ so oft zu wiederholen, bis die Modellqualitätsbewertung mit den Referenzwerten übereinstimmt.

5 Verwandte Arbeiten

Ein weiteres generisches Qualitätsmodell stellt der Goal-Question-Metric-Ansatz (GQM) dar [BCR94]. Durch den Ansatz wird wie bei der Verwendung eines FCM-Modells die Qualität von Software bzw. Modellen objektiv erfassbar.

Zusätzlich zu dem Qualitätsmodell von Cavano und McCall gibt es viele weitere ähnliche Softwarequalitätsmodelle. Zwei Beispiele sind das Softwarequalitätsmodell von Boehm in [BBL76] und die darauf basierende ISO 9126. Allen gemeinsam sind Faktoren wie Zuverlässigkeit, Effizienz, Wartbarkeit und Portierbarkeit. Unterschiede gibt es jedoch bei den weiteren spezifischen Faktoren und dem Zusammenspiel der Kriterien.

Deißenböck et al. [DWP+07] stellen ein Qualitätsmodell für die Wartbarkeit auf. Es enthält nicht nur die Qualitätskriterien, sondern auch deren Einfluss auf die nötigen Wartungsaktivitäten. In einer Fallstudie erweitern sie ihr Qualitätsmodell um Simulink-spezifische Kriterien und Aktivitäten. Die automatisierte Qualitätsbewertung steht nicht in ihrem Fokus.

Einen Ansatz zur Qualitätsbewertung von UML-Modellen beschreiben Lange und Chaudron in [LC05]. Sie unterscheiden zwischen Entwicklung und Wartung, da in den jeweiligen Phasen andere Qualitätskriterien relevant sind. Sie verwenden zum großen Teil OO-Metriken und nur einige modellspezifische Metriken, wie z. B. die Anzahl der überkreuzenden Linien.

6 Zusammenfassung und Ausblick

Wir haben in dieser Arbeit ein Qualitätsmodell zur automatisierten Qualitätsbewertung von Simulink-Modellen vorgestellt. Das Qualitätsmodell enthält auf unterster Ebene Metriken, welche Messwerte auf den Simulink-Modellen messen. Es wurde der schematische Ablauf der Messung vorgestellt und kurz auf ein Metamodell für die Simulink-Modelle eingegangen. Schließlich wurden erste Messergebnisse vorgestellt.

Wie in Abschnitt 4 beschrieben, ist der nächste Schritt die Kalibrierung der einzelnen Metriken, d. h. das Ermitteln der erlaubten Grenzen. Außerdem muss der Zusammenhang zwischen der Problemgröße (z. B. Anzahl der Anforderungen) und den erlaubten Grenzen der Metriken weiter untersucht werden. Des Weiteren zeigen erste empirische Ergebnisse, dass die Ausdrucksmächtigkeit des Qualitätsmodells noch erhöht werden muss, um z. B. mit widersprüchlichen Messwerten besser umgehen zu können. Ebenso ist zu prüfen, ob verschiedene Entwicklungsphasen in dem Qualitätsmodell berücksichtigt werden müssen.

Literatur

[BBL76] B. W. Boehm, J. R. Brown und M. Lipow. Quantitative evaluation of software quality. In *Proceedings of the 2nd International Conference on Software Engineering*, Seiten 592–605, Los Alamitos, CA, USA, 1976. IEEE Computer Society Press.

[BCR94] V. Basili, G. Caldiera und H.D. Rombach. *Encyclopedia of Software Engineering*, Kapitel Measurement, Seiten 646–661. John Wiley & Sons, Inc., 1994.

[CM78] J. P. Cavano und J. A. McCall. A framework for the measurement of software quality. In *Proceedings of the Software Quality Assurance Workshop on Functional and Performance Issues*, Seiten 133–139, 1978.

[DWP+07] F. Deißenböck, S. Wagner, M. Pizka, S. Teuchert und J.-F. Girard. An Activity-Based Quality Model for Maintainability. In *Proc. of the 23rd IEEE International Conference on Software Maintenance*. IEEE CS Press, 2007.

[FHR08] F. Fieber, M. Huhn und B. Rumpe. Modellqualität als Indikator für Softwarequalität: eine Taxonomie. *Informatik-Spektrum*, 31(5):408–424, October 2008.

[FLS01] K. Frühauf, J. Ludewig und H. Sandmayr. *Software-Projektmanagement und –Qualitätssicherung*, Kapitel Software-Nutzen und -Kosten. Teubner Stuttgart, 2001.

[LC05] C. F. J. Lange und M. R. V. Chaudron. Managing Model Quality in UML-Based Software Development. In *Proc. of the 13th IEEE International Workshop on Software Technology and Engineering Practice*, Seiten 7–16. IEEE Computer Society, 2005.

[Rau01] A. Rau. Einsatz von Metriken bei der modellbasierten Software-Entwicklung. *Simulationstechnik - 15. Symposium in Paderborn*, 2001.

[Sch10] J. Scheible. Ein Framework zur automatisierten Ermittlung der Modellqualität bei eingebetteten Systemen. In *Proc. of the Dagstuhl-Workshop: Model-Based Development of Embedded Systems (MBEES), 2010, Schloss Dagstuhl, Germany*, 2010.

[SVEH07] T. Stahl, M. Völter, S. Efftinge und A. Haase. *Modellgetriebene Softwareentwicklung: Techniken, Engineering, Management*. dpunkt.verlag, 2007.

ISO 26262 conformant Verification Plan

Ralf Nörenberg, Ralf Reissing, Jörg Weber*

Specification and Test (GR/PST), Functional Safety (GR/PSP)*
Daimler AG, Group Research and Advanced Engineering
Hanns-Klemm-Str. 45, 71034 Boeblingen
ralf.noerenberg@daimler.com

Abstract: This contribution highlights the challenges of implementing ISO 26262 to an industrial E/E verification and testing environment. A methodology to obtain a verification plan and an adequate test strategy in order to meet ISO 26262 requirements is presented and evaluated in an in-house project.

1 Introduction to ISO 26262

Almost all functionality of a vehicle is realized by embedded systems consisting of networked hardware components with implemented software, and today, a total fraction of 70% to 90% of future automotive innovations is predicted to be based on such systems [Do07]. This increase of software-based functionality however results in higher numbers of possible software faults which may cause system failure [We10]. A 50% ratio of vehicle breakdowns caused by such failures was already estimated by [Be03].

The ISO 26262 [ISO], an automotive domain specific derivation of the generic IEC 61508 standard for functional safety of E/E-systems, addresses this subject [IEC]. Its goal is to further improve the high safety standards of the automotive industry by recommending appropriate methods and tools to contain product safety inherent in specification, design and verification. Central to the specification and design part of the standard are the concepts of risk identification, risk classification and derivation of safety requirements with a corresponding ASIL (Automotive Safety Integrity Level) to avoid or mitigate a particular risk.

2 Challenges to ISO 26262 conformant Verification and Testing

Several OEMs have started pilot projects to integrate ISO 26262 requirements into their industrial environments [Sc09], however, with a strong focus on specification and design processes. This contribution, in contrary, highlights the challenges of integrating ISO 26262 verification by discussing the development of a methodology to 1) to create a project specific “verification plan” according to ISO 26262, part 8, requirement 9.5.1 (short: 8-9.5.1) and 2) derive a sufficient test strategy (8-9.4.2).

The general concept of the ISO 26262 verification approach is the definition of several test objectives to be met on each test level. Further, to accomplish the required objectives, the standard recommends methods for deriving test cases, test methods and coverage criteria to be applied. Foremost, these recommendations are to be understood as a guideline to choose appropriate methods for verification, intentionally demanding a tailoring towards specific work environments. Each OEM thus may elaborate a specific approach, which is unambiguous in interpretation and application and provides a good feasibility within its environment. The three degrees of freedom available by ISO 26262 are: interpretation, selection and combination.

The freedom of interpretation is applicable as ISO 26262 does basically refer to test methods without defining them in detail, e.g. stress testing (4-8.4.2.3.5). Thus, the OEM is free to define the method's exact meaning to its specific environment, or even provide argumentation why an available substitute method applied is sufficient as well (4-4.2).

The freedom of selection is valid as ISO 26262 methods are goal oriented. A right set of "feasible" methods thus is adequate to reach a test objective (4-8.4.1.5; 4-8.4.2.3(.1)). This enables the user to limit methods recommended for a safety requirement to those which are applicable in terms of practicability and usefulness. For example, the method "long term test" may not be reasonable to all safety requirements.

The freedom of combination is given as the standard only recommends an appropriate combination of methods to derive test cases in consideration of the level of integration (recommended test methods), which includes both, selection and combination. Certain combinations of recommended test methods (e.g. 4-8.4.2.3.1) and methods to derive test cases (e.g. 4-8.4.1.5) therefore may be chosen, as some combinations might not even be suitable ("analysis of Boundary values" in dependency with "performance test").

These freedoms enable a variety of verification approaches to be developed. However, the OEM is still in charge to provide (sufficient) argumentation and documentation of its specific approach of functional safety according to ISO 26262: The rationale of the developed test strategy shall be reasonable, the traceability of chosen verification methods to corresponding ISO 26262 requirements and the proof of completeness shall be given. This information is to be documented in the ISO 26262 required verification plan. As ISO 26262 gives only a rough idea on how the document has to be structured, the OEM is strongly advised to define a process to achieve a harmonized and universal verification plan (8-9.5.1). The approach taken at Daimler is shown in Fig 1.

3 Integrated Verification Plan

The integrated verification plan developed at Daimler considers two aspects: First, the work product "verification plan" is achieved by a document template prescribing a standardized, clear and complete framework for documentation. Second, a methodology efficiently guides the developer through filling in the template and to derive an integrated test strategy for a specific project correctly.

The following discussion is focused on ISO 26262 subjects only. However, the great benefit of the presented methodology / verification plan is to be considered the first-time merge of ISO 26262 verification requirements with existing in-house standards and methodologies for systematic derivation of test specifications. Therewith, the methodology also allows the derivation of a tailored test strategy for non ASIL rated systems. Thus, the verification plan is obligatory for all future verification and test activities.

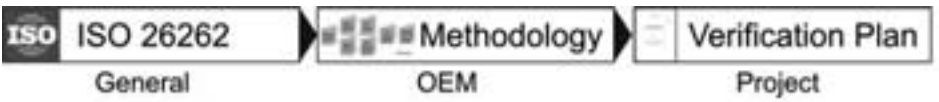


Figure 1: Approach to implement ISO 26262 to a specific project

The supplement of the word integrated to the verification plan describes the generality of the document as all test levels are considered in one document. The integrated view on all test levels provides the most efficient and effective test implementation and execution as it grants identification of test gaps and unnecessary test redundancies over the whole verification process. Moreover, the unified approach allows test objectives to be moved or split up to other possible test levels while ISO 26262 conformance is still met. In case verification activities are distributed over various parties (e.g. OEM and its suppliers), the integrated verification plan also describes the responsibilities of each party and the interfaces between these parties.

3.1 Integrated Verification Plan Template

The integrated verification plan document template is based on the ISTQB interpretation of IEEE 829-1998 test plan [Sp05] [IEEE] and prescribes the structure and relevant content of the template. The 90-page template allows to define test objects, document objects and elements verified elsewhere, define test levels, assign test platforms, to acquire test objectives, select methods to derive test cases and finally, to combine everything into an ISO 26262 conformant test strategy.

The high generality of the document features the ability to be adjusted to specific projects by offering predefined tailoring measures and argumentation choices to be made (for example, consideration of available test platforms) by systematically guiding the developer through ISO 26262 requirements. Of course, in order to assure the sufficiency of ISO 26262 consideration (for argumentation), a separately documented traceability matrix traces the requirements to the proper location in the template. The template is currently provided to in-house pilot projects being subject to ISO 26262 analyses.

3.2 Methodology

The methodology – currently available as a handbook – has been developed to assist filling in the integrated test concept template. In order to guide the user through the most efficient way to fill in the template, a workflow has been defined, whose upper-most level is depicted in Fig 2. As each workflow step is to be completed by determined employees, each step is assigned to certain roles (test manager, test designer, tester).

This masks out all steps not being relevant to the role and thus increases work efficiency. Independent of the role, each available workflow step is split into six subsections: 1) objectives, a description of the main goals 2) required inputs, a specification the required information and documents; 3) procedure, a specification of the individual tasks to performed 4) tools, a list of tools that may assist the tasks 5) work products, a description of intended results and pointers to the locations where to document the results in the template and 6) checklist, to check items to validate the results of this step.

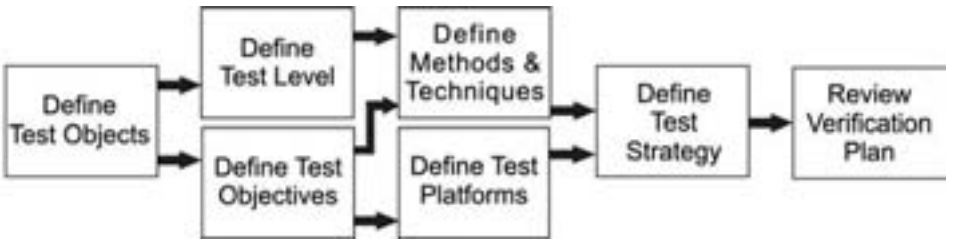


Figure 2: Workflow of the methodology handbook to maintain an integrated verification plan

For a more advanced guidance throughout the topics of the subsections a 3rd level of information has been attached to the individual tasks. This layer consists of 1) background information for general knowledge about the topic, basic ideas and definitions, 2) tutorial on the how to perform the task in detail and how to fill in the template, and 3) an example for a completed section of the verification plan template.

3.3 Towards a Daimler tailored Verification Methodology

The main goal of a Daimler tailored verification methodology is firstly, to provide a highly standardized implementation of the ISO 26262 requirements and simultaneously being unambiguous in interpretation and application, thus being feasible to be put into practice. Secondly, to manage the complex challenge of selecting a right set of methods for verification by using the three degrees of freedom outlined in section 2.

Therefore a twofold approach is chosen: The total quantity of recommended ISO 26262 methods for verification has been evaluated and interpreted to have a standardized and reasonable understanding of the terms and their common sense. Potential substitute methods were also identified. Within the evaluation, the number of methods could be reduced to a smaller number by introducing “test types” – a classification of methods which are similar for all test levels. Defined test types are: function test, interface test, robustness test, performance test, back-to-back test, experience-based test, long term test, and diagnosis test. For example, ISO 26262 defines many interface test methods, which here are summarized by one test type. This approach reduces the number of methods and thus complexity without endangering ISO 26262 conformance.

In order to maintain the link between a test type and ISO 26262 methods it originated from a separate document contains all traceability and background information of the transformation. Of course, knowledge of the original test objective and the method the test type represents at a particular test level is retrievable. In order to document coverage of test methods by a set of test cases (a test specification), test types (multiple choice) are assigned to every test case. Thus, test types offer not only an easy understandable application of the ISO 26262 methods but also an efficient way to meet documentation and organizational requirements.

However, the integration of the freedoms of selection and combination has not been completely resolved yet and is subject to our current focus of research. The approach is to assign safety requirements to a selection of test objectives – an ISO 26262 tailoring similar to the test types is to be considered – which then rule the application of certain verification methods (test types). For example, a safety requirement with a given ASIL is set to meet test objectives "A: validation of robustness" and "B: examination of interfaces" by the safety manager. This now requires the test designer to create test cases with the test types "robustness test" and "function test" (for A) and "interface test (for B). The freedom of selection and combination would thus be reduced by the choice of valid test objectives. The test designer would still have to decide whether all test types ruled by the test objective are necessary or not.

Additionally, this methodology enables an automated evaluation to rate running test activities against defined criteria. Thus if achieved, allowing the conclusion of having obtained – in terms of an entry criterion - an adequate ISO 26262 conformance in the test specification. A manual review therefore may be postponed to the time the entry criterion is achieved, greatly improving personnel capacity and work efforts.

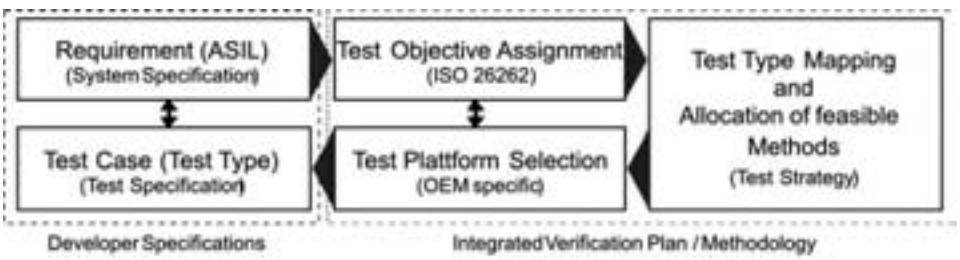


Figure 3: Workflow on how to verify and test ASIL rated requirements with the methodology.

Figure 3 shows the workflow on how an ASIL rated requirement is to be verified and tested. The requirement is assigned to an ISO 26262 test objective which is to be achieved by certain pre-set methods (e.g. test types). The identification of feasible methods is subject to the tailoring guideline of the methodology documented in the verification plan. The assignment to test platforms (on available test levels) is defined by a matrix, however, does allow shifting tests to different test levels in regard to the ISO 26262 test objective to be achieved. The final test cases derived are obtained by the developer/tester and are marked by with the additional attribution "test type" only. The bi-directional traceability to the ISO 26262 methods applied and test objectives covered is provided by the integrated verification plan.

4 Qualitative evaluation of the integrated Verification Plan

Analyses of completed integrated verification plan during a pilot project showed that all strategic and technical information for verification is contained and ISO 26262 compliance is met by using described tailoring mechanisms. The template still needs some improvement on usability at few areas. In order to have the testers complete their own verification plan more efficiently further guidance and automation of the tasks are required. This is subject to current work.

In comparison with ISO WD 26262 (baseline 12), which required an application of all recommended methods to an ASIL rated requirement (and therefore was difficult to be applied in a real life environment due to test and argumentation efforts), the evaluation of the test strategy derived from ISO DIS 26262 shows good conformance for most of the reviewed parts. An attained result of the pilot project therefore is that 1) a general approach on deriving a integrated verification plan containing a 2) project-specific tailoring of the ISO 26262 recommendations becomes necessary in order to achieve ISO 26262 implementation and 3) the ISO DIS 26262 therefore made good progress towards being applicable in practice.

5 Conclusion

The evaluation of the pilot project shows that the challenge of integrating processes and methods required by the ISO DIS 26262 into an existent OEM verification and test environment can be met. A few remaining challenges exist, but are expected to be resolved soon. The introduction of a work-flow oriented methodology to derive an integrated verification plan as well as an ISO 26262 conformant test strategy are on a good development level to face future ISO 26262 requirements. Current research is focused on perfection of the presented test methodology and on an intranet portal solution to present the methodology and its tools in a way that increases usability.

References

[Be03] Behlmer, A.: Discordant notes in the ensemble. Automobil Industrie 48, 16-18 (2003)

[Do07] Dold A., Trapp M.: Herausforderungen und Erfahrungen eines OEM bei der Gestaltung sicherheitsgerechter Prozesse, Lecture Notes in Informatics, (GI), 2007

[IEC] International Electrotechnical Commission: IEC 61508: Generic Standard Functional Safety “Electrical/electronic/programmable electronic safety-related systems”, 2001

[IEEE] Institute of Electrical and Electronic Engineers, Standard for Software Test Documentation (IEEE 829), 1998

[ISO] International Organisation for Standardization: ISO DIS 26262 baseline 15, 2010

[Sc09] Schwarz, J. and Buechl, J.: Preparing the future for functional safety of automotive E/E systems, 21st Int. Tech. Conference on the Enhanced Safety of Vehicles, 2009

[Sp05] Spillner A.: Basiswissen Softwaretest, Page 173, dpunkt.verlag GmbH, 2005

[We10] Weber, J.: An Aspect Driven Approach for the Analysis, Evaluation and Optimization of Safety within the Automotive Industry. SAE 2010 World Congress & Exhibition, 2010

Workshop

eHumanities – Welchen Nutzen hat die Informatik

Prof. Gerhard Heyer, Marco Büchler

GI Annual Meeting 2010

Workshop

eHumanities - How does computer science benefit?

Gerhard Heyer, Marco Büchler
Natural Language Processing Group
Institute for Mathematics and Computer Science
University of Leipzig, Germany
[heyer|mbuechler]@informatik.uni-leipzig.de

In recent years the text-based humanities and social sciences experienced a synthesis between the increasing availability of digitized texts and algorithms from the fields of information retrieval and text mining that resulted in novel tools for text processing and analysis, and enabled entirely new questions and innovative methodologies.

The goal of this workshop is to investigate which consequences and potentials for **computer science** have emerged in turn from the digitization of the social sciences and humanities.

The workshop starts with a series of four invited talks by leading researchers in the field of eHumanities. Their presentations will revolve around the question "How can computer science benefit from eHumanities?". The afternoon will focus on demonstrations and discussions of different solutions to an open challenge, which aims to contrast and compare methods used in computer science with those in the humanities.. In this section, members from both fields of the eHumanities community will apply their own methods and tools on data of their choice to solve a set of previously announced problems. The exact challenges will be made public with the official announcement of the workshop and will be focused on current issues of unsupervised semantic analysis of text which are relevant to computer science, e. g. the handling of unexpected relations and associations, the treatment of rare textual patterns, or the merging of heterogeneous sources.

The date for the workshop has been fixed on Thursday, September 30th, 2010. Prof. Dr. Stefan Wrobel (Director IAIS, Bonn/St. Augustin), Dr. Helge Kahler (Federal Ministry of Education and Research - Department of Humanities), Peter Wittenburg (MPG Nijmegen - Project CLARIN) and Prof. Dr. Gregory Crane (Tufts University, Boston - Project PERSEUS) will be the speakers for the morning session.

Some Challenges Posed to Computer Science by the eHumanities

Gerhard Heyer, Marco Büchler

Natural Language Processing Group
Institute for Mathematics and Computer Science
University of Leipzig, Germany
[heyer|mbuechler]@informatik.uni-leipzig.de

1 eHumanities

To the extent that applications of computer science have always lead to a replacement of analogue by *digital* media and processes, digital media and processing models have an increasing impact also on traditional work flows based on analogue media in the humanities and social sciences. The interdisciplinary combination of methods from computer science and traditional humanities with large amounts of digital data and advanced tools for processing these is commonly known as *eHumanities* [Ca05]. In a broad sense, eHumanities are concerned with any digitized data that are subject to investigation in the humanities and the social sciences, e.g. text, images, and objects (such as in Archeology). However, focusing on text as the main data type in the text oriented humanities helps to highlight the benefit that can be gained from the combination of digital document collections and new analysis tools derived from the area of information retrieval and text mining. Thereby all kinds of historically oriented text sciences as well as all sciences that work with historical or present day texts and documents are enabled to ask completely new questions and deal with text in a new manner. In detail, these methods concern, amongst others,

- The *qualitative improvement* of the digital sources (standardization of spelling and spelling correction, unambiguous identification of authors and sources, marking of quotes and references, temporal classification of texts, etc.);
- The *quantity and structure* of sources that can be processed (processing of very large amounts of text, structuring by time, place, authors, contents and topics, comments from colleagues and other editions, etc.);
- The kind and quality of the *analysis* (broad data driven studies, strict bottom-up approach by using text mining tools, integration of community networking approaches).

To use such computational methods, an individual researcher can proceed by employing two strategies depending on his, or her, own degree of computer literacy. On the one hand, there is the individual software approach. Given a selection of digital text data, the research question is being transferred into a set of issues and methods that can be dealt

with by a number of individual programs. This approach allows for a highly dynamic and individual development of research issues. It requires, however, a high degree of software engineering know-how. On the other hand, there is the standard software approach. For well-defined and frequently encountered tasks, a digital humanities infrastructure will offer solutions that provide the users with data and analysis tools that are well understood, have already delivered convincing results, and can be learnt without too much effort.
Both approaches are interdependent. Probably good solutions in one domain of text oriented humanities can be transferred to other domains by just using different kinds of text. A good infrastructure must be capable of making such solutions accessible as best practices.

2 Interactions with Computer Science

While eHumanities are becoming increasingly popular in the humanities, the focus of this workshop is to investigate which consequences and potentials for **computer science** have emerged in turn from the digitization of the social sciences and humanities. Computer Science and Humanities so far have acted in their working methodologies more as antipodes rather than focusing on the potential synergies. For computer science turning towards the humanities as an area of application may pose new problems that may also lead to rethinking present approaches hitherto favored by computer science and developing new solutions that help to advance computer science also in other areas of media oriented applications.

By way of example, let us consider the so-called digital classics in detail. Historically oriented classical studies (Ancient History, Classical Philology – Latin, Greek, and Byzantine Studies – Epigraphics, Papyrology) nowadays all use digital text corpora that are available in various media (digital WEB libraries, CD-ROM, DVD) and formats (Beta-Code in all varieties, UTF-8, ASCII). Thanks to substantial digitization efforts, most of the ancient greek and latin text corpora are available nowadays over the internet and represent. The public resources of ancient texts include, amongst others, Thesaurus Linguae Graecae (TLG), Perseus, Packard Humanities Institute (PHI), and Bibliotheca Teubneriana Latina (BTL). While the use of digital texts has up to now been mainly restricted to search for, e.g., specific wordings, text passages, or proper names, advancements in text mining make it possible to exploit digital text resources also as raw material for acquiring structured knowledge out of the huge amount of unstructured textual data. A good reference to illustrate the point is *eAQUA*, an interdisciplinary project set up between the departments for classical studies at the University of Leipzig, Heidelberg, and Hamburg, and the division for natural language processing at the computer science department of the University of Leipzig, explicitly addressing the issue of text mining in digital text corpora for the classical studies [HS10]. eAQUA is funded by the German Ministry for Research, Education, and Technology, and aims to explore interdependencies between computer science and the humanities and to lay the foundations for an e-science infrastructure in the humanities. The goals of the project are threefold,

i. to establish a research infrastructure for researchers in the classical studies by setting up a portal that makes specific solutions to research issues accessible to the scientific community by way of best practice applications that can easily be adapted to similar research issues;
ii. to make different digital resources of ancient text accessible to researchers from the eAQUA platform; and
iii. to apply in an experimental way advanced text mining to the ancient texts in order to gain experience which kind of issues in the classical studies lend themselves to this approach, also advancing where possible the classical studies in their detailed research issues.

By using text mining, factual and content related interdependencies can be reconstructed that otherwise could not have been derived as rapidly and exhaustively. Their graphic depiction indicates streams of meaning that visualize historical traditions as well as historical facts in particular contexts like from a bird's eye view. In the traditional way of exploring historical texts such insights on a "higher level" are very time consuming and require an intimate knowledge of and a long time experience in handling ancient text sources.

3 Impacts on Computer Science

From a computer science point of view, we see four main challenges for Computer Science as a result of such interactions, (i) software engineering issues, (ii) impacts on semantic technologies, (iii) visual analytics, and (iv) infrastructure issues.

3.1 Software Engineering Issues

As a general approach to the software engineering issue of how to transform a research issue of the classical studies into a software problem, eAQUA, for example, follows a methodology that distinguishes between Data, Algorithms, and Applications, and that has successfully been applied in the area of natural language processing. Starting with an application as the actual scientific research question, we need data on the one hand, and algorithms on the other. In eAQUA, text data are being imported by eAQUA's standardized interface to available resources. Algorithms are taken from the field of natural language processing and text mining. The main challenge then is to select the suitable algorithms and specify their combination and interaction in order to contribute to a solution of the problems set out in the application scenario. Figure 1 represents a simple instance of this methodology for the reconstruction or the correction of a papyrus or an inscription.

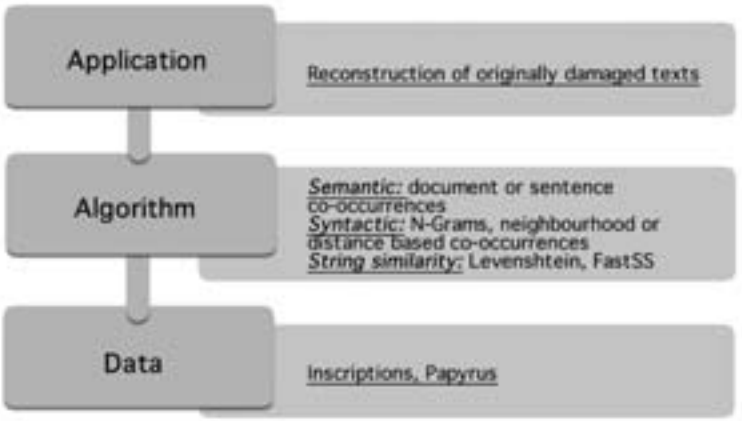


Figure 1: Distinguishing between data, algorithms, and applications for solving the problem of correcting damaged papyri

How can this approach be generalized? And how can it be implemented by using a flexible architecture that allows as much as possible reuse of individual software components, and rapid prototyping of new applications?

3.2 Impacts on Semantic Technologies

Research in Classical Studies is often data driven, i.e. researchers are looking for phrases or words, collect the results and interpret them in view of their field's background knowledge. In contrast, the preferred method of computer scientists is model driven. For dedicated texts, or fragments of text, they try to define significant features that can be used as input for algorithmic analyses of the text, such as classification, or clustering, and latent semantic analysis.

A good example of a possible trade-off between both methods again is eAQUA. On the one hand, researchers in the classical studies need access to digital libraries for searching and comparing texts, on the other hand, the power of model driven analytics can also reshape a research question from a more formal point of view. While a confirmation of what is already known from studying ancient texts always is helpful in scientific work (in particular when it is speeding up research), it is not by itself really innovative and would not justify considering it a novel method of search. The situation is different when by using, e.g., a contextual search we get hints on semantic relations that are neither obvious, nor well known, and that cannot easily be derived by using traditional or otherwise established tools and search strategies like dictionaries, concordances, indices, or dedicated search engines like Diogenes and TLG-online.

Scientific practice in the humanities clearly indicates that we may need to distinguish between knowledge concerning what is obvious and commonly accepted on the one hand, and knowledge concerning particular, or rare events, knowledge that summarizes

the exception from the rules and is distinctive of experts, but that many wish to share. For a number of reasons, including considerations of relevance and optimization, in information retrieval as well as in data and text mining the main methods today are based on statistical methods. By using such methods we can well detect and analyze statistically significant patterns, but these methods are not really suitable for dealing with rare events and their significance within a given context of research. How can this knowledge of rare events be represented and adequately be dealt with? How can we modify and apply semantic technologies in information retrieval to also deal with rare events?

3.3 Visual Analytics

A common assumption in computer science up till now is that first data need to be analyzed, and that visualization then is needed only to visualize the results. However, the necessity to deal with an increasing amount of data may require us to rethink this premise. Visual analytics aims to interactively filter the information relevant to an application and to communicate it to humans in an appropriate way. Again, a good example is eAQUA and its tools for the analysis of textual reuse [BGH10]. Similar to modern publications, classical authors also used the texts of others as sources for their own work. Hence, the analysis of textual reuse plays an important role in Classical Studies research. Leaving aside the technical point of view of semantic technologies for this application, the research of Classicists includes both an application of a *macro view* for Historians as well as one for the *micro view* of Classical Philologists. The visualization dimension of textual reuse is important since text mining approaches typically generate a huge amount of data that can't be explored manually. In effect, visualization helps to better understand the data, allows to gain new insights, and to interact effectively with the data analysis methods. The interaction permits the user to bring his expert knowledge into the data analysis process. From a methodological point of view, however, the semantic technologies and text mining tools also need to get adapted in order to meet the challenge that important characteristics of the data such as dimensionality, homogeneity, topicality, precision, and completeness need to be considered. The visual analysis techniques themselves need to visually convey the quality and relevance of the data and to secure in this way the quality of the gained knowledge.

3.4 Infrastructure Issues

Finally, building an IT infrastructure for digital classics and other digital humanities also is a major challenge to computer science, as there seem to be as many data formats and user interfaces as there are use case scenarios. Often, the tools and resources are designed and suitable just for a dedicated research question. Typically, researchers in the field of Classics are working with more than one tool or web site at once and have to frequently switch between them. Hence, we need to find a way to efficiently deal with distributed and heterogeneous data resources. Most text resources differ in format as they are stored in proprietary and non-standard formats. Only some of the available

resources support the epiDOC standard, the extension of TEI P4 for epigraphic annotations like the Leiden Convention. In addition, the integration of text and image data poses a challenge that equally opens new perspectives as it is difficult. From a research infrastructure point of view, most of the tools for searching are standalone and only allow for local applications. Very rarely text data are complemented with more advanced natural language processing (NLP) applications such as *Morpheus*, the morphological analyzer for ancient Greek and Latin, or the Latin Treebank, that are both part of the Perseus project. In summary, we notice a lack of a suitable component-based and service-oriented framework for language-based resources in the eHumanities. But this in turn also poses a challenge to computer science to develop an infrastructure for the systematic and structured acquisition, generation, processing, administration, presentation, reuse, and publication of *contents*. Content services make available the resources and programs needed for that. Public digital text and data resources may be linked together and made accessible by common standards, while new software architectures integrate digital resources and processing tools to develop new and better access to digital contents.

Literature

[BGH10] Büchler, M.; Geßner, A.; Heyer, G., Eckart, T.: *Detection of Citations and Textual Reuse on Ancient Greek Texts and its Applications in the Classical Studies: eAQUA Project.* In: *Proceedings of the Digital Humanities Conference 2010,* King's College London: London 2010.

[HS10] Heyer, G.; Schubert, C.: *Detection of Citations and Textual Reuse on Ancient Greek Texts and its Applications in the Classical Studies: eAQUA Project.* In: *Proceedings of the Digital Humanities Conference 2010,* King's College London: London 2010.

[Ca05] McCarthy, W.: *Humanities Computing*, Basingstoke (U.K.): Palgrave 2005.

eHumanities - können wir die Komplexität beherrschen?

P. Wittenburg, D. Broeder
MPI für Psycholinguistik
peter.wittenburg@mpi.nl

T. Zastrow, E. Hinrichs
Universität Tübingen

1 Einführung

Sprachressourcen und Werkzeuge entstehen in einem distribuierten und nicht reglementierten Prozess. Diese verteilte Entstehung ist zugleich ein Segen als auch ein Fluch. Denn einerseits wird garantiert, dass nur die Ressourcen und Werkzeuge entstehen, die für die direkten wissenschaftlichen Fragestellungen benötigt werden. Andererseits kann man vermuten, dass verschiedene Ressourcen und Werkzeuge für gleiche Aufgabenstellungen dauernd neu erfunden werden, und dass hinsichtlich Datenformaten und Programmierschnittstellen wenig zusammenpasst. Will man z.B. eine virtuelle Kollektion aus verschiedenen Ressourcen von verschiedenen Linguisten oder eine Verarbeitungskette von Sprachtechnologie-Komponenten verschiedener Teams bilden, so werden diese Vorhaben in den allermeisten Fällen an Inkompatibilitäten oder unterschiedlichen Datenstrukturen scheitern.

e-Humanities, abgeleitet aus dem Begriff e-Science, beschreibt nun genau das daten-basierte Paradigma, in dem Daten und Operationen verschiedenen Ursprungs in einfacher Weise durch jeden Wissenschaftler zusammengefügt werden können, in dem die existierenden Beschränkungen und Hürden - wenn nicht vollkommen abgebaut - so doch erheblich reduziert werden, und in dem die Wissenschaftler kollaborative Umgebungen finden, die eine Annotation, eine Verlinkung und Anreicherung der Daten ermöglichen. Ziel des ganzen Unternehmens ist die Hoffnung, dass sich viele der sogenannten kleinen wissenschaftlichen Fragestellungen schneller beantworten lassen und dass sich darüber hinaus die großen Herausforderungen dieser Zeit, die einerseits mit der Instabilität der Gesellschaften und des Lebens und andererseits mit der zunehmenden Verzahnung unserer Steuerungsmechnismen und der Komplexität unserer Modelle zu tun haben, besser angehen lassen.

Basis eines erfolgreichen e-Humanities Szenarios ist demzufolge eine einheitliche, harmonisierende Infrastruktur, die die beschriebene Fragmentierung überwindet, ohne dabei die verteilten, kreativen und durch eine lange Tradition gefestigten Erzeugungsprozesse zu blockieren. Die sich stellende Frage ist, wie man diese zwei Pole miteinander verknüpfen kann. In den vergangenen Jahren haben wir bereits erleben können, wie umfangreiche Projekte in der Wissenschaft und auch große Firmen diese Frage sehr zum Vorteil der Benutzer bzw. der Gesellschafter haben lösen können.

Im Perseus Digital Library Projekt[1] wird z.B. umfangreiches Material von verschiedenen Quellen über die Geschichte, Literatur und Kultur der griechisch-römischen Zeit in einem uniformen technologischen Rahmen zusammengebracht, so dass es ohne Zweifel zu einer großartigen Fundgrube für die Wissenschaft geworden ist. Von großen Firmen wie FEDEX[2] ist bekannt, dass auch sie mit vielen verschiedenen Partnern zusammenarbeiten und nunmehr einen einheitlichen semantischen und syntaktischen Rahmen definiert haben, um die komplexer werdenden Management- und Austausch-Prozesse auch in Zukunft bewältigen zu können. Bei beiden Beispielen handelt es sich um projektartige Vorgehensweisen, die natürlich jedes in sich selbst wiederum Inseln erzeugen. Allerdings haben Lösungen dieser Art den Vorteil, dass der Rahmen für Vereinbarungen beschränkt ist und im Falle der Firma auch von oben durchgesetzt werden kann.

Eine komplementäre Vorgehensweise ist die, die der ESFRI Roadmap Prozess[3] ausgewählt hat und somit der Definition von J. Taylor folgt: "eScience is about global collaboration in key areas of science and the next generation of infrastructures that will enable it". Dieser Interpretation zufolge bedarf es großer Anstrengungen, um zunächst disziplinspezifisch integrierende und interoperable Infrastrukturen aufzubauen. Sie sind nicht mehr projektgetrieben, sondern sollen disziplinweit die erforderlichen Harmonisierungen erreichen. Auch hier haben wir es mit neuen "Inseln" zu tun, die allerdings breiter gefasst sind. Dabei ist das, was als Disziplin umschrieben wird, durchaus sehr heterogen – man kann vermuten, dass der ESFRI Prozess zunächst erst einmal gute Teams arbeiten lassen will, um dann zu schauen, was letztlich erfolgreich ist und was nicht. In einem Gebiet, in dem Neuland beschritten wird, ist dies sicherlich keine schlechte Strategie. Für den Bereich der Humanities werden im Rahmen der ESFRI Roadmap zwei Infrastrukturprojekte gefördert.

Auf der einen Seite steht das CLARIN Projekt[4], das zunächst nur die vergleichsweise überschaubare Schar von Wissenschaftlern aus der Linguistik und der Sprachtechnologie zu gemeinsamen Absprachen bringen will, die sich mit der Erzeugung von Sprachressourcen und deren Analyse mittels State-of-the-Art Werkzeugen befassen und die bereits auf eine lange Tradition von Meetings und Konferenzen verweisen können. Auf der anderen Seite steht das DARIAH Projekt[5], das sich zum Ziel setzt, die Wissenschaftler aus den Geisteswissenschaften zusammenzubringen, eine ungleich größere und viel heterogenere Gruppe. Zu Anfang wurde auch die EROHS Initiative diskutiert, die direkt alle Sozialwissenschaften und die Humanities zusammenbringen wollte – allerdings haben viele Entscheidungsträger sehr schnell begriffen, dass die Heterogenität bezüglich Terminologie, des Einsatzes der IT, der Organisations-Traditionen und vielerlei mehr zu gross ist.

[1] www.perseus.tufts.edu
[2] www.fedex.com
[3] http://ec.europa.eu/research/infrastructures/index_en.cfm?pg=esfri
[4] www.clarin.eu
[5] www.dariah.eu

Allerdings ist auch allen klar, dass es viele überlappende Aktivitäten geben würde, wenn es keinen Austausch zwischen den verschiedenen Forschungs-Infrastrukturen und den sogenannten horizontalen e-Infrastrukturen (GEANT-Netzwerke[6], EGI-Grids[7], DEISA/PRACE[8]-High-Performance Computing, Data Preservation and Access) geben würde. Dies ist der Anlass für die europäische Kommission und ESFRI, einerseits sogenannte Cluster Initiativen auf den Weg zu bringen, in denen die Initiativen aus den verschiedenen Gebieten (Social Sciences and Humanities, Life Sciences, Environmental Sciences, Particle Physics) zusammenarbeiten und überlappende Ansätze identifizieren und gemeinsam lösen. Alle diese Aktivitäten, wie auch die Zusammenarbeiten mit den e-Infrastrukturen dienen natürlich dem Ziel, dem Ideal eines Öko-Systems der Infrastrukturen näher zu kommen. Der Preis, der zu zahlen ist, ist ein erheblich höherer Synchronisierungs-Aufwand und bisher kann keine Aussage über Erfolg oder Misserfolg dieser Ansätze gemacht werden.

Die entscheidende Frage, die sich in Zusammenhang mit dem Zitat von J. Taylor ergibt, ist demzufolge die nach dem "Scope" dieser Initiativen oder in anderen Worten – die nach dem Grad an Komplexität, der für uns sinnvoll beherrschbar ist. Es gibt genug Projekte, die als Beispiele genommen werden können, um zu zeigen, dass Harmonisierungen erfolgreich in die Tat umgesetzt werden können. Je kleiner der Umfang derartiger Projekte ist, umso einfacher lassen sich Vereinbarungen erzielen, desto höher wird allerdings die Fragmentierung außerhalb der Projekte sein. Auf dem anderen Ende der Skala steht die Anforderung, allgemein gültige Standards zu vereinbaren, an die sich jeder halten muss.

Am Beispiel von CLARIN werden wir zeigen, welche Anforderungen es zu lösen galt und gilt und welche Komplexität damit verbunden ist.

2 Ebenen der Fragmentierung

In der Vergangenheit wurden sprachwissenschaftliche Tools als Kommandozeilen- oder Desktop-Anwendungen zur Verfügung gestellt. Der Anwender musste diese herunterladen und auf seinem lokalen Computer installieren, was durch fehlende Bibliotheken, Inkompatibilitäten etc. zu vielfältigen Problemen führen konnte. Heute bietet das Web 2.0 mit seinen mannigfaltigen Optionen zur Interaktion Mittel und Wege, Applikationen und Daten online zur Verfügung zu stellen, so dass sie vom Benutzer ohne umfangreiche Vorarbeiten direkt im Browser benutzt werden können:

- Webservices stellen verteilt und über das Internet zugänglich einzelne Funktionen zur Verfügung.

[6] http://www.geant.net/pages/home.aspx
[7] http://web.eu-egi.eu/
[8] http://www.deisa.eu/

- Die von den Webservices bereitgestellten Funktionen werden in Webapplikationen gebündelt und mit einer benutzerfreundlichen Oberfläche versehen.

Die so entstandenen Web 2.0-Applikationen und Webservices stellen ganz neue Paradigmen der Entwicklung und Anwendung sprachwissenschaftlicher Daten und Tools dar[9]. Dabei ist der Übergang von rein datengebundenen Angeboten zu online verfügbaren Anwendungen fliessend: Meist treten sie in Kombination miteinander auf, spezielle Tools erlauben den bequemen Zugriff auf die dahinterstehenden Daten. Es ergibt sich die Situation, dass Datenbestände ohne die auf sie zugeschnittenen Tools nicht mehr zugänglich sind und umgekehrt, hochspezialisierte Tools nur im Kontext der jeweiligen Datenbanken anwendbar sind. Hieraus ergeben sich einige neue Fragestellungen, wie die nach einer möglichen Harmonisierung vorhandener Online-Ressourcen und deren Zugänglichkeit bzw. Sichtbarkeit nach aussen.

Bereits heute sieht sich der Benutzer im Netz mit einer Flut verschiedener Anmelde- und Authentifizierungs-Mechanismen konfrontiert. Dies geht einher mit einer immer grösser werdenden Zahl an Benutzernamen und Passwörtern, ohne die der Benutzer keinen Zugang zu seinen personalisierten Daten oder Anwendungen erhält. Hier bietet sich die Anwendung sogenannter "Single-Sign-On"-Systeme (SSO) an: Der Benutzer authentifiziert sich einmalig bei seinem "Identity Provider". Dieser bestätigt nun im Folgenden die Identität des Benutzers gegenüber einer beliebigen Anzahl von "Service Providern", die die eigentliche Funktionalität zur Verfügung stellen. Somit ist gewährleistet, dass der Benutzer nur einen Account mit Username und Passwort zu verwalten hat und gleichzeitig die Service-Provider keine eigenen Nutzer- und Rechte-Datenbanken führen müssen. Soll ein Service Provider auch die Nutzer anderer Identity Provider akzeptieren, so müssen die beteiligten Identity Provider in einer auf gegenseitigem Vertrauen beruhenden Föderation zusammengeschlossen werden.

Ein Nachteil der SSO-Systeme besteht darin, dass ein Nutzer einen Account bei einem registrierten Identity Provider innerhalb der jeweiligen Föderation benötigt, um auf die Angebote der Service Provider zugreifen zu können. Hat er einen solchen nicht, bleiben ihm alle per SSO geschützten Anwendungen verschlossen. Desweiteren stellt ein SSO-System lediglich Funktionen zur Authentifizierung der jeweiligen User zur Verfügung, hiermit ist kein User- oder sonstiges Rechtemanagement verbunden. Ist ein Benutzer einem Service Provider gegenüber authentfiziert, so ist es immer noch diesem überlassen, welche Rechte/Zugriffe dieser dem User einräumt bzw. welche er unterbindet.

[9] Als Beispiel für eine sprachwissenschaftliche, Web 2.0-basierte Service Oriented Architecture sei hier WebLicht genannt: http://www.d-spin.org/weblicht.shtml

Die CLARIN "Service Provider Federation" ist basiert auf dem SAML2[10] Protokoll-Standard, das von Software-Paketen wie z.B. Shibboleth[11] und SimpleSAMLPHP[12] unterstützt wird und einen europaweiten Zusammenschluss nationaler Identitäts-Föderationen ermöglicht[13]. Es wird somit ein grenz- und disziplinübergreifender Zugriff auf europaweit verteilte Tools und Ressourcen ermöglicht. Ist der Benutzer einmalig in der Identitäts-Föderation authentifiziert, stehen ihm eine Vielzahl Web 2.0-basierter Anwendungen und Webservices zur Verfügung. Damit diese nahtlos ineinander greifen können, gilt es weitere Barrieren zu überwinden.

Um Daten zwischen heterogenen Anwendungen austauschen zu können, müssen diese jeweils die gleichen Datenformate verwenden. Diese Problematik umfasst nicht nur den Bereich der "äusseren", die eigentlichen Daten umfassenden (Transport-) Strukturen wie beispielsweise XML-Formate, sondern auch die Beschaffenheit der eigentlichen linguistischen Ressourcen. Dies wird besonders im sprachübergreifenden, europäischen Kontext deutlich, da unterschiedliche Sprachen unterschiedliche Zeichensätze (Encodings) und damit auch Schriftsysteme und -arten zu ihrer Darstellung benötigen. Im sprachwissenschaftlichen Bereich sind in den letzten Jahren viele Datenformate für die unterschiedlichsten Zwecke entwickelt worden. Kann man die Verwendung von UTF-8 als Zeichensatz-Encoding und XML als strukturierendes Containerformat noch als weitgehend akzeptierten Konsens in der Community ansehen, so beginnen die Inkompatibilitäten spätestens bei der Definition ressourcenspezifischer Datenstrukturen. Diese wurden in der Vergangenheit häufig auf einen bestimmten Anwendungszweck bzw. eine bestimmte Ressource hin zugeschnitten und sind dementsprechend nicht kompatibel mit anderen, durchaus ähnlichen Ressourcen.

Diese untereinander zu harmonisieren und verbindliche und tragfähige Datenstandards zu entwickeln, ist Aufgabe der Arbeitsgruppe TC 37/SC4 "Language Resources Management"[14] innerhalb der "International Organization for Standardization" (ISO)[15]. Deren Arbeit geschieht in enger Abstimmung mit weiteren international etablierten Gremien, beispielsweise der "Text Encoding Initiative" (TEI)[16].

Hat der Benutzer mit Hilfe der oben genannten Anwendungen wissenschaftlich relevante Ergebnisse erzielt, so müssen diese anschliessend publiziert bzw. der Community zugänglich gemacht werden. Dies gilt für "klassische" Publikations-Arten wie Papers, Zeitschriftenbeiträge, Monographien und ähnlichem. Neu hinzu kommt die Notwendigkeit, auch Primärdaten und sekundäre Zwischenergebnisse der Untersuchungen nachhaltig zugänglich zu machen: Ohne diese ist eine Wiederverwendung und z.B. die Nachvollziehbarkeit der publizierten Ergebnisse nicht mehr gewährleistet. Aus dieser Anforderung ergibt sich die Notwendigkeit, auch für grössere Datenmengen nachhaltig zugänglichen und entsprechend abgesicherten Speicherplatz zur Verfügung stellen zu müssen. Dieser muss über die reine Storage-

[10] http://saml.xml.org/
[11] http://shibboleth.internet2.edu/
[12] http://rnd.feide.no/simplesamlphp
[13] Zum aktuellen Stand der CLARIN Service Provider Federation, siehe http://www.clarin.eu/node/2965
[14] http://www.tc37sc4.org/index.php
[15] http://www.iso.org/iso/home.html
[16] http://www.tei-c.org/index.xml

Funktionalität hinausgehend eine Versionierung sowie die Möglichkeit, Dokumente im Team bearbeiten zu können, bereit stellen (Collaboration).

Um Ressourcen, Tools und Publikationen in den einschlägigen Such- und Registratursystemen auffindbar zu machen, müssen diese mit persistenten Identifiern (PID) und beschreibenden Metadaten versehen werden. Auch hierfür bedarf es weitgehender Absprachen, denn bisher passte wenig zusammen: (a) Zumeist gibt es überhaupt keine online verfügbaren Metadaten; (b) Oftmals sind für die beschriebenen Ressourcen und Werkzeuge nur wenige Metadaten-Elemente vorhanden; (c) Andere Metadaten sind mittels nicht-standardisierter Elemente und Vokabularien beschrieben und somit nicht kompatibel mit zentralen Metadaten-Registraturen. Sind die entsprechenden Metadaten-Strukturen allerdings standardisiert und registriert, können sie auf Grundlage etablierter Austausch-Protokolle wie OAI-PMH[17] automatisiert von weiteren Registratursystemen übernommen werden (Harvesting).

Das Projekt CLARIN versucht, die in diesem Kapitel dargestellten Aufgabenstellungen exemplarisch im Bereich der Sprachwissenschaft und Linguistik anzugehen und zu lösen. Dabei wurde immer wieder deutlich, dass auch andere Fachdisziplinen mit den gleichen oder ähnlichen Problemen zu kämpfen haben. Gleichzeitig werden sprachwissenschaftliche Ressourcen und Daten durch die erhöhte Zugänglichkeit und Sichtbarkeit auch für Forscher aus anderen Bereichen der Geistes- und Sozialwissenschaften interessant. Diese tragen neue Fragestellungen und Anforderungen an die sprachwissenschaftlichen Daten und Tools heran, so dass die Komplexität der zugrunde liegenden Infrastruktur weiter zunimmt.

3 Lösungsansätze von CLARIN

CLARIN kann bereits auf eine lange Diskussion über Standards und Vereinbarungen verweisen, d.h. in weiten Kreisen gab es bereits eine terminologische Abstimmung, die über die Jahre gewachsen ist und zu einer verschärften Bewusstseinsbildung über Begriffe geführt hat. Als ein Beispiel kann man hier auf die im Jahre 1993 begonnenen EAGLES[18] Diskussionen verweisen, die später im ISLE Projekt[19] weitergeführt wurden. Die 2-jährig stattfindende und im Jahre 1998 begonnene LREC[20] Konferenz hat ebenfalls zu einer Kohärenz beigetragen. Bereits auf der ersten LREC Konferenz gab es einen Workshop zu dem Thema “sharing of resources”.

[17] http://www.openarchives.org/
[18] http://www.ilc.cnr.it/EAGLES/intro.html
[19] http://www.mpi.nl/ISLE/
[20] http://www.lrec-conf.org/lrec2010/

3.1 Standardisierung

Es war daher folgerichtig, dass die Community als Ergebnis der vielen Diskussionen über eine verbesserte Interoperabilität im Jahre 2002 das ISO SubCommittee ISO TC37/SC4 zum Thema "Language Ressource Management" gründete – also 6 Jahre vor dem Beginn der CLARIN Initiative. Die Schwerpunkte lagen auf (a) der Etablierung einer Registratur linguistischer Konzept als eine Referenz-Sammlung, auf die jeder Linguist bei seiner Arbeit verweisen kann und (b) der Absprache von Formaten/Strukturen verschiedener wesentlicher Datentypen wie z.B. Annotationstrukturen und Lexika. Damit wurden die Kernaspekte einer jeden Interoperabilität angesprochen: die Verwendung expliziter und allgemein-verwendbarer Syntax und der Verweis auf weithin akzeptierte semantische Definitionen linguistischer Kategorien.

Damit ist jedoch nur ein Grundstein gelegt und nach 8 Jahren Diskussion können wir auch auf erste verwendbare Ergebnisse verweisen. Die auf dem allgemeinen Modell ISO 12620 basierende Konzept-Registratur[21] ist im Einsatz und bereits mit vielen Kategorien gefüllt. Die Tatsache, dass das Modell relativ eingeschränkt ist, ist einerseits seine große Schwäche und andererseits sein großer Vorteil. Die Definition von Kategorien ist zeitlich überschaubar und damit machbar, allerdings kritisieren einige Experten seine Beschränktheit und hätten anstelle dessen z.B. lieber ein flexibles erweiterbares System von RDF Assertions gesehen. Auch hier galt es wiederum, die Frage nach dem zu stellen, was innerhalb eines begrenzten Zeitfensters machbar ist, um dennoch einen riesigen Schritt vorwärts in Richtung auf Interoperabilität machen zu können.

Ebenso wurde mit der Arbeit an flexiblen und daher generischen Formaten begonnen. Das Lexical Markup Framework[22] kann am ehesten mit einem Baukastensystem verglichen werden, da es dem Wissenschaftler erlaubt, jede gewünschte lexikalische Struktur konstruieren zu können. Einige Experten sagen zurecht, dass uns diese Flexibilität bezüglich der Interoperabilität nicht weiterbringen kann, solange semantische Kategorien in jeden möglichen syntaktischen Kontext eingebunden werden können und sich damit einer einheitlichen Interpretation entziehen. Sie argumentieren für mehr strukturelle Beschränkungen, um den Interpretationsaufwand gering zu halten und folgen damit einer Kritik, die bereits gegenüber dem äußerst flexiblen TEI Modell geäußert wurde. Der neue Schritt in ISO ist jedoch die Kontext-unabhängige Definition von Kategorien, die möglichst weitgehend z.B. in Lexika verwendet werden sollen bzw. auf die aus dem lexikalischen Schema heraus referiert werden soll. Nicht immer werden sich dadurch Interpretations-Unterschiede aufgrund der kontextuellen Einbettung vermeiden lassen. Daher können Empfehlungen für bestimmte Komponenten für Sub-Communities durchaus hilfreich sein.

Insgesamt ist der ISO Prozess ein sehr langwieriger und aufwendiger, und zudem ist vollkommen unklar, ob die entwickelten Standards sich in der Community durchsetzen werden. Das bedeutet auch, dass das Entwickeln von Tools, die auf den sich entwickelnden Standards beruhen, ebenfalls langwierig und nicht ohne Risiken ist.

[21] http://www.isocat.org/
[22] http://www.lexicalmarkupframework.org/

3.2 Temporäre Lösungen

Aufgrund des innovativen Charakters vieler wissenschaftlicher Projekte und auch des hohen Zeitdruckes, dem sie sich stellen müssen, werden sich immer wieder neue Lösungen z.B. auch für Formate ergeben. Dies konnte z.B. in den vergangenen Monaten im WebLicht[23] Teilprojekt [Hi10] von CLARIN gesehen werden, wo man ein einfaches Format benötigte, das bei Prozessketten mitgeführt wird und alle für die einzelnen Schritte erforderlichen Informationen enthält. Früher haben diese zumeist um Tools oder spezielle Ressourcen herum entstandenen Formate nahezu selbstverständlich zu einer eigenen User Community und zu einer eigenen Dynamik geführt, wenn das Tool breit angenommen wurde. Gute Beispiele sind hier z.B. die CHILDES Initiative[24], die bei vielen Sprach-Akquisitionsforschern und darüberhinaus populär war und ist, und das Toolbox Tool[25], welches für Feldforscher sehr attraktiv ist, da es Annotationen und Lexikon sehr eng zu koppeln gestattet.

Natürlich besteht heute die Sorge, dass sich ein neues Format wiederum verselbständigt und das, obwohl die Community sich gerade im Prozess der Einigung auf Standards befindet. Wir erkennen immer wieder diesen Zwiespalt zwischen einerseits dem Druck, der Innovation nachzugeben, und andererseits der theoretischen Einsicht, dass wir die Vielfalt eingrenzen müssen, um die Fragmentation zu überwinden, die Gesamtkomplexität handhaben zu können und die Nutzer nicht mit einer unüberschaubaren Flut von Konvertoren zu überfordern. Natürlich müssen wir als Community den allgemeinen vorherrschenden Einstellungen entgegenwirken, die sich in verschiedenen Formen ausdrücken können: (1) Wir wissen es sowieso besser; (2) Der Aufwand, um alle Dokumente über einen Standard zu lesen, ist zu hoch; (3) Der Standard entspricht nicht ganz meinen Erwartungen; etc.

Kurzfristig ist das Arbeiten mit neuen ad-hoc Formaten sicherlich vertretbar, insbesondere dann, wenn es sich um Tool-interne Formate handelt und wenn das interne Austauschformat mit relevanten ISO Standards voll kompatibel ist, wie im Fall von WebLicht mit dem ISO/DIS 24612 Standard LAF[26]. In dem Augenblick, in dem Formate eine Infrastruktur bereichern sollen, bedarf es einer sorgfältigen Überprüfung ihrer potentiellen Bedeutung im Vergleich zu den bereits existierenden. Erschwert wird eine derartige Überprüfung dadurch, dass wir die wissenschaftliche Innovation nicht behindern sollten.

3.3 Bedeutung generischer Lösungen

Für viele Bereiche gibt es sogenannte generische Lösungen, d.h. Lösungen die auf abstrakten IT-Prinzipien und Überzeugungen beruhen. Um die Problematik zu erläutern, werden wir Beispiele aus dem Bereich Metadaten verwenden.

[23] http://weblicht.sfs.uni-tuebingen.de/weblicht.shtml
[24] http://childes.psy.cmu.edu/
[25] http://www.sil.org/computing/toolbox/
[26] www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=37326

Die im Bibliotheks-Umfeld entstandene DublinCore Initiative[27] hat sich sehr stark um eine Vereinbarung über orthogonale Elemente bemüht, die die Beschreibung aller wesentlichen Eigenschaften von web-basierten Ressourcen erlaubt. Herausgekommen war zunächst ein Satz von 15 Elementen, deren semantische Definitionen notwendigerweise sehr weitgefasst und unpräzise waren. Schon bald wurde der Nachteil derart unspezifischer Definitionen deutlich und es wurden eine Vielzahl von sogenannten Qualified DCMI Elementen hinzugefügt. Obwohl dies sicherlich ein notwendiger Schritt war, konnte dies aus der Sicht wissenschaftlicher Anwendungen nicht zufriedenstellen. Wissenschaftler sehen Metadaten nicht nur als Informationen, mit denen man Ressourcen finden kann, sondern sie sehen sie vor allem als ein Mittel um daten-basierte Forschung betreiben zu können. Sie wollen ihre über viele Jahre gewachsenen Klassifikations-Systeme und Terminologien behalten und sie möchten auch vermehrt mittels automatischer Methoden Selektionen durchführen können, wie z.B. zu einem gegebenen komplexen Ressource-Bündel das passende Werkzeug finden. Folgerichtig entstanden nicht nur in der Linguistik spezifische, auf die Diszplin zugeschnittene Element Sets, die dann natürlich wieder semantisch auf DCMI abgebildet werden mussten, um eine allgemeine Interoperabilität zu erhalten.

Mit CERIF[28] wurde ein recht weitgehendes Modell für Metadaten entwickelt, das auch kontextuelle Informationen z.B. über Personen, Projekte und Publikationen aufnehmen kann. Dieser interessante Ansatz ist momentan für viele Communities nicht primär von Bedeutung, da sie ihre Lösungen zur Beschreibung ihrer Ressourcen und Werkzeuge weitgehend gefunden haben. Man muss hier abwarten, welche Methoden verwendet werden, um z.B. Beschreibungen von Organisationen, Personen und dergleichen zu modellieren und mit denen der Ressourcen zu verbinden. Noch ist nicht klar, ob man aus Komplexitäts-Gründen nicht getrennte Systeme pflegen wird.

Ein weiteres Beispiel ist der UDDI[29] Vorschlag, der in Zusammenhang mit Web Services vom W3C ausgearbeitet wurde. Auch hier gilt, dass die Communities mit ihren spezifischen Ansätzen weit vorangeschritten sind und den generischen UDDI Rahmen als zu komplex und ungeeignet ansehen.

Wie auch andere disziplin-basierte Initiativen hat auch CLARIN diese “generischen” Lösungen nicht übernommen und damit nicht im Sinne einer allgemeinen Interoperabilität gehandelt, sondern die Eigeninteressen, die sich aus der konkreten Forschungspraxis ergeben, höher gewichtet.

[27] http://dublincore.org/
[28] http://cordis.europa.eu/cerif/src/annexes/annex3.htm
[29] http://www.uddi.org/pubs/uddi_v3.htm

4 Ungelöste Probleme

Initiativen wie z.B. Europeana[30] oder auch das Virtual Language Observatory (VLO)[31], als Teil von CLARIN, zeigen uns, wo die großen praktischen Probleme der Integration und Interoperabilität liegen. Beide Initiativen befassen sich mit dem Bilden breit angelegter Portale, auf denen interessierte Wissenschaftler, Studierende, Bürger etc. nach digitalen Ressourcen suchen können. Europeana will ein Portal sein, das auf die verschiedensten Kollektionen von Bibliotheken, Archiven und Museen verweisen kann. Es hat damit einen weitaus umfassenderen Anspruch als das VLO Portal, das auf alle möglichen Sprach-Ressourcen und –Werkzeuge verweisen will.

Beide Initiativen haben das technische Problem des Harvestens von Metadaten und deren Zusammenfügen in einem großen Index weitgehendst gelöst – auch im Sinne einer Skalierbarkeit. VLO umfasst nunmehr bereits Informationen über mehr als 270.000 Ressourcen und Tools, die aus vielen Teilen der Welt kommen. Europeana spricht sogar über mehr als 6 Millionen Digitaler Items, die registriert und mittels Verlinkungen erreichbar sind.

Zumindest bei der VLO können wir nicht von einem ausgereiften Portal sprechen, das wissenschaftlich umfassend nutzbar ist. Die Mängel, mit denen wir konfrontiert sind, sind vielfältiger Art:

- Die Granularität der Beschreibungen ist sehr unterschiedlich. Zu einem großen Teil referieren die enthaltenen Metadaten auf Kollektionen (Korpora) und nicht auf einzelne Ressourcen, obwohl gerade in der heutigen Zeit einzelne Objekte aus Kollektionen in verschiedener Weise und in verschiedenen Kontexten wiederverwendet werden. Noch gibt es keinerlei weithin akzeptierter Empfehlungen und natürlich ist es für Ressourcen-Anbieter einfacher, nur eine Beschreibung auf Korpus Niveau zu erzeugen. Dadurch aber ist der VLO Inhalt unbalanciert und der Benutzer kann zu einem Teil nicht zu dem individuellen Objekt gelangen und es mittels Playern betrachten.

- Die Qualität der Beschreibungen ist sehr unterschiedlich. So werden z.B. Organisations-Namen in verschiedener Weise geschrieben, was eine Suche erheblich erschwert. Eine Kuration derartiger Felder in einem dynamischen Umfeld kommt jedoch einer Sysiphus-Arbeit gleich, solange die Quellen nicht verbessert werden. Eine andere Möglichkeit für VLO besteht darin, Synonymlisten zu führen, was einen nicht unerheblichen Aufwand darstellt, denn aufgrund von Statistiken und manuellen Überprüfungen müssen immer wieder Änderungen vorgenommen werden. Noch schwieriger wird der Umgang mit Elementen, bezüglich derer es keine weithin akzeptierten Vokabulare gibt wie z.B. für “Genre”. Hier bedürfte es einer Ontologie im Hintergrund, die Begriffe aufeinander abbildet etc. Auch dies ist ein ziemlicher Aufwand und es ist von Beginn an absehbar, dass diverse Abbildungen nicht von jedem akzeptiert werden.

[30] http://www.europeana.eu/portal/
[31] http://www.europeana.eu/portal/

- Der Ausfüllungsgrad der Beschreibungen ist sehr unterschiedlich. Viele Ressourcen und Tools werden nur minimal beschrieben (Sprache, Repository, Kontaktnamen) und entziehen sich damit jeder genaueren Analyse, die z.B. eine Longitudinalstudie umfasst, bei der es interessant ist zu wissen, welches Alter die aufgenommenen Personen haben.

- Der Fehlergrad der Beschreibungen ist sehr hoch. Natürlich wird der Service Provider mit sehr vielen Tipp-Fehlern konfrontiert.

Moderne Verfahren wie z.B. der Einsatz von Facetted Browsing oder ein automatischer Profil-Vergleich lassen sich nur sinnvoll anwenden, wenn die oben genannten Mängel behoben werden. Leider sind wir von einem Zustand weit entfernt, indem wir avancierte Verfahren aus dem Semantic Web einsetzen können, um z.B. die Qualität und den Füllungsgrad von Metadaten Beschreibungen automatisch auf der Basis von Kontext-Informationen verbessern können. Momentan sind nur traditionelle nicht skalierende Verfahren einsetzbar.

Einige dieser Mängeln lassen sich mittels Scripts unterschiedlicher Art gepaart mit einem großen manuellen Aufwand beheben, was auf Dauer nicht finanzierbar ist. Grundsätzlich lassen sich diese Probleme nur durch eine Änderung des Bewusstseins über die Erfordernisse eines guten Datenmanagement in der Wissenschaft beheben. Bereits bei der Entstehung müssen Tools das Erzeugen von Metadaten unterstützen, wie es z.B. selbstverständlich für einige Elemente bei Kameras ist. Es sind die Repositorien, die bereits bei einem Deposit-Vorgang auf die Qualität und Kuration der Metadaten achten müssen. Das Etablieren einheitlicher Guidelines für eine geeignete Granularität wird viel Zeit und Überzeugung kosten, denn für viele in der Community ist die Wiederverwendbarkeit einzelner Ressourcen für verschiedene Zielsetzungen und mithin verschiedenen Kontexten kein aktuelles Thema, d.h. sie legen keinen Wert auf einen höheren Granularitätslevel zur Identifizierung und Beschreibung von Objekten.

5 Zusammenfassung

Der Begriff eHumanities ist vom Begriff eScience abgeleitet und beschreibt eine neue Phase daten-getriebener und computationell unterstützter Forschung, die von allen Möglichkeiten des Web extensiv Gebrauch macht. Wie in dem Zitat von J. Taylor angedeutet, bedingt das eHumanities Szenario einen neuen Typus Infrastruktur, der alle möglichen Barrieren bezüglich der einfachen Verwendung und Kombination von web-basierten Ressourcen und Werkzeugen überwindet.

Das Überwinden von Barrieren lässt sich in Projekt-basierten Ansätzen zielstrebig erreichen, wenn ein effizientes Projekt-Management gegeben ist, da diese im allgemeinen einen beschränkten Deckungsbereich haben. In Infrastrukturen, die notwendigerweise breiter angelegt sind, müssen andere Wege beschritten werden. In CLARIN wird auf breite Absprachen gesetzt, die in ISO Standardisierungsbemühungen münden. Dabei wird zumeist nicht auf Lösungsvorschläge mit einem disziplin-unabhängigen Gültigkeitsanspruch gesetzt, sondern auf Lösungen, die sich aus der Arbeit in der Community ergeben. Damit wird vor allem auch den Traditionen und der Innovationsdynamik in der Disziplin Rechnung getragen.

Der Aufwand, um derartige Lösungen bereits innerhalb einer Community zu definieren und zu breiter Akzeptanz zu führen ist enorm, und der Weg ist mit Risiken behaftet. Auch diese Lösungen müssen in gewisser Weise "generisch" sein, da sie für eine ganze Community gelten sollen und von den verschiedenen Anwendungsszenarien von Teil-Communities abstrahieren müssen. Konkrete Beispiele wie das Virtual Language Observatory zeigen, dass mit jeder Integration ein enormer Aufwand verbunden ist, der nur zu einem kleinen Teil technisch gelöst werden kann, aber langfristige Community-Prozesse umfasst.

Die entscheidende Frage, mit der wir bei dem Überwinden von Barrieren konfrontiert sind, ist die nach dem Abdeckungsbereich, den wir erzielen wollen, und damit nach der Komplexität, die wir innerhalb eines beschränkten Zeitfensters zu bewältigen in der Lage sind. Der Bereich eHumanities umfasst derart viele Disziplinen, die alle mit ihren eigenen Terminologien und Traditionen behaftet sind, dass alles umfassende Lösungen zunächst nur sehr allgemeiner Art sein können, wie es z.B. durch XML und UNICODE beschrieben wird. Es wird größere Zeiträume bedürfen, um gemeinsame Lösungsansätze für den gesamten Bereich der Humanities umzusetzen, die über basale Aspekte hinausgehen und direkt mit den Inhalten zu tun haben – der Form der Repräsentationen und der Codierungen der relevanten Phänomene. Diese Aspekte können von Initiativen, die Projekt-bezogen definiert sind und eine gewisse Breite haben, oder durch Infrastrukturen mit einem breiteren Deckungsgrad vorangetrieben werden.

Im Sinne eines Öko-Systems von Infrastrukturen, das aus Kostengründen letztlich unumgänglich sein wird, bedarf es dann vorsichtiger Versuche, die Erkenntnisse aus den verschiedenen Disziplinen immer wieder zusammenzufassen und nach abstrakteren Lösungen zu schauen. Da diese Abstraktion oft nur in enger Zusammenarbeit mit den wirklichen Experten geleistet werden kann, sind uns auf diesem Weg zeitliche Beschränkungen auferlegt, denn diese Experten sind auch genau diejenigen, die die Integrations- und Interoperabilitäts-Anstrengungen in den Disziplinen vorantreiben.

Referenzen

[Hi10] Hinrichs,M., Zastrow, T, Hinrichs, T.: WebLicht: Web-based LRT Services in a Distributed eScience Infrastructure, in: Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10), Valetta, Malta. Published by European Language Resources Association (ELRA), 2010

Corpus Linguistics, Treebanks and the Reinvention of Philology

David Bamman, Gregory Crane

Perseus Project, Department of Classics
Tufts University, Medford MA, 02140USA
david.bamman@tufts.edu
gregory.crane@tufts.edu

Abstract: The fields of corpus and computational linguistics address fundamental goals – and challenge us to rethink the structure – of humanistic research. All work with historical languages is, in some sense, an exercise in corpus linguistics. The Greek and Latin Treebanks illustrate changes in intellectual practice. Linguistic annotation of historical corpora serves a different community and offers a different combination of challenges and opportunities. On the one hand, historical languages such as Greek and Latin have, by definition, no native speakers. At the same time, these corpora have been, and remain, objects of intensive study. The Greek and Latin Treebanks thus have spawned three areas of activity, each of which differs from what we find in corpus linguistics and which collectively constitute a new form of intellectual activity, one that draws upon both the most traditional goals of philology and upon emerging fields such as corpus and computational linguistics.

1 Introduction

Humanists in general and students of the Greco-Roman world in particular have been working with digital materials for a generation but the emerging digital world has, in this first generation, so far exerted relatively little effect upon the goals, practices and general intellectual culture of the humanities. Students of the past have used new tools to ask the same questions and to enhance well-established activities – they have used their large collections as giant concordances and email has accelerated, rather than changed, the flow of electronic publication. The Greek and Latin Treebanks being developed by the Perseus Project at Tufts University has begun to reflect more fundamental changes. Treebanks are collections of text with extensive morphological, syntactic and similar categories of annotation and are familiar instruments for corpus and computational linguistic research. In building Treebanks for historical languages such as Greek and Latin, we found a new intellectual space that combined elements from computational and corpus linguistics and from the ancient discipline of philology. The paper below outlines work on the Treebanks and then describes the implications of this work for Greek, Latin and other historical languages.

2 Syntactic Analysis

The resurgence of statistical methods in computational linguistics over the past twenty-five years has given rise to a great investment in the creation of treebanks – large, syntactically annotated corpora. Much of the work has focused on English (Marcus et al. 1993) and other modern languages, including Czech (Hajic 1998), German (Brants et al. 2002), Spanish (Moreno et al. 2000), French (Abeillé et al. 2000), Italian (Montemagni et al. 2000) and Japanese (Kurohashi and Nagao 1998), but several have arisen recently for historical languages as well, including Middle English (Kroch and Taylor 2000), Early Modern English (Kroch et al. 2004), Old English (Taylor et al. 2003b), Medieval Portuguese (Rocio et al. 2000), Ugaritic (Zemánek 2007), and several Indo-European translations of the New Testament (Haug and Jøhndal 2008).

Funding from the National Science Foundation allowed us to begin developing a treebank for Classical Latin in 2006. The results of this work directly led to private funding for the development of a 400,000-word treebank for Ancient Greek poetry. As of July 2010, we have publicly released over 280,000 syntactically annotated words from these two languages (230,953 words of Ancient Greek and 53,143 words of Latin).[1] Since Latin and Ancient Greek are both highly inflected languages with a high degree of variability in word order, we have based our annotation style on the dependency grammar used by the Prague Dependency Treebank (Hajic 1998) for Czech (another non-projective language), which has since been widely adopted by a number of annotation projects for other languages, including Arabic (Hajic et al. 2004), Slovene (Džeroski et al. 2006) and Modern Greek (Prokopidis et al. 2005). Figure 9 illustrates one such dependency tree in the Ancient Greek Dependency Treebank, taken from the first line of Homer's Iliad.

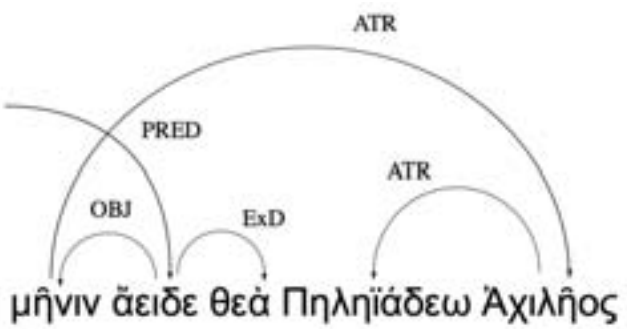


Figure 9: Dependency tree of μῆνιν ἄειδε θεὰ Πηληϊάδεω Ἀχιλῆος ("Sing, goddess, of the rage of Achilles, the son of Peleus"), Homer, *Iliad* 1.1. Arcs are drawn from heads to their dependents.

[1] All syntactically analyzed data is publicly available at:
http://nlp.perseus.tufts.edu/syntax/treebank/

3 Annotation Infrastructure

The efficient annotation of Latin and Ancient Greek is hindered by the fact that no native speakers exist and the texts that we have available are typically highly stylized in nature. To help this with problem, we have embedded our annotation environment within the Perseus Digital Library. Established in 1987 in order to construct a large, heterogeneous collection of textual and visual materials on the archaic and classical Greek world, Perseus today serves as a laboratory for digital library technologies and is also widely used by students, academics and others to access information on the Greco-Roman world (Crane 1987a, Crane 1987b, Crane 1998, Crane et al. 2006, Crane et al. forthcoming).

The scholarship that has attended historical texts since their writing has produced a wealth of contextual materials to help non-native speakers understand them, including commentaries, translations, and specialized lexica. The Perseus reading environment presents the Greek or Latin source text and contextualizes it with these secondary publications along with a morphological analysis of every word in the text and variant manuscript readings as well. Figure 10 presents a screenshot of the digital library with a syntactic annotation tool built into the interface. In the widget on the right, the source text in view (the first chunk of Tacitus' *Annales*) has been automatically segmented into sentences; an annotator can click on any sentence to assign it a syntactic annotation. Here the user has clicked on the first sentence (*Vrbem Romam a principio reges habuere*); this action brings up an annotation screen in which a partial automatic parse is provided, along with the most likely morphological analysis for each word. The annotator can then correct this automatic output and move on to the next segmented sentence, with all of the contextual resources still in view.

Our collaboration with the Alpheios Project has also allowed us to integrate a graphical treebank editor into our annotation process to make the construction of trees more intuitive and to provide annotators with greater flexibility as to their preferred input method. Figure 11 shows a tree in the process of being constructed, with a single word (Romam) being dragged onto its syntactic head.

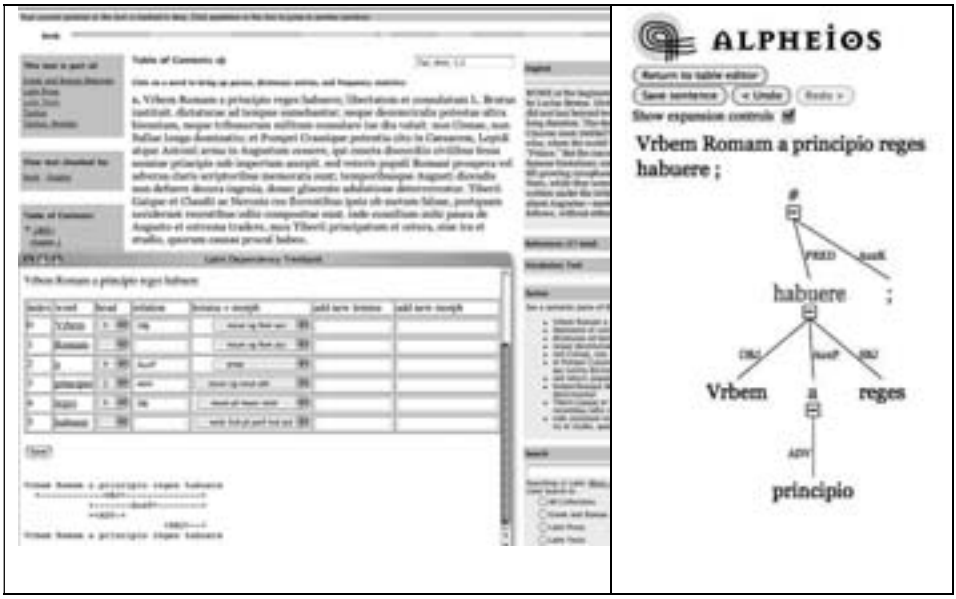


Figure 11: (left) A screenshot of Tacitus' *Annales* from the PDL; (right) Alpheios graphical treebank editor.

In addition to providing morphological analysis and digitized dictionaries such as Lewis and Short's Latin Dictionary or the LSJ, Perseus' translations and commentaries are also especially helpful in the understanding of a text. By situating our annotation environment in the middle of these contextualizing resources, we are providing support for non-native speakers of the language to maximize their contributions to the treebank, and are lowering the barriers of entry for contributing to our work. Such contextual information has greater impact on beginning students than on experts, but is of use to any annotator who wants to consult the published interpretations of authorities in the field.

By embedding our annotation environment within this online infrastructure, we have been able to build a network of annotators who are distributed not only across the United States but across the world as well (our annotators are based not only at Tufts University, the University of California-Berkeley, the University of Pennsylvania, and many other institutions in between, but are also based in Hungary, the United Kingdom and Australia). Ongoing collaboration with several Classics professors has allowed us to introduce treebanking into classrooms at Tufts University, the University of Missouri-Kansas City, Furman University, The College of the Holy Cross, and the University of Nebraska-Lincoln.

4 Methods of Annotation

In developing our work on the Latin and Ancient Greek Dependency Treebanks, we have leveraged three different method of annotation. The "classroom" production method involves soliciting annotations from students in class (e.g., a Greek course on Homer's *Iliad*), which are then reconciled by the professor; the "standard" production method involves soliciting annotations from two independent and heavily trained annotators, whose differences are then reconciled by a third; and the "scholarly" method follows the tradition of creating a critical edition, in which a single scholar with extensive training in the subject area creates a syntactic annotation for a work and is solely responsible for it as an act of interpretation.

4.1 "Classroom" Production Method

We have supported the use of treebanking in classrooms in six universities across the United States – Tufts, Brandeis, the College of the Holy Cross, Furman University, the University of Missouri at Kansas City, and the University of Nebraska at Lincoln. The primary motivation for this work has been pedagogical, since instructors and students both find the act of treebanking useful for learning complex grammatical phenomena. In addition to this fundamental utility, we have also leveraged the resulting annotations as raw material for our published treebanks. Under this method, the students provide multiple primary streams of annotation that the professor, as an expert, is then responsible for correcting and submitting.

In a study to evaluate the potential for this kind of contribution, we evaluated the annotations of a group of thirteen undergraduates at the College of the Holy Cross. Unlike the annotators in the standard production model, who undergo months of training with constant feedback on their performance, this group was provided with only limited training by their professor and access to an online handbook of annotation guidelines. While the overall inter-annotator accuracy averaged only 54.5% due to the different skill levels of students in the class, we importantly found that different students have (naturally) different skill sets – while they all can perform very well on some tasks (such as attributive modification, with an average 91.9% F-measure across the entire class), on other tasks the accuracy varies widely. Figure 12, for example, charts these users' ability to correctly identify participial attachment (i.e., distinguishing an adverbial use of a participle such as "*Reclining* on the bed, I read the book," from an attributive one, such as "the *wandering* king"). Here we see a much wider range of accuracy (reported again as F-measure), from 0% (user 12) all the way to 89.0% (user 3).

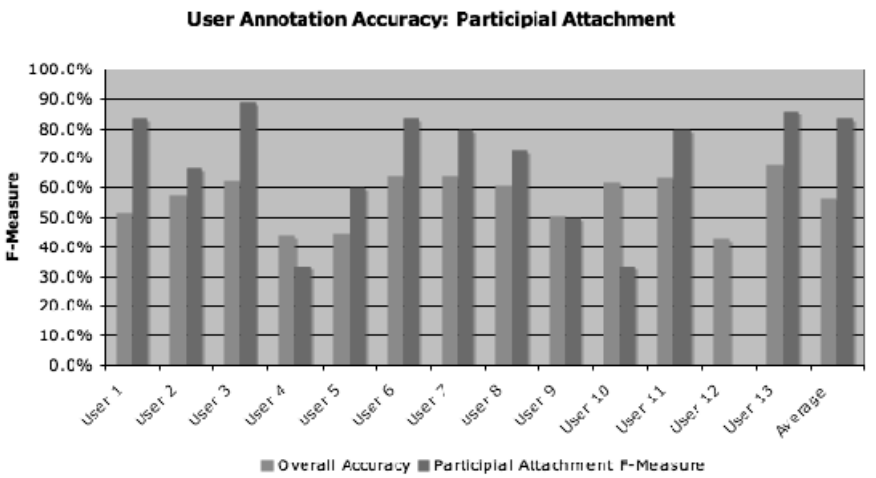


Figure 12: User annotation accuracy for participial attachment.

One pedagogical reward of incorporating treebanking into the classroom is the ability to automatically identify the strengths and weaknesses of individual students – Figure 12, for example, identifies that student 12 clearly needs more assistance in comprehending participial attachment in Greek. We can leverage this work simultaneously for the production of high-quality syntactically annotated data in two ways: first, a professor can either correct the streams of annotation produced by the students in the class, and submit that as the final, finished annotation (in which the entire class and the professor is acknowledged as the owner); second, we use the classroom annotation to help the professor identify the best-performing students, who can then go on to receive more training and provide the primary annotations in the "standard" model). "Standard" Production Method

Under the standard model of treebank production, the annotators who contribute to our existing Greek and Latin treebanks undergo extensive training with constant feedback on their performance. The backgrounds of these annotators range from advanced undergraduate students to recent PhDs and professors, with the majority being students in graduate programs in Classics. In addition to an initial training period, annotators are actively engaged in new learning by means of an online forum2 in which they can ask questions of each other and of project editors; this allows them to be kept current on the most up-to-date codifications to the annotation guidelines while also helping bring new annotators up to speed. Two independent annotators annotate every sentence and the differences are then reconciled by a third. This reconciliation (or "secondary" annotation as it is encoded in the XML release) is undertaken by a more experienced annotator/editor, typically a PhD with specialization in the particular subject area (such as Homer).

[2] The Latin and Ancient Greek forums can both be found here: http://treebank.alpheios.net/forum/

Expert analyses, however, are slow and expensive to create, especially given the difficulty and historical distance of Classical texts. The Penn Treebank can report a productivity rate of between 750 and 1000 words per hour for their annotators after four months of training (Taylor et al. 2003a) and the Penn Chinese treebank can report a rate of 240-480 words per hour (Chiou et al. 2001), but there are no native speakers of historical languages such as Greek. Our annotation speeds are therefore significantly slower, ranging from 92 words per hour to 224, with an average of 130.

4.2 "Scholarly" Production Method

With our treebank of the complete works of Aeschylus, we investigated a new mode of production: that of a single scholar completing a syntactic annotation for an entire work and treating it as an self-standing interpretation of the text.

The motivation for this work is the fundamentally different nature of historical treebanks compared to modern ones. While an article from the Wall Street Journal is certainly more representative of how native English speakers actually speak than Homer's epic Iliad is for ancient Greeks, the Iliad has been a focused object of study for almost 3,000 years, with schoolchildren and tenured professors alike scrutinizing its every word, annotating its syntax, semantics and other linguistic levels either privately in the margins of their books or as published commentaries. While ambiguity is of course present in all language, the individual ad hoc decisions that annotators make in resolving syntactic ambiguity when creating modern treebanks have, for heavily studied Classical and other historical texts, been debated for centuries; dissertations and entire careers have been made on the study of a single work of a single author.

Figures 12, 13 and 14, for example, illustrate the complexity that surrounds textual interpretation of a single of Aeschylus' Agamemnon (τὸν φρονεῖν βροτοὺς ὁδώσαντα, τὸν πάθει μάθος θέντα κυρίως ἔχειν, "[Zeus] ... who put men on the path of wisdom, who established that the law 'learning through suffering' shall be in force," lines 176-8).

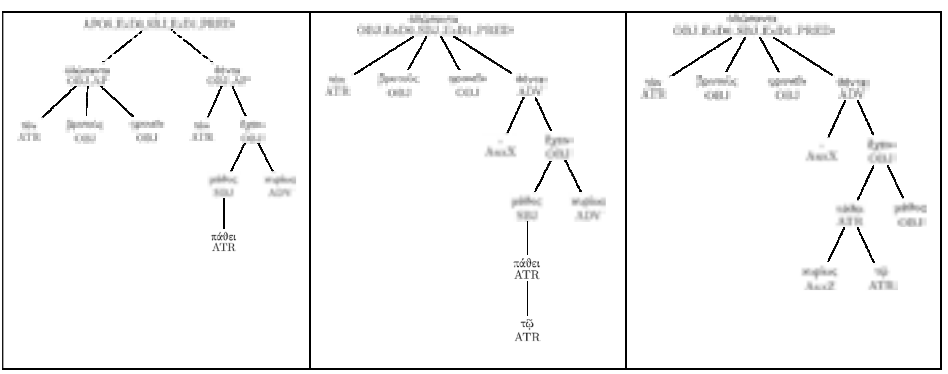


Figure 13: Three different interpretations of a sentence from Aeschylus' *Agamemnon* as machine-actionable syntactic analyses. Syntactic tree of Ag. 176-8 (Denniston-Page, Fraenkel, and Bollack).

Though the formula πάθει μάθος ("learning through suffering") is both quoted and commented upon in many general introductions to the theater of Aeschylus (it was even quoted by Robert F. Kennedy in his speech on the assassination of Martin Luther King Jr. (Kennedy 1968), both the text and syntactic interpretation of the sentence are highly controversial (Bamman et al. 2009). The three most recent commentaries on the play – Fraenkel (Fraenkel 1950), Denniston-Page (Page 1957) and Bollack (Bollack 1981) – have adopted three very different solutions based on their own weighing of the philological evidence, each resulting in a markedly different syntactic tree.

The variety of textual and syntactic interpretations for just these three lines of Aeschylus begins to point out the shortcomings of a standard treebank production model for texts of ongoing scholarly debate. In creating an annotated corpus of a language for which no native speakers exist (and for which we subsequently cannot rely on native intuitions), we are building on a mountain of prior scholarship that has shaped our fundamental understanding of the text.

5 Conclusion

The Greek and Latin Treebanks are not simply databases for research but the catalysts for new intellectual life. Two implications in particular stand out. First, they open for undergraduates in Greek and Latin opportunities similar to those familiar to their counterparts in many of the sciences for making tangible contributions. Second, the Treebanks are not just databases of industrially produced data but repositories of machine-actionable interpretations. A syntactically analyzed sentence is a new form for the publication of scholarly conclusions about language – a form that is itself largely language independent. Whether the researcher's preferred language of publication is English or German, Arabic or Chinese, the parse tree looks the same. The Greek and Latin Treebanks thus have opened new possibilities for students and for advanced researchers to participate more fully in the sudy of Greco-Roman culture than was feasible in print.

Literaturverzeichnis

Abeillé, A., L. Clement, A. Kinyon, and F. Toussenel (2000), "Building a Treebank for French," in: Proceedings of the Second Conference on Language Resources and Evaluation (Athens), pp. 87-94.

Bamman, D., F. Mambrini and G. Crane (2009), "An Ownership Model of Annotation: The Ancient Greek Dependency Treebank," in: Proceedings of the Eighth International Workshop on Treebanks and Linguistics Theories.

Bollack, J. and P. Judet de La Combe (1981), L'Agamemnon d'Eschyle: le texte et ses interprétations. Presses universitaires de Lille, Lille, 1981.

Brants, S., S. Dipper, S. Hansen, W. Lezius, and G. Smith (2002), "The TIGER Treebank," in: Proceedings of the Workshop on Treebanks and Linguistic Theories (Sozopol, Bulgaria).

Chiou, F., D. Chiang, and M. Palmer (2001), "Facilitating Treebank Annotation Using a Statistical Parser," in: Proceedings of the First International Conference on Human Language Technology Research HLT '01, pp. 1-4.

Crane, G. (1987), "Clay Balls and Compact Disks: Some Political and Economic Problems of New Storage Media," Favonius Supplement, 1, pp. 1-6.

Crane, G. (1987), "From the Old to the New: Integrating Hypertext into Traditional Scholarship," in: Hypertext '87: Proceedings of the 1st ACM conference on Hypertext, pages 51-56.

Crane, G. (1998), "New Technologies for Reading: The Lexicon and the Digital Library," Classical World, pages 471-501.

Crane, G., D. Bamman, L. Cerrato, A. Jones, D. M. Mimno, A. Packel, D. Sculley, and G. Weaver (2006), "Beyond Digital Incunabula: Modeling the Next Generation of Digital Libraries," in: Proceedings of the 10th European Conference on Research and Advanced Technology for Digital Libraries (ECDL 2006), pp. 353-366.

Crane, G., D. Bamman, and A. Jones (forthcoming), "Philology in an Electronic Age," in: Greek Lexicography After Liddell and Scott. Pre-publication version: http://geryon.perseus.tufts.edu/data/Papers and Props/Philology.pdf

Page, D. (1957) Aeschylus Agamemnon. Edited by the late John Dewar Denniston and Denys Page. Clarendon Press, Oxford.

Džeroski, S., T. Erjavec, N. Ledinek, P. Pajas, Z. Žabokrtsky and A. Žele (2006), "Towards a Slovene Dependency Treebank," in: Proceedings of the Fifth International Conference on Language Resources and Evaluation, ELRA, Genoa.

Fraenkel, E. (1950), Aeschylus. Agamemnon. Clarendon Press, Oxford, 1950

Hajič, J. (1998), "Building a Syntactically Annotated Corpus: The Prague Dependency Treebank," In : E. Hajičová, editor, Issues of Valency and Meaning. Studies in Honor of Jarmila Panevová. Prague, Charles University Press.

Hajič, J., O. Smrž, P. Zemánek, J. Šnaidauf, and E. Beška (2004), "Prague Arabic Dependency Treebank: Development in Data and Tools," In Proceedings of the NEMLAR International Conference on Arabic Language Resources and Tools.

Haug, D.T.T., and M.L. Jøhndal (2008), "Creating a Parallel Treebank of the Old Indo-European Bible Translations," In Proceedings of the Second Workshop on Language Technology for Cultural Heritage Data (LaTeCH 2008).

Kennedy, Robert F. Statement on the assassination of Martin Luther King, Indianapolis, Indiana, April 4, 1968

Kroch, A., and A. Taylor (2000), Penn-Helsinki Parsed Corpus of Middle English, second edition. http://www.ling.upenn.edu/hist-corpora/PPCME2-RELEASE-2/

Kroch, A., B. Santorini, and L. Delfs (2004), Penn-Helsinki Parsed Corpus of Early Modern English. http://www.ling.upenn.edu/hist-corpora/PPCEME-RELEASE-1

Kurohashi, S., and M. Nagao (1998), "Building a Japanese Parsed Corpus while Improving the Parsing System," Proceedings of the First International Conference on Language Resources and Evaluation (Granada).

Marcus, M. P., M. A. Marcinkiewicz and B. Santorini (1993), "Building a Large Annotated Corpus of English: The Penn Treebank," Computational Linguistics 19, pp. 313-330.

Montemagni, S., F. Barsotti, M. Battista, N. Calzolari, O. Corazzari, A. Lenci, A. Zampolli, F. Fanciulli, M. Massetani, R. Raffaelli, R. Basili, M. T. Pazienza, D. Saracino, F. Zanzotto, N. Mana, F. Pianesi, and R. Delmonte (2000), "The Italian Syntactic-Semantic Treebank: Architecture, Annotation, Tools and Evaluation," in: Proceedings of the COLING Workshop on "Linguistically Interpreted Corpora (LINC-2000).

Moreno, A., R. Grishman, S. López, F. Sánchez and S. Sekine (2000), "A Treebank of Spanish and its Application to Parsing," in: Proceedings of the Second Conference on Language Resources and Evaluation.

Prokopidis, P., E. Desipri, M. Koutsombogera, H. Papageorgiou, and S. Piperidis (2005), "Theoretical and Practical Issues in the Construction of a Greek Dependency Treebank," In Proceedings of the 4th Workshop on Treebanks and Linguistic Theories (TLT), pages 149–160.

Rocio, V, M. A. Alves, J. Gabriel Lopes, M. F. Xavier and G. Vicente (2000), "Automated Creation of a Medieval Portuguese Partial Treebank," in Anne Abeillé (ed.), Treebanks: Building and Using Parsed Corpora (Dordrecht: Kluwer Academic Publishers), pages 211-227.

Taylor, A., M. Marcus, and B. Santorini (2003), "The Penn Treebank: An Overview." In: Anne Abeille, editor, Treebanks: Building and Using Parsed Corpora, pages 5-22. Kluwer Academic Publishers.

Taylor, A., A. Warner, S. Pintzuk and F. Beths (2003). York-Toronto-Helsinki Parsed Corpus of Old English Prose. University of York.

Zemánek, Petr (2007), "A Treebank of Ugaritic: Annotating Fragmentary Attested Languages," In Proceedings of the Sixth Workshop on Treebanks and Linguistic Theories (TLT2007), pages 213–218, Bergen.

eHumanities: Intelligent Analysis and Information System for Humanities and Culture (Extended Abstract)

Sven Becker, Marion Borowski, Melanie Gnasa, Kai Stalmann, Stefan Wrobel

Fraunhofer-Institut für Intelligente Analyse- und Informationssysteme IAIS
Schloss Birlinghoven
53757 Sankt Augustin

{sven.becker, marion.borowski, melanie.gnasa, kai.stalmann,
stefan.wrobel}@iais.fraunhofer.de

1 Introduction

While electronically available data have played an important role in the natural sciences for a long time, the impact of data and information on the humanities and culture has become clear only recently. Progress in computer science has made it possible to digitize the objects of research from these disciplines so that they can be made available electronically. This digitization assists scientific discourse as well as electronic networking. Key challenges arising in this connection derive from the type of material and the overwhelming amount of objects which need to be digitized. In the following, we discuss the main challenges, goals and solutions in an exemplary fashion for the so-called German Digital Library (DDB) project. This project, being part of the European Library initiative Europeana (www.europeana.eu) is aiming at making available the core objects of 30.000 German cultural and scientific institutions in an online version. We argue that the tasks of analysis, interpretation, and scientific networking can be supported considerably by adequately designed systems that meet the requirements of the humanities and the cultural sector.

The number of cultural and scientific institutions in Germany is estimated at around 30.000. They preserve more than 30 million cultural assets and scientific documents. The exact number of books, paintings, movies, musical compositions, historical monuments, archival documents, and others is not known. Right now, only an extremely small percentage of objects is made available in digital form. Using recent technical developments, however, digitization and indexing will open new opportunities to give access to cultural and scientific objects to a broader audience.

2 Related Work

The German Digital Library will add momentum to the current trend towards digitization by implementing a portal that aggregates, analyzes and retrieves digital data with its describing metadata. In doing so, a very high resolution digital reference copy of objects is stored in a decentralized way on the web sites of the individual libraries, archives, museums and various other types of public institutions, whereas a lower resolution version is stored directly on the DDB platform to enable a quick overview and approximate browsing. Today's digital cultural and scientific assets are distributed and often only accessible via different smaller portals. The German Digital Library's intention is to aggregate the various knowledge data in one retrieval and collaboration platform.

Libraries, scientists and 'Fachinformationszentren' (FIZ, centers of specialized information) have always had strong interests to drive digitization, e-publishing and information retrieval. Space and financial limitations are strong motivations for libraries. Archives have a need for exchanging data, but slightly differ from libraries, because they usually have an awarding authority paying for a particular archival duty. Scientists have the strongest need for direct and fast access to information that may be held in a library, archive or museum on the other side of the globe. These are some of the reasons why library and information science started embracing computer science, software engineering and computational linguistics. The FIZ in Germany, some of which have become technology and service providers with competitive computer centers and strong engineering teams, have combined efforts to launch vascoda.de[1] [Pi03], a portal making some 75 million records of scientific literature accessible. Vascoda, which is hosted at Technische Informationsbibliothek und Universitätsbibliothek Hannover, contributes to WorldWideScience.org [HJ08], a US maintained web gateway providing a federated search service on databases spread over the world with more than 200 million records. A first attempt to integrate sources from different cultural heritage institutions is the BAM-Portal [Ma02] (www.bam-portal.de). It enables the search in the collections of the participating libraries, archives, museums and other sources in Germany. To date this portal manages more than 45 million digital objects; about 1 million of those objects are available in digital form.

[1] It should be remarked that the funding of this project is expired and its future service is uncertain.

Libraries, archives, and museums grow and adapt themselves to a changing world, but these institutions also are the most important holders of cultural heritage. In 2004 Google started the Google Book Project (books.google.com) that aims at porting every book in every language into the Google Digital Library. In succession of this ongoing activity many publicly founded digitization projects and digital libraries have been set off. More recent examples are the Max Planck Digital Library (mpdl.mpg.de), or the Biodiversity Heritage Library (bhl.org). One of the first digital library projects ever, the Internet Archive (archive.org), is privately founded and dates back to 1996. National libraries followed up and started crawling digital content from websites. One of the most prominent European digital library projects, Europeana (europeana.eu), was initiated in 2005 and currently holds metadata from aggregators with links to about 6 million digital items from all over Europe.

3 Challenges

The German Digital Library was set out at the end of 2009 with the long-term goal of ingesting up to 300 million items from up to 30.000 national institutions within the next decades. The project is carried out by Fraunhofer IAIS in collaboration with the German National Library, a network of cultural and scientific institutions, the FIZ Karlsruhe, and other partners. When first released to the public, which is expected to happen in 2013, the library shall serve as a web portal and also as a platform that can contribute to other projects and libraries, for example to europeana.eu.

Like Europeana, the German Digital Library is conceived of supporting networking activities of the partners involved, but other than Europeana, the German Digital Library is deemed to hold not only metadata but content also. Furthermore, it shall endeavor to contribute to the semantic web and offer value added services based on metadata and content. One of the virtues of these projects is that the ingested data could be rich and of high quality. But in fact quality may vary: certain collections contain metadata annotations at item level, others at collection level only. Page level metadata is currently hardly provided at all. This task is addressed by the CONTENTUS [Pa09] project which is an application scenario of German THESEUS research program [Th10]. It aims at digitizing text and multimedia collections in order to annotate them semantically. Besides the heterogeneity of metadata and content in terms of quality and formats, the challenges for projects like CONTENTUS, Europeana or the German Digital Library are manifold. From a usability perspective the most demanding problems are precision and recall, name disambiguation, clustering of (near) duplicates, and multilinguality of metadata and queries. First results from the CONTENTUS project for name disambiguation [Pi09] and the detection of higher level ontology concepts for words and phrases [PR09] are already available and can be applied to the German Digital Library. A custodian might for good reasons emphasize those requirements that ensure an adequate representation of objects provided. Scientific use on the other hand does not necessarily rely on easy usability and beautiful representation, but requires the ability to safely identify, localize, and evaluate items. Sharing objects in a distributed environment or workspace for collaborative work is highly desirable for scientists but may also fit well into a web2.0 and semantic web savvy user's world.

Linked data has to be maintained as certain providers come up with rich metadata that has inner links and also may refer to external targets like authority files and other resources on the web. The key challenge is the metadata mapping, partly because of the heterogeneity of the input received by providers, but also because a one-to-one mapping and harmonization for metadata and ontologies is close to impossible. Experience and best practices must be applied when transforming and consolidating formats into an internal knowledge model which is needed for clustering and reasoning. Besides from a state of the art multilingual and faceted search, not only scientists would profit from queries including transitions, like "Which person is related to event (x)?" or "When was (x) first used?" Queries like these that run against billions of triples on a public platform and that may end up with more than 1.000 parallel users are not common technical requirements.

The fact that user generated content (objects, links, valuation) can seriously enrich a platform has already been demonstrated by other projects, for example by the portal of the Australian newspapers (http://newspapers.nla.gov.au). Commercially exploitable services and products will be offered at a later stage but require an architecture well suitable for the integration of services implementing e-commerce functionality and payment systems. A cornerstone towards monetization is that IPR and other regulations are strictly respected. Sealing of digital objects, a technique firstly applied in ancient times to physical objects using wax, may be needed to ensure the authenticity of cultural heritage objects.

More advanced technical features are related to data mining and text/image/audio processing, like automatic metadata extraction from textual or binary content, clustering based on raw content and automatic linking. One of the preconditions is to tightly cooperate with those institutions that offer services like persistent identifiers and URN resolvers, that maintain authority files and vocabularies, or already do work on platforms like Fedora [PL98] and eSciDoc [Ra09], on frameworks and tools. While decentralization is part of the challenge, it also bears a clue for the solution. A lesson learned from earlier projects is that mapping and visualization of the cultural heritage objects can only be done in an adequate way by using a decentralized approach: It is the provider who has the best knowledge of the digital objects. In addition, domain specific consulting is needed to recognize and establish human task processes, patterns of usage and templates for visualization and mapping. Thus the acceptance of the project stands or falls with the usability that end users experience, but this experience is heavily reliant on understanding how cultural heritage is managed in traditional modern knowledge managing institutions.

Acknowledgement

The work presented here was funded by the Federal Government Commissioner for Culture and the Media.

References

[HJ08] Hitson, B. A.; Johnson, L. A.: WorldWideScience.org: Bringing Light to Grey. Tenth International Conference on Grey Literature: Designing the Grey Grid for Information Society, 8-9 December 2008.

[Ma02] Maier, G.: Gemeinsames Internetportal für Bibliotheken, Archive und Museen– BAM-Portal. World Library and Information Congress: 68th IFLA Council and General Conference, 2002.

[Pa09] Paaß, G. et.al.: Text Mining and Multimedia Search in a Large Content Repository. In: Proceedings of the Sabre Conference on Text Mining Services, TMS 2009, Leipzig, 2009.

[Pi03] Pianos, T.: Vascoda - a Portal for Scientific Resource Collections Created by German Libraries and Information Centres. World Library and Information Congress: 69th IFLA General Conference and Council, 2003.

[Pi09] Pilz, A., et. al.: Entity resolution by kernel methods. In: Proceedings of the Sabre Conference on Text Mining Services, TMS 2009, Leipzig, 2009.

[PL98] Payette, S.; Lagoze, C.: Flexible and Extensible Digital Object and Repository Architecture (FEDORA). Research and Advanced Technology for Digital Libraries, Volume 1513, p. 517, 1998.

[PR09] Paaß, G.; Reichartz, F.: Exploiting semantic constraints for estimating supersenses with crfs. In: Proceedings of Ninth SIAM International Conference on Data Mining 2009.

[Th10] Theseus. http://theseus-programm.de/theseus-basic-technologies.html. 2010

[Ra09] Razum, M. et. al.: eSciDoc Infrastructure: A Fedora-Based e-Research Framework. Research and Advanced Technology for Digital Libraries. Volume 5714, p227-238, 2009.

eHumanities aus Sicht eines Förderers

Helge Kahler

Referat 425 (Geistes-, Sozial- und Kulturwissenschaften, Akademien, Forschungsmuseen)
Bundesministerium für Bildung und Forschung
53170 Bonn

helge.kahler@bmbf.bund.de

Abstract: Das Bundesministerium für Bildung und Forschung (BMBF) fördert eine Reihe von Projekten aus dem Bereich der eHumanities. Der Vortrag berichtet über aktuelle und anstehende Aktivitäten.

1 eHumanities – Chance für die Geisteswissenschaften

„Die Geisteswissenschaften“ sind ein Sammelbegriff für eine Reihe von Disziplinen mit unterschiedlichen Gegenständen und Methoden. Bei allen Unterschieden wächst bei vielen Geisteswissenschaftlern die Erkenntnis, dass Informations- und Kommunikationstechnologie dazu beitragen kann, alte wissenschaftliche Fragen besser und einige neue Fragen überhaupt erst beantworten können. In diesem Sinn kann die angemessene Verwendung von Informations- und Kommunikationstechnologie die Methoden und auch die Erkenntnisse der Geisteswissenschaften erweitern und diese so zu „erweiterten Geisteswissenschaften“, i.e. „enhanced Humanities“, i.e. eHumanities machen.

Die Erweiterung kann sich auf alle Bereiche der Datenerfassung, -analyse, -verwertung und –speicherung beziehen. Gängige Beispiele sind die Digitalisierung von Text oder Bildern, ihre Annotation mit Metadaten, ihre Visualisierung oder ihre Bearbeitung mit Hilfe geeigneter Applikationen.

Das Bundesministerium für Bildung und Forschung unterstützt seit einigen Jahren ausgewählte Pilotprojekte aus dem Bereich der eHumanities und plant, diesen Bereich noch stärker in den Fokus der Projektförderung zu nehmen.

2 eHumanities – Chance für die Informatik

Die Unterstützung der Geisteswissenschaften mit Methoden der Informatik bietet aber auch der Informatik Chancen. In den vergangenen Jahrzehnten sind eine Reihe von anwendungsorientierten Spielarten der Informatik entstanden, die der Einbringung von Methoden und Werkzeugen der Informatik in anderen Disziplinen folgten: technische, praktische und angewandte Informatik oder Wirtschafts-, Medizin- und Umweltinformatik.

Die Geisteswissenschaften können einerseits von der Informatik profitieren, sie können ihr aber auch andererseits ein hochinteressantes Anwendungsfeld anbieten. Und ebenso, wie es in der Vergangenheit begabte Schüler in die Informatik zog, die an der Unterstützung von Betrieben durch Workflow-Systeme oder an der Simulation der Entstehung von Galaxien interessiert waren, können und sollen zukünftig immer mehr Informatikstudierende ihre Lieblingsanwendung in den Geisteswissenschaften finden. Dabei können sie von so unterschiedlichen Fragen getrieben sein, wie man über die Verbindung archäologischer Daten mit Georeferenzdaten eine historische Gegebenheit verortet oder die Frage, wie man mit statistischen Methoden und geeigneter Visualisierung interaktiv bestimmen kann, welcher Autor einen bestimmten aramäischen Text geschrieben hat oder wie man bei modernen Autoren oder barocken Komponisten herausfindet, wer von wem abgeschrieben hat.

Weitere Fragen werden zweifellos von kreativen Geisteswissenschaftlern und Informatikern generiert werden, wenn sie jeweils mit den Möglichkeiten, die Informatikmethoden und –werkzeuge für Geisteswissenschaften bieten, vertrauter werden.

Im Übrigen bieten die Geisteswissenschaften mit ihren großen mehr oder weniger gut strukturierten Datenmengen durchaus nicht nur eine rein feingeistige und ansonsten wenig nützliche Tätigkeit für Informatiker. Die Fähigkeit, insbesondere aus großen Text- und Bilddatenmengen sinnvolle Information zu extrahieren und Möglichkeiten zur Weiterverarbeitung zu bieten, ist mit zunehmender Digitalisierung und Vernetzung immer stärker gefragt.

In der Kooperation von Informatik und Geisteswissenschaften liegt noch ein riesiges Potential, das nicht unausgeschöpft bleiben sollte.

Specific polysemy of the brief sapiential units

Marie-Christine Bornes Varol
département Etudes Hébraïques
Inalco Paris. France

Jean-Daniel Gronoff
www.DualSemantics.Org Paris. France

Marie-Sol Ortola Querol
Université Nancy2 Nancy. France

Abstract: In this paper we explain how we deal with the problems related to the constitution of the Aliento database, the complexity of which has to do with the type of phrases we work with, the differences between languages, the type of information we want to see emerge. The correct tagging of the specific polysemy of brief sapiential units is an important step in the preparation of the text within the corpus which will be submitted to compute similarities and posterity of the units.

1 Introduction

Proverbs form an essential part of the oral and written culture of the Mediterranean basin since the oldest antiquity. They represent a specific type of philosophical knowledge. Many of these sayings, maxims and proverbs are still used nowadays.

The aim of the project Aliento[1] is thus the study of proverbs, sentences, maxims... from the Iberian Peninsula (IXth - XVth C.), their circulation between three cultures and several languages, their sources and their posterity. The spatial area has been chosen for its central role in the circulation of knowledge during the Middle Ages. The database will establish how transmission occurs from the ancient sapiential sources to the modern and contemporary Mediterranean books of proverbs. This is why we must cross and intersect texts from distinctive cultural traditions, written in different languages, annotated collectively in a uniform way. The database assembles and compares digitalized sapiential texts in Latin, Arabic, Hebrew, Greek and Romance languages and it is build and accessed by the way of the collaborative website 'Aliento.eu'. Interrogation languages are Spanish, French and English.

[1] The project Aliento (linguistic & intercultural analysis of brief sapiential statements of the Iberian Peninsula (IXth – XVth C.) and their transmission Orient / Occident & Occident / Orient) is a project from MSH Lorraine, INALCO – Paris, Université Nancy2, Belunet & DualSemantics.Org.

The extension and variety of the texts we deal with require a critical pinpointing and a precise classification of the relevant brief statements. The themes of the discourses are sapiential from beginning to end. However, what we want to see emerge is the brief sapiential units only, that is to say, a subgroup within a textual corpus. Each unit is meaningful within a context or a succession of statements from which it can't be separated. These brief units evolve depending on particular temporal periods; they become proverbs, independent statements or not.

Relevant operating parameters exist, such as the nature of the brief sapiential unit, the language and the culture which it belongs to, its dating, the geographical area which it comes from (from within the Peninsula or outside).

This information is brought by textual annotations (concerning manuscript, critical edition, and so on) relative to the language of the text, its origin, its author, its dating, all this data coming from the scholarly tradition.

The text being taken as a whole, we must build a subgroup of comparable units. An essential work is to set apart the statements to be tagged, for their relevance in terms of their structure and their significance.

We must also take into consideration the fact that a statement found in our texts becomes a "proverb" in one culture, and is not yet such a thing in another, but it belongs to a lengthy expression containing an advice or an edifying discourse.

It is important to indicate that we have decided to tag differently the different parts of the text, at this stage of our study. The narrative parts of the text left aside are indicated by the tag: <text.del>; the eventual exempla: <exemplum>; similarities: <simil>; edifying discourse: <de>. We don't rule out the possibility of coming back to them. We have categorized in a different way the sapiential statements, depending on their degree of proverbialisation: <pr?> for the units potentially proverbial; <pr>: good candidates to proverbialisation; <pr.est>: established proverbs.

To compute similarities of meaning between the multilingual brief sapiential units of the database and to compute their posterity, the knowledge representaion is based on embedded XML annotations in order to represent the specific "polysemy" of each brief sapiential unit. XML annotations allow for possible reexamination depending on the results of the research, and for tracing back the evolution, the freezing process, the puns on set expressions, the analysis of the sapiential units set in a cultural, diachronic, synchronic perspective, and so on.

2 Building a computational knowledge container for each brief sapiential unit

To show how we build the knowledge representation of the specific polysemy of brief sapiential units, we used the following list of five brief sapiential units extracted from five texts belonging to the tradition assembled in our corpus of reference:

1. (Arabic) : **وقد كنت مرتفعاً فأصبحت متّضعاً.**
2. (Spanish_1) : E ante érades mucho alto, e agora sodes mucho baxo.
3. (Spanish_2) : Quanto fue alta la su sobida tanto fue mas baxa la su cayda.
4. (Latin) : Quanto altitudo Alexandri excellencior fuit, tanto gravior est casus.
5. (Hebrew): והיית נכבד והינך שפל.

The units of the list are collected by specialists to correspond to the same kind of sapiential lesson. They belong to a group of texts closely related, as shown by scholars of the different fields of studies, which interest Aliento project directly.

This is the list of the English literal translation of the five brief sapiential units:

1. and you were on the top and you are down
2. And before you were very high, and now you are very low
3. the higher was his hight the lower was his fall
4. greater was Alexander's hight, more painful is his fall
5. And you were an eminent person and now you are down

We first link each unit to a meaningful set of information, such as translations of the literal and figurative meaning in French, English and Spanish (languages of interrogation of the database), junctions to nowadays used proverbs, which represent the principal form of questioning in current paremiological databases; the "romanisation" of the units written in languages with characters other than latin. Romanisation is a system of simplified transliteration, which avoids diacritic signs, for example. Difficulties inherent to Semitic languages graphical output have been jointly solved by Arabists and Hebraists, all members of Aliento. In order to compare units and to compute them as 'similar', we link each one to a number of information. All added information is represented as an XML entity linked to the appropriate unit.

This is for example the list of the XML entities linked to the unit "E ante érades mucho alto, e agora sodes mucho baxo" at the beginning of the process of comparison. We would like to emphasize at this point that the tags are specific to the Aliento project and that the encoding in such a way is made to create semantic clusters that will be analyzed by algorithmic calculation.

<**pr.all**>
<**pr**>e ante érades mucho alto, e agora sodes mucho baxo</**pr**>
<**pr.type.es**>cuanto más alto subas mayor será la caída</**pr.type.es**>
<**pr.type.fr**>…plus dure sera la chute</**pr.type.fr**>
<**sl.fr**>et avant tu étais très haut, et maintenant tu es très bas</**sl.fr>**
<**sl.en**>And before you were very high, and now you are very low</**sl.en**>
<**sf.fr**>Vivant vous étiez au sommet, mort vous êtes en dessous de tous</**sf.fr**>
<**sf.es**>Vivo estabas en la cumbre, muerto estas debajo de todos</**sf.es**>
<**sf.en**>Alive you were above all, dead you are below everyone</**sf.en**>
<**lec.fr**>La mort réduit le puissant à rien</**lec.fr**>
<**lec.es**>La muerte aniquila al potente</**lec.es**>
<**lec.en**> death annihilates the mighty</**lec.en**>
<**key.fr**> puissance déchéance mort chute </**key.fr**>
<**key.es**> potencia decadencia caída muerte </**key.es**>
<**key.en**> might decay fall death </**key.en**>
<**lem.es**>y antes ser muy alto ahora ser muy bajo</**lem.es**>
<**str.ling**>Sadv1 SV Sadv2 / Sadv opp à 1 SV Sadv opp à 2</**str.ling**>
<**str.form**>parallélisme binaire : avant/après haut/bas </**str.form**>
<**str.poet**>8 / 8 ass a/o</**str.poet**>
</**pr.all**>

<**pr.all**>: the pair of tags (<pr.all> at the top of the list and </pr.all> at the end of the list) includes all the XML entities wich refer to the same brief sapiential unit. <pr.all> builds a kind of knowledge container where a lot of information linked to the brief sapiential unit can be freely included. The brief sapiential unit is the first XML entity of the container.

<**pr**>: the brief sapiential unit. This unit must be unique at the scale of the project. The unicity can be computed and certified later.

<**pr.type**> characteristic proverb or expression which is directly linked to the studied unit (all languages can appear but the main languages are the three languages of interrogation of the data base)

<**sl.fr**>: literal translation of the brief sapiential unit (<pr> entity), written in Spanish, into French.

<**sl.en**>: literal translation of the brief sapiential unit into English.

<**sf._**>" stands for figurative sense, and the suffix '._>' refers to the language of the entity ('.en>' '.fr>' '.sp>').

<**lec._**>" stands for sapiential lesson, and the suffix '._>' refers to the language of the entity ('.en>' '.fr>' '.sp>').

<**key._**>" stands for ideological keywords, and the suffix '._>' refers to the language of the entity ('.en>' '.fr>' '.sp>').

<**lem.es**>: the root "<lem" stands for lemmatization the language of the sapiential unit (Spanish).

Lemmatization is the traditional way for human specialists to compare multilingual entities. It can be used also for multilingual computed comparisons when it is an XML entity of the container.

Why do we use lemmatization? : Terms spelling and material forms vary even within the same language or in a same text, because spelling is not normalized neither fixed yet. Words can be as well characterized by an important phonetic and morphologic variation: “eres, sodes, sos” are the verb to be (ser) 2nd pers. sing. If we add to this the possibility of person, mode, time variation, lemmatization becomes very important and uncovers a number of homological ambiguities with terms that are in no way related.

Aliento’s linguists solve graphic and morphological variation through lemmatization, which consists in the linking together a graphic form of the text (a word) and a “lemma”, that is to say, an unmarked form of the word, a sort of “neutral basis”, or a stem without any temporal, aspectual mark, and no distinction of voice, number, gender and syntactic agreement. This resolution leads to the constitution of an alphabetical dictionary of lemmas which will help find all the occurrences of a word in all the texts with all its morphological and orthographical variants.

For Aliento, the XML entity container represents a complex lexical encoding done in a well-thought and efficient way in order to simplify the computing of lexical categorization. It should, however, be mentioned that, since this categorization functions at the level of the word as an autonomous entity, it has a limited relevance at certain levels of the calculation.

<**str.ling**>: the root '<str.' stands for structure and the suffix '.ling' stands for linguistic. This entity uses a specific formal convention to describe the linguistic structure of the sapiential unit.

<**str.form**>: we indicate that we are dealing with a dialogue form (question/answer), an imperative sentence (exhortation), a parallel sentence structure, and so on.

Formal structures vary even if a certain number of them recur and are identifiable. This is the case of “si…” (“if…” or “whether…”); “como” with the meaning of “as” , “such as” or “like”; “quanto mas… tanto mas” (“the —er [the more]…, the—er…”);

<**str.poet**>: the root '<str.' stands for structure and the suffix '.poet' stands for poetic. This entity uses a specific formal convention to describe the poetic structure of the sapiential unit.

The development of the formal structure, the linguistic structure and of the poetic structure had first to do with the description of a formal and logico-semantic model, the type of model proposed by Fernando Bello in 1981 and quoted by Salah Mejri in his book intitled Le figement lexical (1997). But it is an idealized model, which can't be used as such in computational modeling. This is why we have decided to encode differently what belongs to the form and the meaning, what is intracultural and intercultural. This explains the use of tags indicating: ideological keywords <key._>, figurative meaning <sf._>, literal meaning <sl._>, the lesson <lec._>.

We also take into account diachronic and areal aspects in appended databases concerning the works from which we extract the brief sapiential units, the authors, translators, compilers that are to be found in the encyclopedic index cards coming from critical introductions, prosopographies. All this information is directly linked to the encoded texts.

3 Using the specific polysemy of brief sapiential units

In the description of the knowledge container, we mentioned the specific computational use of some XML entities, like the <lem.es> entity for example.

But we can define also the knowledge container as a computational container, where some XML entities are used to drive specific semantic calculations.

This dual conception of the knowledge representation is mentioned by Markman in his analysis of the requirements of a model of knowledge representation:

> "It makes no sense to talk about representations in the absence of processes... Only when there is also a process that use the representation does the system actually represent, and the capabilities of a system are defined only when there is both a representation and a process". [Markman]

Ultimately, the container can be thought as a dialogical resource shared as well by specialists than by any computing process [Gronoff]. This dialogic process allows to improve and validate XML knowledge representation in order to compute similarities between brief sapiential units characterized by several kinds of irregularities or between multilingual search for similarities [Wierzbicka].

Many sapiential units integrate in their linguistic representation phonetic elements to make the unit efficient to give an advice, to produce great-sounding effect when it is declaimed, to help memorisation, and /or to make puns. A great sapiential unit integrates the whole – good advice, musical and poetic declamation, replay and enjoyment – in a specific textual blending.

This kind of transcription can be analyzed as a polysemic system depending on the quality of the blending. A high quality blending is an efficient polysemic system where the borders of the different components of the system are melted and allow to slide between meaning, aesthetics and ergonomy. The quality of the polysemic system can be thought as a part of the success story of some sapiential units.

Such blended polysemic systems can be analyzed with the use of specific formal conventions working at the scale of the construal unit - in our case the brief sapiential unit -, instead of at the scale of the words, as it is the case with statistiscal methodologies or, for instance, in [Arbesman S.; Strogatz, S. H.; Vitevitch, M.], although the authors deal with "the presence of many smaller components (referred to as islands)..." to compare phonologically and semantically Spanish and English.

Literaturverzeichnis

[An00] Anscombre, J.-C., (ed.): Langages – La parole proverbiale, Nr. 139, 2000.

[ASV10] Arbesman, S.; Strogatz, S. H.; Vitevitch, M. S.: "Comparative Analysis Networks of Phonologically Similar Words in English and Spanish." *Entropy* 12, Nr. 3, 2001; pp. 327-337.

[Be84] Bello F.:Proposition d'une méthodologie d'analyse des logiques d'un corpus proverbial, *Richesse du proverbe. Typologie et fonctions*. Vol2. Etudes réunies par F. Suard et C. Buridant. Université de Lille III, 1984.

[BOG10] Bornes Varol, M.-C.; Ortola, M.-S.; Gronoff, J.-D., Aliento - The Mukhtar al- Hikam et ses sources. Journée d'étude Aliento (11 juin 2010). Inalco 201.

[Ch10] Chabrolle-Cerretini,,A.-M.: Réflexions théoriques sur les aspects linguistiques de la base Aliento- Europhras 2010 Cross-Linguistic and Cross-Cultural Perspectives on Phraseology and Paremiology (Granada, June 30th – July 2nd) 2010.

[CR04-09]Croft, W.; Cruse, D. A.: Cognitive Linguistics. Cambridge University Press, New-York, 2004-2009.

[Go96] Gouvard, J.-M.: Les formes proverbiales - Langue française, Linguistique et poétique: après Jakobson, Nr. 110, 1996.

[GL10] Gronoff, J.-D.; Lorquet, S.: Man / Machine cognitive interaction: constructing semantic precision, plasticity and dialogue with annotation. http://DualSemantics.Org (to be published in september 2010), 2010.

[Ma99] Markman, A. B.: Knowledge representation. Lawrence Erlbaum Associates, Mahwah N.J, 1999.

[Me97] MEJRI, S.: Le figement lexical : descriptions linguistiques et structuration sémantique. Série linguistique X, Publications de la Faculté des lettres de la Manouba, 1997.

[WI87] Wierzbicka, A.: Boys will be boys - Language 63, 1987; pp.:95-114.

Der *Thesaurus Linguae Aegyptiae* – die Verknüpfung eines elektronischen Corpus ägyptischer Texte und dem Wortschatz der ägyptischen Sprache

Dr. Ingelore Hafemann

Akademienvorhaben Altägyptisches Wörterbuch
Berlin-Brandenburgische Akademie der Wissenschaften
Jägerstrasse 22/23
10117 Berlin
hafemann@bbaw.de

Abstract: The *Thesaurus Linguae Aegyptiae* (TLA) is an electronic corpus of texts written in Ancient Egyptian language (from about 2500 B.C. to 300 A.D). The TLA is today with actually about 800.000 text words the largest corpus of Egyptian texts available online. It is a lexicon based corpus - every word of the corpus is linked to a corresponding lemma in the electronic lexicon that comprises meanwhile 26.450 lemma entries with lexicographical description. The original intention of the project was mainly to create a considerably large corpus. The more the corpus grew larger the more it became obvious that it would be necessary to improve the analytical tools both for the corpus and the lexicon.

Das lexikographische Projekt

Die Projektarbeit des Akademienvorhabens Altägyptisches Wörterbuch knüpft an das größte lexikographische Unternehmen in der Ägyptologie an - dem Wörterbuch der Ägyptischen Sprache, das von 1897 bis 1925 vorbereitet und schließlich in fünf Hauptbänden von 1926-31 publiziert wurde. Dieses Wörterbuch gründete sich auf 1.2 Millionen Belegzettel, die in der Regel den Volltext der Textquelle erfassen. Diese Zettelsammlung liefert sämtliche Belege eines Wortes in seinen Kontexten (Abbildung 1). Dem Prinzip nach ist das ein KWIC-Index. Seit 2002 steht diese Zettelsammlung komplett im Internet als indiziertes *Digitalisiertes Zettelarchiv* und wird von der internationalen Ägyptologie weltweit genutzt [Se2000].

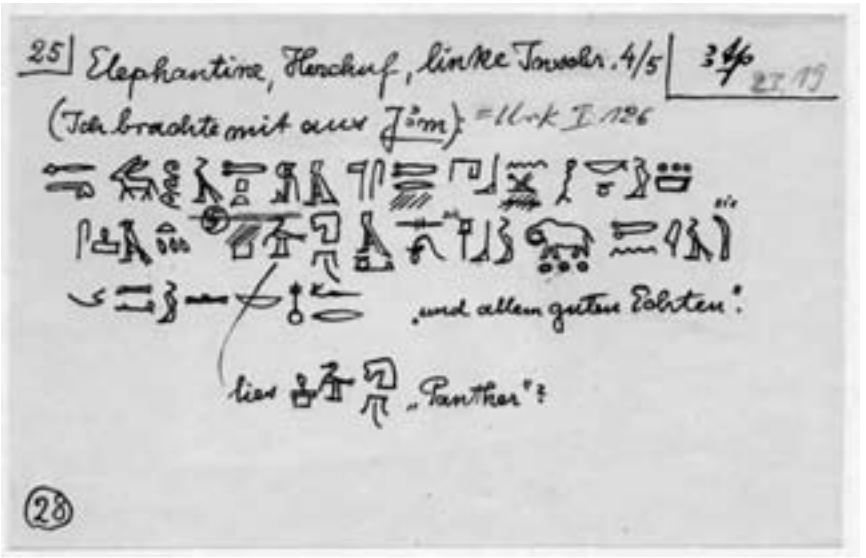


Abbildung 1: Textzettel mit *node word* in seinem Kontext - rot unterstrichen und in der rechten oberen Ecke für die lexikalische Einsortierung noch einmal in Transkription notiert

Jeder Zettel steht hinter einer Lemmakarte und eine lexikalische Feinsortierung gliedert die Zettel nach Gebrauchsweisen, d.h. nach phraseologischen und syntaktisch-semantischen Aspekten. Dieses methodische Grundprinzip - die Darstellung des Wortschatzes ganz und gar aus den Quellen zu erarbeiten - und die Entscheidung, die lexikographische Arbeit auf ein umfassendes und erschlossenes Corpus zu gründen, hat sich als fruchtbar, auch für das neue Projekt erwiesen.

Der *Thesaurus Linguae Aegyptiae*

Das Prinzip corpusbasierter Lexikographie ist für die Ägyptologie also nicht neu. In diesem Ansatz liegt das Spezifikum der Berliner Wörterbucharbeit bis heute. Das neue Projekt, das 1990 mit Vorüberlegungen und 1992 mit der realen Arbeit an der Berlin-Brandenburgischen Akademie der Wissenschaften startete, entschied sich für die Anlage eines Corpus ägyptischer Texte in einer relationalen Datenbank. Die Schwierigkeiten der Organisation großer Materialmengen sind jetzt beherrschbarer. Allerdings erfordert der Einsatz moderner Datenverarbeitung genaues Vorausdenken - Strukturierung und Kontrolle der Dateneingabe anhand von standardisierten Beschreibungsbegriffen waren erforderlich. Grundsätzlich sind zwei Datengruppen zu erfassen - die extralinguistischen Beschreibungsdaten und die Sprachdaten selbst (Abbildung 2):

1. Beschreibungsdaten der ägyptischen Texte und Wörter (Begriffsthesauri)
2.a. Texte als fortlaufende Wortsequenzen mit Satzsegmentierung
2.b. Lexikalischer Thesaurus als Lemmaliste, mit hierarchischer Binnenstruktur.

Eine Bilddatenbank soll Wörterbuch (Bildwörterbuch) und Textcorpus (Faksimiles, Photos der originalen Denkmäler) ergänzen.

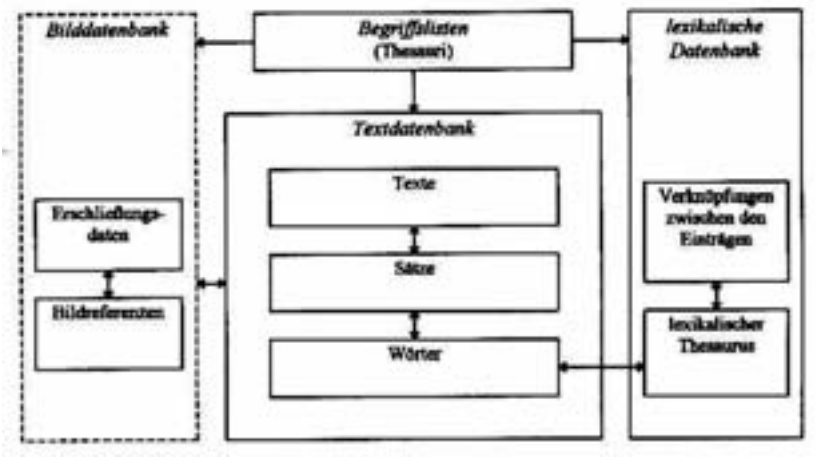


Abbildung 2: Strukturskizze der Datenbank des Altägyptischen Wörterbuches

Die Texterfassung war und ist das Kernstück der Arbeit. Mit Anwachsen der Datenbank wurden zunehmend Filter und Analysefunktionen notwendig, die parallel entwickelt und implementiert wurden. Im Jahr 2004 startete der Internetauftritt mit dem *Thesaurus Linguae Aegyptiae* als der neuen Publikationsplattform des Projektes [Se2004] und ist seitdem unter der Adresse http://aaew.bbaw.de/tla erreichbar.

Die Struktur des *Thesaurus Linguae Aegyptiae*

Die Texte

In der Texterfassung wird der ägyptische Text, der in Hieroglyphen oder einer Kursive, dem Hieratischen, niedergelegt ist, in einer Umschrift notiert. Dabei werden die Wörter des Textes in Folge erfasst und durch *blanks* voneinander getrennt. Die Wortsequenzen werden segmentiert in Sätze oder satzähnliche Strukturen, die semantisch-syntaktische Sinneinheiten widerspiegeln. Jedem Satz / Sinneinheit wird in einem separaten Feld eine Übersetzung in Deutsch, Englisch oder Französisch zugefügt. Außerdem erfolgt eine grammatische Annotation der Wortform (Flexionsform).

In einem halbautomatischen Lemmatisationsvorgang wird jedes Textwort mit einer elektronischen Lemmaliste verknüpft.

Das Lexikon

Das Wörterbuch steht in Gestalt einer elektronischen Lemmaliste zur Verfügung. In ihr ist der gesamte bekannte Wortschatz des Ägyptischen erfasst. Die Lemmaliste bietet je Eintrag die ägyptologische Transkription, die hieroglyphische Standartschreibung, eine oder mehrere Übersetzungen in Deutsch und Englisch, bibliographische Angaben (Wörterbücher, Glossare, Einzeltextzitate) und ist mit einigen lexikographischen *labels* versehen wie einer Wortartangabe und teilweise einer Klassifizierung von Eigennamen als Personennamen, Götternamen, Titel und Epitheta, Ortsnamen und Namen von Institutionen. Diese Wortliste hat auch eine hierarchische Binnenstruktur. Die Hierarchie bildet semantische Lesarten oder sogar teilweise phraseologische Wortverbindungen ab (Abbildung 3)



Abbildung 3: Lemmaliste des Altägyptischen Wörterbuches mit Binnengliederung

Die vielen syntagmatischen und paradigmatischen Wortverbindungen sind nie in einem Wörterbuch abzubilden und jeder Lexikograph muss hier selektieren. Daher ist die Verknüpfung mit dem gesamten Textcorpus von unschätzbarem Wert. So lässt sich der vielfältige Gebrauch von Wörtern, verteilt über Epoche oder Textsorten abfragen.

Die Recherche-*tools* des *Thesaurus Linguae Aegyptiae*

Da mit nunmehr fast 800.000 laufenden Textwörtern das größte elektronische Corpus ägyptischer Texte geschaffen wurde, das weiter anwächst, sind erstmals auch Frequenzdaten für Wortvorkommen verfügbar. Das Corpus ist zur Zeit für statistische Analysen im Ganzen noch nicht geeignet, da es kein ausgewogenes Corpus darstellt. Für viele Texte und Textgruppen aber sind eine Reihe von Abfragen und Analysen möglich, die über die lexikalische und thematische Struktur dieser Texte informieren.

Folgenden Analysen sind aufgrund der relationalen Datenbankstruktur möglich:

- Jedes Wort des Wortschatzes kann in der Lemmaliste nachgeschlagen werden. Die Suche erfolgt über die Transkriptionsmorpheme, auch unter Verwendung regulärer Ausdrücke sowie alternativ über die hieroglyphischen Schreibungen. Die Selektion nach Wortarten oder Spezialwörtern (Eigennamen) ist möglich.
- Für jedes Lemma der Wortliste können die Belege des Corpus aufgelistet werden. Die Ausgabe kann satzweise oder als KWIC-Datei, sortiert nach rechtem oder linkem Kotextwort, ausgegeben werden.
- Für jeden Textbeleg ist ein direkter Zugang zur gesamten Wortfolge des jeweiligen Textes möglich. Jedes Textwort dieses Textes ist erneut durch seine Verknüpfung mit der Lemmaliste analysier- und recherchierbar.
- Es können Wortindizes für alle Texte oder Corpussegmente erstellt werden und der Spezialwortschatz kann aufgelistet werden.
- Es kann nach dem gemeinsamen Auftreten zweier konkreter Wörter gesucht werden. Die Suche nach dem kombinierten Auftreten von Wörtern kann auch auf Wortarten bezogen werden.

Darüber hinaus sind eine Reihe statistischer Analyse-*tools* im TLA implementiert:

- Kollokationsanalysen zu einem Lemma
- Analyse der *type/token*-Ratio und der Zusammensetzung nach Wortarten
- Analyse der Verteilung der Häufigkeiten der Wörter, einschließlich ihrer Verteilung auf Wortarten
- Schlüsselwortanalyse (*key-words*) für einzelne Texte oder Corpussegmente
- Analyse der lexikalischen Gravitation eines Lemmas

Der *Thesaurus Linguae Aegyptiae* als virtuelles Wörterbuch

Der lexikonbasierte Ansatz setzt ein gutes Lexikon voraus. Viele Wörter sind polysem, wie Verben, die - abhängig von beteiligten Präpositionen und Substantiven - den propositionalen Gehalt des Satzes bestimmen und verändern können. Hier unterscheiden Lexika oft mehrere Lesarten eines Lexems. Die Lesartendifferenzierung stellt immer wieder ein Problem dar. Lexika bieten hier oft sehr verschiedene Ansätze für dasselbe Lexem. Bei der Ansammlung von Belegen auf die ägyptischen Lemmata während der Texterfassung nutzen wir in vielen Fällen eine lexikalische Differenzierung, die in Form eines Untereintrages zum Hauptlemma zur Verfügung gestellt wird. Beim Ansetzen solcher Lesarten spielen gewöhnlich Frequenzdaten eine Rolle. Oft können wir uns auf Vorarbeiten der Bearbeiter des alten Wörterbuches (Zettelarchiv mit Zettelzahlen) beziehen. Viele solcher Untereinträge ergeben sich aber auch erst im Laufe der Arbeit am Corpus durch wachsende Belege. Um Bedeutungswandel zu erfassen, sind solche Fragen von großer Bedeutung. Da sowohl das alte Wörterbuch der Ägyptischen Sprache (1926-1931) als auch der *Thesaurus Linguae Aegyptiae* das Ägyptische in allen seinen Sprachstufen abbilden ist auch die elektronische Lemmaliste diachron angelegt. Bedeutungs- und Sprachwandel zu erforschen ist eine zentrale Aufgabe. Wir hoffen, mit einem dynamischen Corpusansatz und der modernen Texterschließung neue Wege für die ägyptologische und die allgemeine Sprachforschung zu öffnen. Das Ägyptische bietet sich hier als die am längsten überlieferte Schriftsprache der Welt in besonderer Weise an.

Literaturverzeichnis

[Se00] Seidlmayer, S. u.a.: Das Zettelarchiv des Wörterbuches der Ägyptischen Sprache. Aufbau, Digitalisierung, Erschliessung und Konsultation im Internet, Thesaurus Linguae Aegyptiae 1, Achet Verlag, Berlin 2000.

[Se04] Seidlmayer , S.: Der Thesaurus Linguae Aeyptiae im Internet. In Göttinger Miszellen 203, Göttingen 2004, S. 99-104.

Salton und Wittgenstein in den Humanities: Über die Semantik in Philosophischen Texten

Marco Büchler, Gerhard Heyer
Natural Language Processing Group
Institute for Mathematics and Computer Science
Leipzig University, Germany
$[mbuechler|gheyer]@eaqua.net$

Abstract: In der Informatik wird die Semantik durch *diskriminierende Terme* beschrieben. Jedoch fehlen oftmals speziell in philosophischen Texten genau diese gewichtigen Terme. Ausgehend von der oft eingesetzten *diskriminierenden Semantik* wird am Problem der Sinn- und Weisheitssprüche eine *kontrastive Semantik* vorgestellt. Die eingeführte Methode stellt ein Lessons Learnt aus dem eAQUA-Projekt [BHG08, HBB+10] im Umgang mit antiken Texten dar.

1 Einführung

Dem Thema Semantik kann sich im Bereich der *Automatischen Sprachverarbeitung* auf verschiedenste Weise genähert werden. Aus der Sicht des *Information Retrieval* werden Suchmaschinen darauf optimiert, mittels möglichst weniger Eingabewörter ein relevantes Dokument zu finden (Semantik einer Textpassage). Dazu werden die Terme einer Textpassage gewichtet und repräsentieren somit den semantischen Raum. Auf der anderen Seite werden bspw. im *Text Mining* signifikante Assoziationen zwischen Wörtern berechnet (Semantik eines Wortes), die wiederum deren semantische Nutzung in einem Textkorpus wiedergeben.

Wird das Thema Semantik aus der Sicht des *Text Reuse* und *Knowledge Transfer* betrachtet, spielen beide Aspekte - Semantik einer Textpassage und Semantik eines Wortes - eine entscheidende Rolle. Im Kontext der eHumanities muss jedoch zwischen den *historischen* und *philosophischen Zitationsspuren* unterschieden werden. Während beim historischen Wissenstransfer oftmals eindeutig diskriminierende Terme wie *Orte*, *Personen* oder *Ereignisse* bestimmt werden können, ist das Vokabular der Philosophie sehr stark von Allgemeinsprache geprägt [Pie10], welches den Einsatz von semantischen *Reuse*-Verfahren deutlich erschwert. Dies kann am folgenden Spruch von William Shakespeare verdeutlicht werden.

To be, or not to be this is the question.
William Shakespeare in Hamlet

Auch wenn dieser Spruch von Shakespeare von vielen Menschen wiederverwendet wird,

ist es mit *Information-Retrieval*-Methoden sehr schwer eine semantische Repräsentation zu bestimmen, da er nahezu komplett aus Stoppwörtern besteht und dementsprechend keine oder nur schwach diskriminierende Terme enthält. Werden beispielsweise die Termgewichte nach dem $tf.idf$-Maß [SWY75] für diesen Spruch ausgerechnet, dann haben bis auf *question* alle Terme ein Gewicht von 0 (Stoppwörter). Des Weiteren liefern auch Verfahren wie die *Differenzanalyse* oder das *Log-Likelihood-Ratio* (beide siehe Abschnitt 2) im Vergleich zu einem Referenzkorpus keine ernsthafte semantische Repräsentation.

Ferner ist das Sprachvokabular oftmals so allgemein, dass sich philosophische Texte nur schwer vom sprachlichen Niveau eines Grundschülers unterscheiden. So entspricht das oben vorgestellte *Shakespeare*-Zitat nach dem *Dale Chall Readability Index* gerade einmal dem Sprachniveau eines Schülers der Klassenstufe 2 bis 3.

Da Methoden der *Semantik durch diskriminierende Terme* philosophische Texte nur sehr schwer beschreiben können, wird in diesem Papier ein einfaches Verfahren vorgestellt, welches *Semantik durch Kontrast* misst. Hierbei wird eine Textstelle nicht durch diskriminierende Terme beschrieben, sondern durch einen möglichst großen Kontrast mindestens zweier Wörter innerhalb dieser Textstelle.

Gerade in philosophischen Texten ist dies von größtem Interesse, da in ihnen oftmals Lebensweisheiten bzw. gesellschaftliche und soziale Wertungen enthalten sind, die aktiv von Wissenschaftlern aus den Geisteswissenschaften gesammelt wurden und immer noch werden. Ein Spezialfall dieser philosophischen Texte stellen die punktierten *Sinn- und Weisheitssprüche* (Gnomologien) dar [Pie10, VOG10, Rou10]. Speziell hierbei werden Konzepte in Relation zueinander gestellt, die sowohl nicht erwartet als auch oftmals semantisch kontrastive Terme enthalten. Vielmehr bilden sie Wissensmuster ab, die alltägliche Konzepte in nicht-alltäglichen Zusammenhängen miteinander verbinden, die wiederum durch einen Lerneffekt gewonnen worden sind.

2 State Of The Art

Aus der Sicht des semantischen *Text Reuse* und *Knowledge Transfers* gibt es zwei Sichten auf das Modellieren von Semantik:

- *Semantische Repräsentation von Textpassage*: Hierbei wird eine Textpassage durch ihre diskriminierenden Terme beschrieben und dementsprechend repräsentiert bzw. gewichtet. Während Salton's $tf.idf$ [SWY75] auf einem einzelnen Korpus angewendet wird, können Methoden wie die *Differenzanalyse* [Wit04] bzw. eine entsprechende Modifikation des *Log-Likelihood-Ratios* [Wit04, Dun93] eingesetzt werden, um entsprechende Abweichungen bzgl. eines Referenzkorpus zu messen. Neben den *probabilistischen* gibt es auch *linguistische* und *vorwissensbasierte Verfahren*, auf die an dieser Stelle aber nicht im Detail eingegangen werden soll.
- *Semantischer Kontext eines Wortes*: Unabhängig von den einzelnen Textpassagen kann die semantische Umgebung eines Wortes bestimmt werden, um dessen Bedeutung innerhalb eines Korpus zu berechnen [Wit53]. Als gängige Methode hat

sich diesbezüglich die *Kookkurrenzanalyse* bewährt [HQW08, Büc08, Büc05]. Dabei wird die Assoziationsstärke zwischen zwei Wörtern gemessen. Die Menge aller assoziationsstarken Terme zu einem Wort repräsentieren dessen semantische Bedeutung innerhalb eines Korpus, die wiederum dazu genutzt werden kann, um ähnlich benutzte Wörter zu bestimmen [Bor07].

Unabhängig von der semantische Repräsentation einer Textpassage oder eines Wortes wurden *Readability*-Tests eingeführt, um Texte nach ihrem sprachlichen Niveau zu bewerten. So werden in den USA Scores wie der *Dale Chall Readability Index* [Cen10b], *Coleman Liau Readability Index* [Cen10a] und der *Automated Readability Index* [SS67] eingesetzt, um Textdaten altersgerecht und dem sprachlichen Niveau entsprechend einer Altersstufe zuzuordnen [Cen10b]. Methodisch messen solche Verfahren letztendlich immer zwei Merkmale: Einerseits spielt die *Satzlänge* eine wichtige Rolle. Andererseits die Menge *inhaltsbehafteter Wörter*. Je nach Maß wird dies über die Wortlänge, die Anzahl der Silben oder die Anzahl der Wörter, die nicht zu den 3000 häufigsten Wörtern zählen, gemessen. All diese *Readability*-Maße haben gemeinsam, dass sie den Score erhöhen, sobald sich diskriminierende Terme häufen und die Sätze länger werden.

3 Methodologie

Im eingangs erwähnten Spruch von William Shakespeare sind weder die im Abschnitt 2 genannten diskriminierenden Terme enthalten, noch scheint der Spruch inhaltlich schwierig zu sein. So kann nach dem *Dale Chall Readability Index* [Cen10b] für diesen Spruch ein Score von $DCI = 4.1821$ berechnet werden. Dies entspricht nach internationalen Standards und der in den USA aktiv eingesetzten Klassifikation dem Sprachniveau eines Grundschülers der 2. - 3. Klasse. Da jedoch dieser Spruch allgemein bekannt und oft zitiert ist, stellt sich die Frage, ob semantische Relevanz immer durch diskriminierende Terme gemessen werden kann.

Speziell in philosophischen Texten können nur selten diskriminierende Terme aus dem *Iota*-Bereich der Wortverteilung (seltene Wörtern) genutzt werden. Vielmehr werden tendenziell häufige und allgemeine Terme der *Delta*- und *Zeta*-Bereiche (Details zu *Delta*-, *Zeta*- und *Iota*-Wörtern in [RE10]) beobachtet. *Burrow's Delta* [Bur07] wurde als Methode im Bereich des *Authorship Attribution* eingeführt. Hierbei werden speziell die Stoppwörter des *Delta*-Bereichs genutzt, um nach stilistischen Unterschieden zu suchen [Arg08]. Da der Schwerpunkt nicht auf stilistischen Merkmalen von Termen des *Delta*-Bereiches, sondern auf Semantik speziell des *Delta*- und *Iota*-Bereiches liegt, wird nachfolgend eine Adaption von *Burrow's Delta* genutzt, um nach möglichst großem semantischen Kontrast zweier Terme innerhalb einer Textstelle mit möglichst geringem Abstand im Text zu suchen.

Hierzu werden in einem ersten Schritt die semantischen Kookkurrenzen K zu jedem Wort bestimmt [Büc08, Büc05]. Als Signifikanz-Maß wurde das Log-Likelihood-Ratio sim_{lgl} mit einem Schwellwert von 6.63 und einer Mindestkookkurrenzfrequenz von 2 gewählt.

Basierend auf dem berechneten Kookkurrenzgraphen K werden anschließend die paar-

weisen Ähnlichkeiten zweier Wörter w_i und w_j mit dem Dice-Koeffizient berechnet.

$$sim_{dice}(w_i, w_j) = 2 * \frac{|K_{w_i} \cap K_{w_j}|}{|K_{w_i}| + |K_{w_j}|} \tag{1}$$

Hierbei entsprechen K_{w_i} und K_{w_j} den Kookkurrenzen der Wörter w_i und w_j. In diesem Schritt könnte auch ein Ähnlichkeitsmaß wie das *Cosinus*-Measure benutzt werden. Da dafür jedoch entsprechende Termgewichte nötig sind, ist im Rahmen dieses Papiers aus mehrfach genannten Gründen verzichtet worden.

Im Gegensatz zum Bestimmen von Wörtern mit ähnlichen Kontexten [Bor07] werden im zweiten Schritt genau die Wörter mit sehr ähnlichen Kontexten entfernt, um Kandidaten für Assoziationen zu bestimmen, die einen semantischen Kontrast repräsentieren.

$$contrast(w_i, w_j) = \begin{cases} 1 - sim_{dice}(w_i, w_j) & \text{if } sim_{dice}(w_i, w_j) \leq eps \\ 0 & \text{if } sim_{dice}(w_i, w_j) > eps \end{cases} \tag{2}$$

In der konkreten Anwendung hat sich im Altgriechischen ein $eps = [0.1, 0.15]$ als praktikabel herausgestellt.

Da die Menge der Assoziationen kontrastiver, unähnlicher bzw. unerwarteter Kookkurrenzprofile $C_{cand} = \cup_{i,j \in V} contrast(w_i, w_j)$ mit dem Vokabular V nicht zwangsweise auch im Text zusammen vorkommen, werden in einem dritten Schritt aus der Menge C_{Cand} diejenigen *kontrastiven Assoziationen* selektiert, die innerhalb eines Textfensters (hier Satz) auch zusammen auftreten. Dies entspricht dem Durchschnitt $C = K \cap C_{Cand}$ der beiden Menge K und C_{Cand} mit der zusätzlichen Bedingung.

$$dist(w_i, w_j) \leq eps_{dist} \text{ aus } (w_i, w_j) \in C \tag{3}$$

Hierbei entspricht die Restriktion $dist(w_i, w_j)$ dem Abstand der beiden Wörter im Text.

Visuell kann sich diese Methode wie folgt vorgestellt werden: Es sei angenommen, dass sich die Semantik eines Wortes durch Farben ausdrücken lässt. Dann bestimmt der Algorithmus in den beiden ersten Schritten einen möglichst großen farblichen Unterschied (Kontrast), der im dritten Schritt auf einen kleinen Raum beschränkt wird, um bspw. die Kante eines Objektes bzw. Gegenstandes auf einem Bild zu erkennen.

4 Ergebnisse, Lessons Learnt und Scope

Das im Rahmen dieses Papiers vorgestellte Verfahren misst semantischen Kontrast. In Anlehnung an das Beispiel aus dem Bereich des Image Mining gibt es mehrere Ergebnis-Cluster, die im Wesentlichen von der Textsorte abhängig sind wie zum Beispiel

- *Philosohie*: beispielsweise Gnomologien,
- *Komödie*: Sarkasmus und Zynismus,

- *Historie/Geschichtsschreibung*: unerwartete historische Zusammenhänge,
- *Sentiment Analysis*: Künstliche Doppeldeutigkeit, wobei nur die Doppeldeutung und nicht deren positive oder negative Wertung erkannt werden.

Die vorgestellte Methode ist genau genommen in vielerlei Hinsicht gegenteilig zu existierenden Ansätze. Auf der einen Seite entspricht der Algorithmus einer neuen Klasse von Verfahren. In der Automatischen Sprachverarbeitung werden Sprachmodelle benutzt, um basierend auf Trainingsdaten Vorhersagen zu machen, was als wahrscheinlichste Assoziation gilt. Sei es auf der syntaktischen Ebene bei den Markov-Ketten oder auch bei den Kookkurrenzen auf dem semantischen Level. Herkömmliche Sprachmodelle messen immer das Offensichtliche bzw. Wissen, das als gesichert angesehen werden kann. In den Humanities jedoch, ist dieses Wissen bekannt und kann nach jahrhundertelanger Forschung als gegeben angesehen werden. In einer solchen geisteswissenschaftlichen Anwendungen werden *latente Sprachmodelle* benötigt, um einen Mehrwert zu generieren. Andererseits werden in der Informatik Graph-Partitionierungsalgorithmen angewandt, um semantische Cluster zu bilden. Der in Abschnitt 3 vorgestellte Algorithmus bewertet jedoch genau diejenigen Kanten, die ein solcher Partitionierungsalgorithmus entfernt.

In Anlehnung an die zugrunde liegenden philosophischen Texte kann eine signifikante Überlappung zwischen den *kontrastiven Relationen* aus diesem Papier und dem *Text Reuse* bzw. *Knowledge Transfer* ausgemacht werden. In über 90% einer kontrastiven Relation wird auch ein *Text Reuse* gemessen. Diese Beobachtung ist insofern interessant, als dass dadurch erstmals nicht die Frage nach dem *Wie wird Text Reuse gemessen* im Mittelpunkt steht, sondern *Warum wird Text wiederverwendet*. Des Weiteren kann beobachtet werden, dass aufgrund ihrer philosophischen Reife (philosophisch gut überlegten Formulierungen) solche Weisheitssprüche sehr stark am Original wiederverwendet werden. Das Kernproblem der syntaktischen Verfahren ist jedoch, dass nicht zwischen einem statistisch signifikanten und häufig benutzten N-Gramm wie *im Namen unseres Herren Jesus Christus* und einem Zitat unterschieden werden kann. Mittels der *kontrastiven Relationen* kann zwischen allgemeinen Phrasen und potentiellen Zitaten eine Unterscheidung gemacht werden.

Im konkreten Beispiel sei auf das *Korpus der arabischen und syrischen Gnomologien* [Pie10] verwiesen. Aus den deutschen Übersetzungen der Weisheitssprüche ist die Abbildung 1 für das Wort *Körper* visualisiert worden. Es gibt insgesamt 8 verschiedene semantische Cluster dieses Wortes. In 5 dieser Cluster, die jeweils für einen Weisheitsspruch stehen, kann ein offensichtlicher Kontrast durch paarweise Antonyme wie z.B. *lebend* und *tot*, *besitzen* und *verteilen* oder auch *Stärke* und *Schwäche* erkannt werden.

5 Further Work

Für die Informatik hat sich zwischen den *kontrastiven Relationen* und dem *Text Reuse* ein unerwarteter Zusammenhang ergeben. Im Rahmen der weiteren Arbeiten im Bereich des *Text Reuse* und *Knowledge Transfers* wird die vorgestellte Methode verbessert. So kann

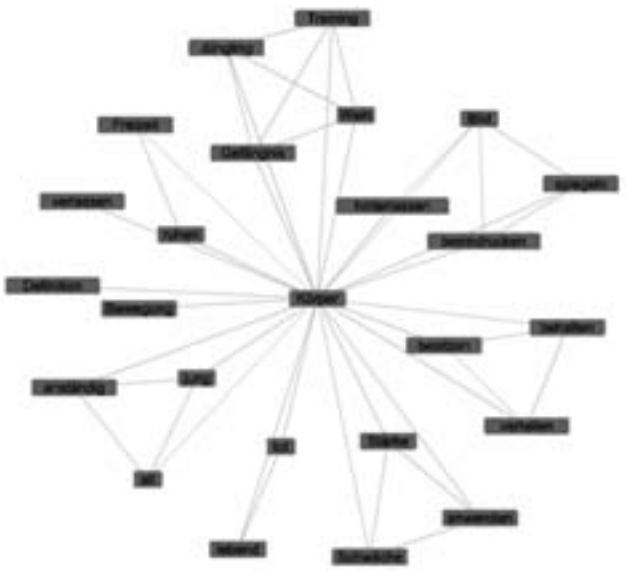


Abbildung 1: Die semantischen Cluster des Wortes *Körper* in den deutschen Übersetzungen aus dem *Korpus der arabischen und syrischen Gnomologien* [Pie10]. Jedes Cluster entspricht einem Weisheitsspruch. 5 der 8 Cluster beinhalten kontrastive Relationen wie *jung* und *alt*.

auf geshuffelten Texten gezeigt werden, dass das vorgestellte Verfahren durch einen Verzicht auf jeglichen probabilistischen Ansatz Schwächen hat. So kann der obere Grenzwert auch zufällig wie auf geshuffelten Texten sein. Daher wird bereits an einem Verfahren gearbeitet, welches einerseits die Kantengewichte nach dem *Log-Likelihood-Ratio* sowie die Topologie und der damit verbundene Dichte eines Wortes berücksichtigt. Hierbei werden diejenigen kontrastiven Assoziationen bevorzugt, die aus zwei sehr stabilen und gesicherten sowie möglichst unterschiedlichen Kontexten kommen.

Bei den *Readability*-Tests soll ein Text anhand der nötigen kognitiven Leistungsfähigkeit eines Menschen klassifiziert werden. Hierbei gibt es im Wesentlichen die Parameter der Satzlänge und die Menge an längeren Wörtern. Basierend auf diesen Kennzahlen werden die Texte klassifiziert. Jedoch muss angenommen werden, dass im Kontext eines Sprachmodells immer erwartbare bzw. leicht verständliche semantischen Dependenzen im Text vorkommen. So wird das eingangs erwähnte Zitat von Shakespeare auf das sprachliche Niveau eines Grundschülers eingestuft, der dieses Zitat sicher auch lesen aber jedoch wohl eher nicht verstehen kann. In diesem Sinne wird außerhalb der Arbeiten zum *Text Reuse* diese Methode zu einem *philosophischen* bzw. *semantischen Readability-Test* weiterentwickelt. Als konkrete Anwendung hierzu stehen aktive Forschungsarbeiten in eAQUA an, die sich mit der Frage nach dem *warum sind bestimmte Werke wichtig* beschäftigen und damit heutzutage noch erhalten, während andere nur noch in Fragmenten vorliegen.

6 Zusammenfassung

In diesem Papier wird eine neue Methode im Umgang mit Semantik beschrieben. Während Semantik bisher immer mit *diskriminierende Semantik* durch stark inhaltsbezogene Features beschrieben wird, ist ein Verfahren vorgestellt worden, das *kontrastive Semantik* durch semantische Differenz formuliert. Hierbei liegt der Arbeitsschwerpunkt auf philosophischen Texten mit der Spezialisierung auf den Sinn- und Weisheitssprüchen, die durch *diskriminierende Semantik* aufgrund eines eher allgemein bekannten Vokabulars eher mäßig beschrieben werden können. Ferner wurde skizziert, dass die kontrastiven Relationen gute Indikatoren für den Arbeitsbereich des *Text Reuse* sind, da sie nicht das *Wie*, sondern das *Warum* messen. Hierbei wird der Mehrwert durch semantischen Kontrast gemessen.

Literatur

[Arg08] Shlomo Argamon. Interpreting Burrows's Delta: Geometric and Probabilistic Foundations. *Lit Linguist Computing*, Seite fqn003, 2008.

[BHG08] M. Büchler, G. Heyer und S. Gründer. Bringing Modern Text Mining Approaches to Two Thousand Years Old Ancient Texts. In *e-Humanities – an emerging discipline: Workshop in the 4th IEEE International Conference on e-Science*, 2008.

[Bor07] S. Bordag. *Elements of Knowledge-free and Unsupervised Lexical Acquisition*. Dissertation, Universität Leipzig, 2007.

[Büc05] M. Büchler. Medusa Release Homepage, 2005. URL: http://mbuechler.eaqua.net/medusa/ last accessed Feb. 14th, 2010.

[Büc08] M. Büchler. *Medusa: Performante Textstatistiken auf großen Textmengen: Kookkurrenzanalyse in Theorie und Anwendung*. Vdm Verlag Dr. Müller, 2008.

[Bur07] John Burrows. All the Way Through: Testing for Authorship in Different Frequency Strata. *Lit Linguist Computing*, 22(1):27–47, 2007.

[Cen10a] RFP Evaluation Centers. Coleman-Liau Grade Level Readability Score, reading scores, 2010. URL: http://rfptemplates.technologyevaluation.com/readability-scores/coleman-liau-readability-score.html last accessed Jul. 21th, 2010.

[Cen10b] RFP Evaluation Centers. Dale-Chall 3000 Simple Word List, Readability Grade Score, 2010. URL: http://rfptemplates.technologyevaluation.com/dale-chall-list-of-3000-simple-words.html last accessed Jul. 21th, 2010.

[Dun93] T.E. Dunning. Accurate Methods for the Statistics of Surprise and Coincidence. *Computational Linguistics*, 19(1):61–74, 1993.

[HBB+10] G. Heyer, M. Büchler, V. Boehlke, C. Utschig-Utschig und C. Schubert. Aspects of an Infrastructure for eHumanities. In *Journal of Computing and Cultural Heritage, http://jocch.acm.org/)*, 2010.

[HQW08] Gerhard Heyer, Uwe Quasthoff und Thomas Wittig. *Text Mining: Wissensrohstoff Text – Konzepte, Algorithmen, Ergebnisse*. W3L-Verlag, 2008.

[Pie10] Ute Pietruschka. Corpus der arabischen und syrischen Gnomologien, 2010.

[RE10] Jan Rybicki und Maciej Eder. Deeper Delta Across Genres and Languages: Do We Really Need the Most Frequent Words? *Digital Humanities 2010 - Conference Abstracts. Centre for Computing in the Humanities, King's College London*, 2010. URL: http://dh2010.cch.kcl.ac.uk/academic-programme/abstracts/papers/pdf/book-final.pdf last accessed Jul. 21th, 2010.

[Rou10] Charlotte Roueché. Sharing Ancient Wisdoms, 2010.

[SS67] E. A. Smith und R. J. Senter. *Automated Readability Index (ARI)*. Wright-Patterson AFB, OH: Aerospace Medical Division. AMRL-TR, 66–22, 1967.

[SWY75] G. Salton, A. Wong und C.S. Yang. A Vector Space Model for Automatic Indexing. *Communications of the ACM*, 18(11):613–620, 1975.

[VOG10] Marie-Christine Bornes Varol, Marie-Sol Ortola und Jean-Daniel Gronoff. Aliento project - Intercultural Analysis of Sapiential statements and Transmission, 2010. URL: http://www.aliento.eu/en/node/63 last accessed Jul. 21th, 2010.

[Wit53] Ludwig Wittgenstein. *Philosophical Investigations*. Blackwell, Oxford, 1953. Translated by G.E.M. Anscombe.

[Wit04] F. Witschel. Text, Wörter, Morpheme - Möglichkeiten einer Terminologie-Extraktion. Diplomarbeit, Universität Leipzig, 2004.

Semantic Ageing of Complex Documents: A Case Study from Built Heritage Preservation

Christoph Schlieder and Peter Wullinger

Lehrstuhl für Angewandte Informatik in den
Kultur-, Geschichts- und Geowissenschaften
Otto-Friedrich-Universität Bamberg
christoph.schlieder@uni-bamberg.de

Abstract: At least two types of ageing processes affect digital records: media ageing and semantic ageing. We argue that the main challenges for digital preservation arise from semantic ageing, that is, the evolution of data formats. The article analyzes a type of document that is particularly vulnerable to semantic ageing: the digital maps produced by researchers on built heritage which combine spatial and thematic data. We describe a solution for format migration based on the ontological modeling of the thematic data.

1 Introduction

The preservation of digital records has been a concern of memory institutions such as libraries or museums ever since they admitted born-digital content into their collections [Kun97]. Around the year 2000, several research initiatives on digital long-term preservation started in the United States, in Europe, and in other parts of the world [Bor06], [Sol10]. Owing to these research efforts, a number of organizational and technological solutions have emerged for the preservation of digital texts and images. However, for more complex types of documents, the situation remains unsatisfactory. From a computational perspective, the most complex documents are not produced by the text-oriented disciplines which form the core of the "digital humanities" but by disciplines such as archeology or historic geography for which material records of cultural processes constitute a central data source. The digital maps which researchers on built heritage produce in order to document historic buildings provide an example of such complexity.

We argue that the computational challenge for long-term preservation of these documents is caused by semantic variability. This means that the semantics of the data is shared by only a small group of users and, additionally, it shows a high tendency to evolve. Any approach for the long-term preservation of complex documents has to somehow address this issue of semantic variability. The main contributions of this article are the following: We analyze the semantic ageing processes that affect the digital maps produced in the context of built heritage conservation. (section 2). Furthermore, we describe a solution for format migration based on ontological modeling (section 3). The article concludes with a comment about a recent discussion on semantic ageing.

2 Semantic Ageing Processes of Digital Records

Any concrete instance of a digital document is bound to a physical medium. The term *media ageing* denotes the process of physical deterioration that affects the medium until at some point the document's original bitstream cannot be recovered anymore. Although private computer users often experience media ageing, e.g. disk failures, as the unique source of their digital preservation problems, this is certainly not true for memory institutions. The standard approach for addressing media ageing consists in detaching the bitstream from the medium and in copying it onto another medium – a task that is easily handled by an adequate preservation planning scheme [Bor06].

A much more serious challenge arises from *semantic ageing*, that is, the evolution of data formats. Knowledge about the semantics of the data is quickly lost if it is not explicitly specified and maintained. The effects of semantic ageing become visible often too late, when the curators of a digital collection realize that no application software on the current platform is able to access a document via an import filter. Unfortunately, standardized data formats do not provide a long-term solution to semantic ageing [Sch10]. Over a period of 50 years even character formats evolve beyond any standard (ASCII, ISO 8559, Unicode). The solutions for the semantic ageing problem that the digital preservation initiatives found to be most reliable are based on two types of approaches: emulation or migration – or sometimes a combination of both.

Emulation recreates the runtime environment of the application software which generated the documents in the first place. Since it also permits to re-enact a user experience from the past, emulation is usually chosen for collections of interactive media such as video games or when the original look-and-feel of document manipulation is of particular importance. Archivists at Emory University, Atlanta, for instance, emulated the 20-year old software environment used by the writer Salman Rushdie to give literary scholars insight into the working conditions during the time when the writer was forced to live underground [Coh10].

While emulation guarantees a maximum of authenticity, it does not help to integrate past contents into the knowledge-based workflows of the present. Being able to run decade-old application software does not make it interoperable with today's technologies. Workflow integration requires more, namely a translation process that converts outdated data formats into currently supported formats, in other words, *migration*.

Documentation in built heritage preservation is based to a large extent on architectural drawings or maps that are generated by preservation scientists while inspecting the building [Sch02]. Digital workflows are supported by documentation systems running on a mobile computer, e.g. a TabletPC. An example for such a system is the Mobile Mapping System in its archiving edition (MMSarchive) which has been developed at the University of Bamberg and is currently used at several Middle European cathedrals and other monumental buildings ([Mat05], [Wul08], [Fre09]). The resulting digital documents include *inventory maps* describing the different parts of the building as determined by more or less extensive measurement procedures (tachymetry, laser scan) and *damage maps* which report the damages of the built structure.

A digital map associates spatial and thematic data as illustrated in Figure 1. The map consists of a collection of spatial objects, for instance, polygons representing individual ashlars, that is, the dressed stone blocks which make up the masonry. Each spatial object is associated with thematic data. Figure 1 shows an input dialog that permits to enter the preservation state, the construction phase, the surface variety, and the stone type.



Figure 1: Digital map with recorded facade damages, St. Stephan, Passau

It is important to note that the thematic data differs considerably between preservation sites. Semantic heterogeneities arise not just by the fact that different sets of properties are recorded but also from an incompatible logical structuring. For instance, while the data model for the Passau cathedral has the stone type encoded in the value of an attribute, the data model for the Vienna cathedral uses explicit mapping objects for each particular stone type. Semantic ageing occurs because the requirements of the documentation task evolve. New types of objects, properties and relations are introduced others become obsolete. Sometimes, the modeling is just restructured to reflect not new concepts but a different understanding of their logical organization. In this process of format evolution, however, the data does not become invalid. There is a need to reuse digital maps in scientific workflows decades after they have been generated.

Digital preservation has studied semantic ageing mainly for single media documents although approaches for preservation planning for mixed media documents have been proposed [Hun06]. The digital maps of historic buildings are more complex than the typical mixed media document which is an aggregate of different medias (e.g. an MPEG2 container encapsulating a video stream and an audio stream encoding). Fortunately, metadata for built heritage maps possess two properties that can be leveraged to find suitable solutions for format migration: first, built heritage map metadata can be represented as formal ontologies [Mat05], [Wul08]. This enables us to utilize technologies that deal with ontology change, most importantly techniques for ontology evolution and ontology matching. Secondly, because of their origin in

datasheet oriented workflows, built heritage maps share a simple, yet expressive meta-structure. Representing built heritage metadata as formal ontologies gives us the ability to represent translations between different document formats as formal translation rules between different ontologies.

3 Format Migration as Ontology Change Problem

[Flo08] identify various subfields of ontology change: Methods from ontology evolution, debugging and versioning may be employed in the design and re-design phases, when an existing document format is adapted to new requirements and correctness as well as traceability of the changes need to be guaranteed. To find translation rules between two different but similar ontologies, it is possible to employ methods from the subfields of ontology matching and mapping. Ontology matching [Euz07] is the process of finding correspondences between matching elements of two different ontologies. Ontology mapping consists in applying the mappings to transform information modeled in the source ontology into the target ontology.

Various methods for ontology matching and mapping have been developed. Unfortunately, most of these methods are still limited to simple correspondences. A simple correspondence can only map single entities (concepts, roles, instances) from the source ontology onto entities from the target ontology. Simple correspondences cannot capture translations that require to consider more than one ontological element at the same time. Only limited research is available with regard to the derivation of complex mappings between ontologies [Stu08], [Švá09]. *Model based refinement* is a novel approach developed by the second author to detect complex ontology mappings. Starting with the initial mapping of two core concepts, model based refinement tentatively applies a set of refinement rules to obtain more complex mappings between source and target models. This is done by heuristically exploring the set of possible source and target models to find suitable matching models. Automated description logic reasoning is used to limit the set of detected models to only consistent models.

Figure 2 shows two snippets from built heritage ontologies. The snippets represent different modeling variants for a facade stone (ashlar). In both representations, two features of an ashlar are recorded: the integration epoch of the stone and the stone type. The modeling variant seen left in Figure 2 represents these features as a simple flat taxonomy using direct subclasses only. The ontology shown right in Figure 2 evolved from the first. It differs in a number of respects.

Some information is represented differently and an extension has been performed. The integration epoch is represented not as a subclass but as a scalar-valued attribute (i.e. a description logic data property). The values of the scalar attribute now refer to the start of the appropriate century and not the ordinal of the century. The second ontology also features stone *Type* as a standalone concept connected to *Ashlar* via a concept-concept relationship (a description logic object property). Additionally, a third stone type variant *Nürnberger* has been introduced which has no appropriate representation in the first ontology. To determine a mapping between the two ontologies represented in Figure 2,

we start with an initial mapping: *Ashlar* in the first ontology is most likely equivalent to *Ashlar* in the second ontology. Such simple mappings may be obtained using a traditional ontology mapper. Such a mapping corresponds to a list of matching models. In the initial case, a single individual of type *Ashlar* from the source ontology is mapped onto a single individual from the target ontology of the same time.

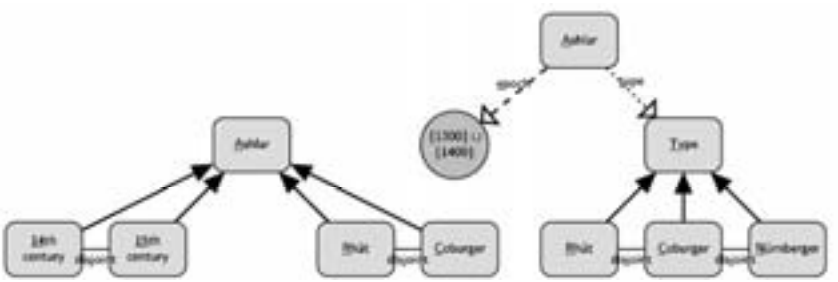


Figure 2: Ontology evolution which restructures the conceptual modeling

A stepwise refinement of the initial mapping is computed by expanding the matched models on both sides. In the first ontology, we refine the initial match by introducing subclasses. Using existing information about concept disjointness, a total of four different refinement models may be obtained: (*A*, 14th, *R*), (*A*, 15th, *R*), (*A*, 14th, *C*), (*A*, 15th, *C*). With regard to the second ontology, there are multiple refinement choices – to introduce the *epoch* attribute or introduce a new concept *Type* linked to *Ashlar*. Proceeding with this stepwise refinement process, we can additionally introduce a new *type* link at the target side and introduce appropriate subclasses of *Type*. This yields a total of six models at the target side: (*A*, 1300, *s*, *S*, *R*), (*A*, 1400, *s*, *S*, *R*), (*A*, 1300, *s*, *S*, *C*), (*A*, 1400, *s*, *S*, *C*), and (*A*, 1300, *s*, *S*, *N*), (*A*, 1400, *s*, *S*, N*)*. The desired mapping is now easily extracted.

Model based refinement is well suited to match axiomatized ontologies arising from ontology evolution because it is possible to use a description logic reasoner to eliminate inconsistent refinement early on. Additionally, comparing expanded description logic models is much simpler than comparing abstract axiomatic descriptions.

3 Discussion and Conclusions

We presented and analyzed a type of documents that is particularly vulnerable to semantic ageing – the digital maps produced by researchers who are studying built heritage. Since the maps combine spatial and thematic data, they are more complex than the mixed media documents that current preservation management approaches do handle. Furthermore, the conceptualization of the thematic data tends to evolve because the requirements of the documentation task change over time. We have shown that an novel ontology mapping technique, model based refinement, can be used to compute the format transformations needed to implement the migration of formats which semantic ageing requires.

Recently, [Ros10] has questioned whether format evolution constitutes a real threat because of market mechanisms that would prevent the obsolescence of formats. However, there exist a number of document-centred workflows in the digital humanities that are supported by special purpose software only such as the task of documenting built heritage. [Ros10] speaks of "immature markets" because the software is adopted just by a small community of expert users in which case, he admits, format obsolescence is very common. We have shown that in such cases there exist technological means to compute the necessary transformations from obsolete data formats into formats which can be accessed by current application software.

References

[Bor06] Borghoff, U., Rödig, P., Scheffczyk, J., and Schmitz, L.: Long-term Preservation of Digital Documents: Principles and Practices, Springer, Berlin, 2006.

[Coh10] Cohen, P.: Fending Off Digital Decay, Bit by Bit. The New York Times, C1. March 16, 2010

[Euz07] Euzenat, J. & Shvaiko, P.: Ontology Matching, Springer-Verlag New York, Inc. Secaucus, NJ, USA, 2007

[Flo08] Flouris, G.; Manakanatas, D.; Kondylakis, H.; Plexousakis, D. & Antoniou, G.: Ontology change: classification and survey, The Knowledge Engineering Review, Cambridge Univ Press, 2008, 23, 117-152

[Fre09] Freitag, B. & Schlieder, C.: MonArch--Digital Archives for Monumental Buildings, Künstliche Intelligenz, 2009, 4/2009, 31-35

[Hun06] Hunter, J. and Choudhury, S.: PANIC – An Integrated Approach to the Preservation of Composite Digital Objects using Semantic Web Services. International Journal on Digital Libraries, 6(2), 2006, pp. 174–183

[Kun97] Kuny, T. (1997) A Digital Dark Ages? Challenges in the Preservation of Electronic Information. In: 63rd IFLA Council and General Conference, Workshop on Preservation and Conservation, http://archive.ifla.org/IV/ifla63/63kuny1.pdf (2 Apr 2010).

[Mat05] Matyas, S. & Schlieder, C.: Generating Content-related Metadata for Digital Maps in Built Heritage, Proceedings of the First on-Line conference on Metadata and Semantics Research (MTSR'05): Advances in Metadata Research, Rinton Press Inc., 2005

[Ros10] Rosenthal, D.: Format Obsolescence: Assessing the Threat and the Defenses, Library Hi Tech, 28, 2, 2010, pp. 195-210

[Sch10] Schlieder, C.: Digital Heritage: Semantic Challenges of Long-term Preservation, Semantic Web Journal, 1, 2010

[Sch02] Schuller, M. Building Archaeology, Monuments and Sites: VII. International Council on Monuments and Sites (ICOMOS), 2002.

[Sol10] Solvberg, I. and Rauber, A.: Digital Preservation in Solvberg, I. and Rauber, A. (eds), Digital Preservation, European Research Consortium for Informatics and Mathematics, 2010, pp. 12–13.

[Stu08] Stuckenschmidt, H.; Predoiu, L. & Meilicke, C.: Learning Complex Ontology Alignments: A Challenge for ILP Research, International Conference on Inductive Logic Programming (ILP2008), 2008

[Švá09] Šváb-Zamazal, O. & Svátek, V.: Towards Ontology Matching via Pattern-Based Detection of Semantic Structures in OWL Ontologies, Proceedings of the Znalosti Czecho-Slovak Knowledge Technology conference, 2009

[Wul08] Wullinger, P. & Schlieder, C.: Digital Maps of Historical Buildings: Preservation Issues and Solutions, Proceedings of IS&T's Archiving 2008 Conference, 2008

Time Series of Linguistic Networks in the Patrologia Latina*

Alexander Mehler, Rüdiger Gleim, Ulli Waltinger & Nils Diewald
Bielefeld University

Abstract: We analyze a corpus of historical texts in terms of complex network theory. This is done by means of the *Patrologia Latina*, a collection of Latin documents that were written over a period of more than 1,000 years. We perform a lemmatization of this corpus and map its documents onto the end of the productive period of the corresponding authors. By means of this temporal ordering, we perform a transitivity analysis on the level of lexemes and sentences as a function of time. This analysis shows a remarkable law-like behavior of transitivity on the lexical and sentential level.

1 Introduction

A long-term historical text corpus comprises natural language texts that were produced over a long period of time in the past. An example is the *Patrologia Latina* (PL) [Mig55], which is a collection of writings of the Church Fathers and other ecclesiastical authors that were written over a period of more than 1,000 years. With the growing availability of tools for automatic text analysis, the processing of historical corpora comes more and more into the focus of text-technology [LPF05]. There are four areas of advancements in this area: *corpus building*, *resource formation*, *corpus management*, and *corpus mining*. With a focus on Latin, these approaches can be reviewed as follows:

- *Corpus building and resource formation:* a multitude of projects aim at building text collections and related lexical and syntactic resources. This includes, for example, corpora of Latin texts of the early modern time as provided by the *Camena* project [Sch01]) as well as full-text databases of Latin literature as included in the *Bibliotheca Teubneriana Latina* (BTL-1) [Tom99] and its companion editions. As one of the most sophisticated projects in this area, the *Perseus Digital Library* [SRCC00] provides not only historical corpora, but also syntactic annotations [BPBC08]. See [Kos05] and [PD10] for related endeavors. Note that all these projects integrate full-form lexica of Latin.
- *Corpus management and mining:* the Perseus project has been built as a digital library that operates on historical corpora of several languages. Recently, two additional projects have started that also integrate facilities for corpus management in order to support text mining on Latin texts. Firstly, this relates to the *eAQUA* project

*Financial support of the German Federal Ministry of Education and Research (BMBF) through the project *Linguistic Networks* (see www.linguistic-networks.net) is gratefully acknowledged.

[BHG08], which aims at extracting linguistic knowledge from ancient resources by means of explorative data analysis. Secondly, the *Linguistic Networks Project* and its eHumanities Desktop [GM10] provides, amongst others, access to the PL by means of related methods. Both approaches leave the narrow focus on present-day corpora in order to explore historical corpora and, thus, gain access to language change.

In this paper, we analyze the PL as a long-term historical corpus of Latin texts. This is done with the help of mapping all PL documents onto the dates of the end of the productive period of their authors. By means of this temporal ordering we derive a subcorpus of 1,000 texts in order to perform a cluster analysis on the level of lexemes and sentences as a function of time (Section 3): starting from a graph model of layered linguistic networks (Section 2), we utilize complex network theory to get insights into the temporal dynamics of the PL as a long-term historical corpus (Section 4). We show a remarkable law-like behavior of clustering on both the lexical and the sentential level. In order to provide comparative values, we analyze a corpus of present-day texts in the same framework.

2 A Graph Model of 2-layer Linguistic Networks

A natural language is a multiresolutional system that is structured on various, interwoven linguistic levels (e.g., the morphological, lexical or syntactic level). From a formal point of view, such a system can be modeled as an instance of the class of *multilayer graphs*. Generally speaking, a directed *k-layer graph*

$$D \quad = \quad (V, A, B) \tag{1}$$

is a digraph whose vertex set is partitioned into non-empty, pairwise disjoint subsets $V_1, \ldots, V_k$ such that the following conditions hold:

- *Layer-external arcs:* $\forall a \in A \exists! V_i, V_j \in \{V_1, \ldots, V_k\} : \mathrm{in}(a) \in V_i \neq V_j \ni \mathrm{out}(a)$.
- *Layer-internal arcs:* $\forall a \in B \exists! V_i \in \{V_1, \ldots, V_k\} : \mathrm{in}(a) \in V_i \ni \mathrm{out}(a)$.

The ith layer of a k-level graph, $i \in \{1, \ldots, k\}$, is defined by the subgraph $D(i) = D_i = (V_i, B_i)$ of (V, B) such that $a \in B_i \leftrightarrow \mathrm{in}(a) \in V_i \ni \mathrm{out}(a)$. Likewise, the bipartite graph that is induced by the layers i, j is defined by the subgraph $D(i, j) = D_{ij} = (V_i \cup V_j, A_{ij})$ of (V, A) such that $a \in A_{ij} \leftrightarrow (\mathrm{in}(a) \in V_i \wedge \mathrm{out}(a) \in V_j) \vee (\mathrm{in}(a) \in V_j \wedge \mathrm{out}(a) \in V_i)$ (note that $D_{ij} = D_{ji}$). Subsequently, we experiment with two linguistic layers, the lexical and the syntactic layer. That is, we deal with 2-layer graphs $D = (V, A, B)$ where A is the set of arcs that either connect words with sentences or vice versa, while B is the union of lexical and sentential arcs. Thus, we have to induce three graphs D_1, D_2 and D_{12} to build a 2-layer graph where D_1 is a co-occurrence graph, D_2 is a graph of sentential relations and D_{12} links sentences with their lexical constituents:

- $D_1 = (V_1, B_1)$ is induced as a lexical co-occurrence network by means of the co-occurrence measure $\sigma : V_1 \times V_1 \to \mathbb{R}_0^+$ of [HQW06]. That is, vertices $v \in V_1$ denote

lexical units that are output by lemmatizing the PL, while arcs denote significant lexical co-occurrences in the sense of σ. This procedure of network induction is described in detail in [MDW$^+$10]. Note that in contrast to [HQW06], we do not operate on word forms, but on lexemes. Note further, that for every arc $a \in B_1$ there is an arc $b \in B_1$ such that $\mathrm{in}(a) = \mathrm{out}(b)$ and $\mathrm{in}(b) = \mathrm{out}(a)$ since lexical associations are, as mapped here, symmetric.

- $D_{12} = (V_{12}, A_{12}) = (V_{12}, A)$ is induced as follows: the vertex set of D_{12} is the union of the set of lexemes V_1 and the set of sentences V_2. Further, for any lexeme $v \in V_1$ for which there exists a word form as a lexical constituent of the sentence $w \in V_2$, we generate two non-parallel but multiple arcs as elements of A_{12} that end at v and w, respectively.

The third step is to induce the sentence network $D_2 = (V_2, B_2)$ for which $B_2 = B \setminus B_1$. Our starting point is to connect sentences that express similar concepts, that is, sentences with similar conceptual meanings. Evidently, the automatic detection of these meanings is out of reach so that we need a heuristic. We do this by means of the following hypothesis:

> **H1:** *The more lexical units two sentences have in common and the higher the degree of semantic specificity of these units, the higher the contribution of these common lexical constituents to the conceptual similarity of both sentences.*

Obviously, sentences that have no lexical constituents in common can still be semantically related, for example, by a relation of entailment. Such relations are currently out of reach of being automatically detected in huge corpora such as the PL. Therefore, we start from H1 by saying that the probability of a conceptual similarity of two sentences is a function of the semantic specificity and the number of the lexical constituents they share. This approach can be implemented straightforwardly. In set-theoretical terms, sentences can be represented as collections of lexical units so that multisets can be used to model the first parameter, that is *number*. Further, the notion of semantic specificity can be modeled by means of *idf*-scores [SB88]. Now, let $v_i, v_j \in V_2$ be two sentences that are represented as multisets S_i and S_j. Then, the *lexical overlap* of v_i and v_j is computed as

$$\omega(S_i, S_j) = \frac{\sum_{x \in S_i \cap S_j} i(x)}{\sum_{x \in S_1} i(x) + \sum_{x \in S_2} i(x) - \sum_{x \in S_1 \cap S_2} i(x)} \tag{2}$$

where $S_i \cap S_i$ is the multiset intersection of the set representations of v_i and v_j. $i(x)$ is the *inverse document frequency* (idf) of x in the underlying reference corpus, that is, the PL. σ has the property that if S_1 and S_2 are identical, then $\sigma(S_1, S_2) = 1$. Otherwise, if their intersection is empty, then $\sigma(S_1, S_2) = 0$. Note that we do not say that the conceptual similarity of sentences equals their lexical overlap. Rather, we use $\omega(S_i, S_j)$ to approximate the impact of shared lexical constituents to conceptual similarity in the sense of H1.

Variable	Value	Variable	Value
number of authors	1,320	number of sentences	7,727,864
number of texts	4,555	number of tokens	121,722,687
number of paragraphs	674,718	number of word forms	1,094,850

Tabelle 1: Quantitative characteristics of C1, the subset of the PL without commentaries.

3 The Corpora

In order to instantiate our model of 2-layer graphs, we start from a subset of 4,555 texts of the PL by excluding all commentaries (see Table 1). Henceforth, this subcorpus is denoted by *C1*. It is input to a preprocessing module that includes a sentence boundary detection, named entity recognition, lemmatization and an annotation of the logical document structure of all texts in *C1*. The details of this preprocessing and of the preceding formation of a full-form lexicon are described in [MDW$^+$10]. In this section, we concentrate on describing the process of network induction based on the PL. In order to arrive at a manageable time series analysis, we perform two steps of corpus selection: firstly, we rank all texts in *C1* by their number of tokens and select the 1,000 highest ranked documents that follow the first 14 ranks. This subcorpus is denoted by *C2*. As we induce for each of the documents in *C2* a 2-layer graph, we have to analyze 2,000 networks in terms of their topological characteristics. Obviously, this is a huge comparative network analysis. Thus, the reason to build *C2* is to reduce computational effort and to keep the corpus more balanced. Secondly, we perform a mapping of the documents of the PL onto the date of the end of the productive period of the corresponding author. This is done to get a temporal ordering of *C2*. As a result of additionally applying a random ordering of documents with the same time values, we get a linear ordering $C = (x_1, \ldots, x_{1000})$ of *C2*. The time values are annotated manually where we use the *Documenta Catholica Omnia* to get the author information. As a result of this selection, we get the subcorpus *C3* of 1,000 documents that are ordered by the end of the productive period of their author. *C3* is input to induce a separate 2-layer graph for each of the 1,000 documents according to Section 2. That is, we map *C3* onto a corresponding time series of 2-layer graphs

$$(D^{[1]}, \ldots, D^{[n]}) \tag{3}$$

where for each $t \in \{1, \ldots, n\}$: $D^{[t]} = (V^{[t]}, A^{[t]}, B^{[t]})$, $n = 1000$, is the two-level graph of lexical and sentential networking in document x_t at time t such that $D_1^{[t]} = D^{[t]}(1)$ is the lexical and $D_2^{[t]} = D^{[t]}(2)$ is the sentential layer of $D^{[t]}$.

As a basis of comparison we analyze a corpus of newspaper articles of the *Süddeutsche Zeitung* (SZ) from 1994 by the same procedure. The difference is that 2-layer networks are not induced for each article as this would get tiny graphs, but not complex networks. Instead, we deal with each daily issue as a single document for which we induce a lexical and sentential network as described above. This gives a time series of 301 2-layer networks.

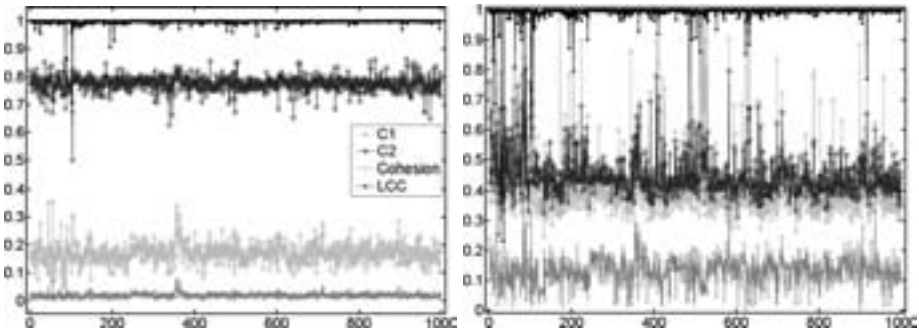


Abbildung 1: Time series of indices of lexical (left) and sentential networks (right) in the PL.

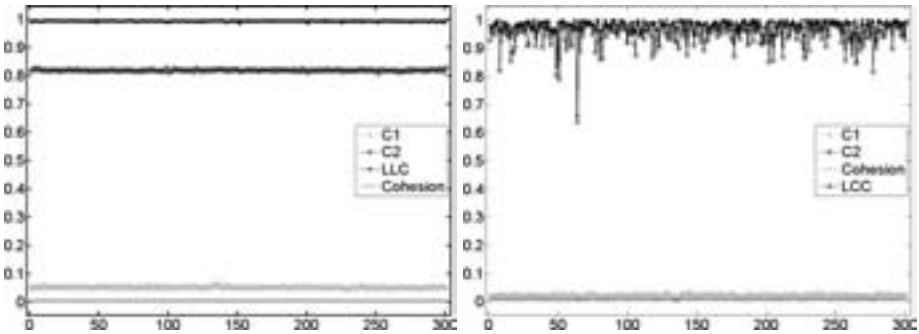


Abbildung 2: Time series of indices of lexical (left) and sentential networks (right) in the SZ.

4 Variation of Clustering of Textual Units in Time

In order to get insights into the temporal dynamics of the PL by example of the time series of 2-layer graphs (see Equation 3) we compute four topological indices of complex networks. The density or cohesion of a graph $G = (V, E)$ is the number of its edges in relation to the number of all possible edges in G: $coh(G) = \frac{\sum_{v \in V} d_G(v)}{|V|^2 - |V|} \in [0, 1]$ where $d_G(v)$ is the degree of $v \in V$. Note that $coh(G)$ is computed for the undirected variant of $D_1^{[t]}$ and $D_2^{[t]}$, respectively, $t \in \{1, \ldots, n\}$. Our expectation is that linguistic networks of the sort considered here are sparse networks by analogy to small-world graphs [New03]. This expectation is confirmed on the lexical and sentential level by example of the newspaper corpus (see Figure 2). In both cases we observe a cohesion near to zero. Interestingly, this observation correlates with the fact that the fraction lcc of vertices that belong to the *Largest Connected Component* (LCC) is near to 1 – although in the sentential network it is smaller than in the lexical network. Thus, lexical and sentential networking results in highly sparse though connected graphs in the case of present-day language.

In the case of the PL, we observe a remarkably high value of lcc and a near to zero cohesion on the lexical level. However, if we look on the sentential level, we observe a difference

to the newspaper corpus: although a fraction of nearly 100% of vertices that belong to the LCC is retained in almost all cases, the cohesion is much larger. One reason for this difference is that while in the case of the Pl, sentential networks are computed by including all links of at least average specificity (in the sense of Equation 2), in the case of the SZ we retain a connectedness of nearly 100% if we include only those sentence links whose weight is $\geq \mu + 5\sigma$. If we apply the same criterion to the sentence networks of the PL, we get disconnected networks where $lcc \leq 70\%$. Thus, we conclude that networking is much more stable in the present-day corpus than in the historical corpus. However, we also observe a remarkably stable network pattern in the case of historical lexical networks. To a minor degree this also holds for the historical sentential networks.

The second pair of indices that we consider relates to what is called transitivity in quantitative sociology, that is, the probability by which neighbors of the same vertex (e.g., a person) are themselves neighbors (e.g., friends). More formally, the transitivity or cluster value of an undirected graph $G = (V, E)$ is as follows [WS98]:

$$C_2(G) = C_{ws}(G) = \frac{1}{n}\sum_{i=1}^{n} c_{ws}(v_i) = \frac{1}{n}\sum_{i=1}^{n} adj(v_i) / \binom{d_G(v_i)}{2} \in [0,1] \tag{4}$$

where $adj(v_i)$ is the number of edges ending only at neighbors of v_i. [BR03] discuss an interesting alternative to C_2 that is computed as follows:

$$C_1(G) = C_{br}(G) = \left(\sum_{v \in V} \binom{d_G(v)}{2} c_{ws}(v)\right) / \sum_{v \in V} \binom{d_G(v)}{2} \tag{5}$$

The difference of C_2 and C_1 lies in the fact that the latter is a weighted mean in contrast to C_2, which is an arithmetic mean. C_1 weights the impact of the transitivity of each vertex v by its degree: the higher $d_G(v)$, the higher the impact of $c_{ws}(v)$ to C_1.

If we look on Figure 1 and 2 we observe a remarkably stable distribution of C_2 and C_1 on the level of lexical networks (though to a minor degree in the case of lexical networking in the PL). This patterned distribution is also observable in the case of sentential networks derived from the present-day corpus and to a much smaller degree in the case of the historical corpus. Interestingly, this exception correlates with the fact that in the case of sentential networking in the PL, C_1 and C_2 have a similar spectrum of values. In all other cases we observe a tremendous divergence of both cluster values. This is a hint that in these networks, high-degree vertices have small cluster values, while in the sentence networks of the PL, high-degree and low-degree vertices tend to have the same impact. These findings are in support of [BR03] who stress the need to use weighted cluster coefficients, which are more expressive in terms of network topology.

In summary: with the exception of historical sentence networks, we observe an almost constant transitivity of linguistic networks *irrespective of time*. Thus, clustering, cohesion and connectedness evolve as stable topological indicators that hint at a law-like linguistic networking. However, we also observe remarkable differences between historical and present-day language. There may be at least four reasons for these differences: in the case of the newspaper corpus we can expect much more stable patterns of text production.

There is certainly a tendency to highly stabilized text types due to a standardized process of text production. This level of standardization may result in the almost constant cluster coefficients of lexical and sentential networks. Secondly, we have to expect less heterogeneous genres in the case of newspaper articles in contrast to the PL with its indices, sermons, handbooks etc. Thus, we need a more precise *text-typological* analysis of the PL and a sophisticated look at *clustering in different genres* in order to underpin our findings. Thirdly, the SZ is a short-term corpus, while the PL is a long-term corpus. This difference in temporal scale may be an additional source of differences in the sense that the PL manifests language change that is not present in the SZ. Finally, the PL-based networks are computed per document, while the SZ-based networks are computed per daily issue. The corresponding differences in text coherence may also be a source of the deviations.

5 Conclusion

We analyzed a subset of the Patrologia Latina in terms of its temporal dynamics. We found a pattern of clustering on the lexical level and on the sentential level, though to a minor degree. In the case of lexical networking, this pattern is stable irrespective of the period of time. Future work aims at investigating this pattern using larger sets of reference corpora.

Literatur

[BHG08] Marco Büchler, Gerhard Heyer und Sabine Gründer. eAQUA: Bringing modern Text Mining approaches to two thousand years old ancient texts. In *Proc. of the e-Humanities workshop at the 4th IEEE Int. Conf. on e-Science*, 2008.

[BPBC08] David Bamman, Marco Passarotti, Roberto Busa und Gregory Crane. The Annotation Guidelines of the Latin Dependency Treebank and Index Thomisticus Treebank. In *Proceedings of LREC 2008*, Marrakech, Morocco, 2008. ELRA.

[BR03] Béla Bollobás und Oliver M. Riordan. Mathematical Results on Scale-Free Random Graphs. In Stefan Bornholdt und Heinz Georg Schuster, Hrsg., *Handbook of Graphs and Networks*. Wiley-VCH, Weinheim, 2003.

[GM10] Rüdiger Gleim und Alexander Mehler. Computational Linguistics for Mere Mortals — Powerful but Easy-to-use Linguistic Processing for Scientists in the Humanities. In *Proceedings of LREC 2010*, Malta, 2010. ELDA.

[HQW06] Gerhard Heyer, Uwe Quasthoff und Thomas Wittig. *Text Mining: Wissensrohstoff Text*. W3L, Herdecke, 2006.

[Kos05] Cornelis H. A. Koster. Constructing a Parser for Latin. In Alexander F. Gelbukh, Hrsg., *Proceedings of CICLing 2005*, Seiten 48–59, 2005.

[LPF05] Anke Lüdeling, Thorwald Poschenrieder und Lukas C. Faulstich. DeutschDiachron-Digital. *Jahrbuch für Computerphilologie*, Seiten 119–136, 2005.

[MDW+10] Alexander Mehler, Nils Diewald, Ulli Waltinger, Rüdiger Gleim, Dietmar Esch, Barbara Job, Thomas Küchelmann, Olga Pustylnikov und Philippe Blanchard. Evolution of Romance Language in Written Communication. In *Arts, Humanities, Complex Networks: a Leonardo satellite symposium at NetSci 2010*, 2010.

[Mig55] Jacques-Paul Migne, Hrsg. *Patrologiae cursus completus: Series latina*, Jgg. 1–221. Chadwyck-Healey, Cambridge, 1844–1855.

[New03] Mark E. J. Newman. The structure and function of complex networks. *SIAM Review*, 45:167–256, 2003.

[PD10] Marco Passarotti und F. Dell'Orletta. Improvements in Parsing the Index Thomisticus Treebank. In *Proceedings of LREC 2010*, 2010.

[SB88] Gerard Salton und Chris Buckley. Term Weighting Approaches in Automatic Text Retrieval. *Information Processing Management*, 24(5):513–523, 1988.

[Sch01] Wolfgang Schibel. CAMENA-Neulateinische Dichtung im WWW. *Neulateinisches Jahrbuch*, 3:211–219, 2001.

[SRCC00] David A. Smith, Jeffrey A. Rydberg-Co und Gregory R. Crane. The Perseus Project. *Literary and Linguistic Computing*, 15(1):15–25, 2000.

[Tom99] Paul Tombeur, Hrsg. *Bibliotheca Teubneriana Latina (BTL 1)*. Teubner/Brepols, Stuttgart, Leipzig/Turnhout, 1999.

[WS98] Duncan J. Watts und Steven H. Strogatz. Collective Dynamics of 'Small-World' Networks. *Nature*, 393:440–442, 1998.

Workshop

Protokoll-basierte Modellierung von Geschäftsinteraktionen

Johannes Reich, Prof. Dr. Andreas Speck

Vorwort zum Workshop "Protokoll basierte Modellierung von Geschäftsinteraktionen"

Dieser Workshop war als Fortsetzung des Workshops "Spiele, Geschäftsprozesse und Interaktionsmodelle" der 39. Jahrestagung der Gesellschaft für Informatik 2009 in Lübeck gedacht.

Die Durchdringung moderner Unternehmen mit Software, die direkt in Geschäftsinteraktionen involviert ist, ist in den letzten Jahren enorm gewachsen. Es ist deshalb eine naheliegende Forderung, dass die - aus unternehmerischer Sicht selbstverständliche - Ausrichtung unternehmerischen Handelns an den notwendigen Geschäftsinteraktionen, sich möglichst einfach in der Software abbilden lassen sollte, die in den Unternehmen die entsprechenden Geschäftsinteraktionen unterstützen.

Damit die Beschreibung von Geschäftsinteraktionen auch aus Applikationssicht laufzeitrelevant werden kann, ist es notwendig, diese Interaktionsbeschreibungen einer formalen Interpretation zugänglich zu machen, aus der ihre Beziehung zu berechenbaren Systemen eindeutig hervorgeht. Von einer solchen laufzeitrelevanten Interaktionsbeschreibung sind vielversprechende Impulse hinsichtlich der Automatisierung der Implementierung von Geschäftsinteraktionen zu erwarten. Insbesondere die KMU-Unternehmen, die momentan von den immer noch zu hohen Anfangsinvestitionen in diesem Bereich abgeschreckt werden, sollten davon profitieren.

Ausgangspunkt dieses Workshops ist die Annahme, dass die Ökonomie mit ihrem spieltheoretisch fundierten Spielbegriff dieselben Interaktionen beschreibt, wie die Informatik bei ihrer Beschreibung der nichtdeterministischen Interaktion von Geschäftsprozessen. Wegen der engen formalen Beziehung von (ökonomischem) Spiel und (informatischem) Protokoll liegt es nahe, den Protokollbegriff nicht nur informell zur Illustration zu verwenden, wie dies viele heutige Ansätze tun, sondern die formale Basis für die informatische Beschreibung von Geschäftsinteraktionen in der Protokoll-Theorie zu suchen, im Sinne einer Theorie der konsistenten finiten Interaktion von (ggfs. finiten) Systemen.

In Ihrem Vortrag "A computational study of an automated negotiation scheme to solve multiple criterion single machine scheduling problems" stellen René Ramacher und Lars Mönch einen Mediator-basierten Verhandlungsmechanismus vor, zur Lösung des Problems der Ablaufsteuerung einer einzelnen Maschine nach multiplen Kriterien. Zwei Agenten mit unterschiedlichen Zielen konkurrieren um diese Ressource. Der eine zielt auf Minimierung der mittleren Bearbeitungszeit, der andere auf Minimierung der maximalen Latenz. Es wird das Verhalten von egoistischen und kooperativen Agenten untersucht. Die Performance des vorgeschlagenen Verhandlungsprotokoll wird experimentell ermittelt und mit einem theoretischen Modell verglichen.

Herwig Unger und Coskun Akinalp modellieren in ihrem Vortrag "Berücksichtigung menschlicher Charaktereigenschaften in geschäftlichen Transaktionen" den Einfluss des Charakters auf den Ausgang einer Geschäfts- bzw. Spielsituation mit Wettbewerbscharakter. Ihrem Charaktermodell legen sie die drei Parameter Balance, Stimulanz und Dominanz zugrunde. Die weitere Untersuchung geschieht mittels eines in ein Nullsummenspiel abgewandeltes El-Farol-Bar Problem (eine Form eines Minoritätenspiels), bei dem die Spieler ihre Aktivitäten je nach Erfolg im Spiel ändern können.

Johannes Reich beweist in seinem Artikel ”Finite system composition and interactions”, dass eine protokoll-basierte Interaktion zwischen zwei finiten Systemen dann zur Bildung eines Supersystems führt, wenn das eine System die Aktionen des anderen determiniert. Damit ergibt sich die Eigenschaft eines Systems ein Subsystem zu sein aus dem Kontext der Interaktion. Seine Arbeit wirft damit ein interessantes Licht auf Fragen der Systemkomposition, insbesondere bezüglich der Bedeutung von ”loser Kopplung” interagierender Systeme.

In ihrem Artikel "Realization of Business Processes"demonstrieren Andreas Speck, Andreas Rusnjak und Marcel Schulz einen Ansatz, wie ausführbare Workflow Modelle (Intershop Pipelines), in eine Automatenbeschreibung transformiert werden können als Basis für eine formale Verifikation. Als Verifikationstechnologie wird Modelchecking vorgeschlagen, bei der die Automaten gegen Spezifiaktionsausdrücke der Computational Tree Logic (CTL) getestet werden. Die Autoren geben an, gute Erfahrungen mit der Formulierung entsprechender CTL-Ausdrücke durch die zuständigen Qualitätsexperten gemacht zu haben.

Alexander Pokahr und Lars Braubach zeigen in ihrem Artikel ”Reusable Interaction Protocols for Workflows”, wie man vordefinierte, domänen-unspezifische Teil-Protokolle mittels BPMN beschreibt und diese in domänen-spezifischen Prozessen (wieder-)verwendet. Die mittels BPMN beschriebenen Regeln des Nachrichtenaustauschs werden mit Entscheidungsstellen versehen, in der die domänenspezifische Logik mit abstrakten Prozessen modelliert wird. Die Wiederverwendung eines solchen Teil-Protokolls erfordert dann eine Referenz auf die konkrete Prozessimplementierung der domänenspezifischen Logik. Als Beispiel werden contract-net Verhandlungen basierend auf mehreren Teil-Protokollen modelliert.

In seinem Vortrag "Formalizing Dependencies in Business Process Models Using Constraint Satisfaction"beschreibt Wolfgang Runte die Verwendung von Bedingungserfüllungsproblemen (Constraint Satisfaction Problems) im Kontext des Compliance Managements von Geschäftsprozessen. Es werden beispielhaft verschiedene Geschäftsprozesse mittels Bedingungserfüllung (constraint satisfaction) beschrieben und eine Kategorisierung der Bedingungen vorgeschlagen.

Die Mitglieder im Programmkomitee waren (in alphabetischer Reihenfolge) Lars Braubach (Universität Hamburg), Alejandro Buchmann (Technische Universität Darmstadt), Bernd Finkbeiner (Universität Saarbrücken), Hans-Jörg Kreowski (Universität Bremen), Martin Leucker (Technische Universität München), Christoph Liebig (SAP AG, Walldorf), Lars Moench (Fernuniversität Hagen), Ralf Peters (Martin-Luther-Universität Halle-Wittenberg), Alexander Pokahr (Universität Hamburg), Elke Pulvermüller (Universität Osnabrück), Johnnes Reich (SAP AG, Walldorf), Norbert Ritter (Universität Hamburg), Stefan Sackmann (Martin-Luther-Universität Halle-Wittenberg), Mareike Schoop (Universität Hohenheim), Andreas Speck (Christian-Albrechts-Universität zu Kiel) und Herwig Unger (Fernuniversität in Hagen). Damit bleibt uns zu guter Letzt, uns herzlich für ihr Interesse und Engagement zu danken. Ohne sie wäre dieser Workshop nicht möglich gewesen.

Andreas Speck und Johannes Reich

A Computational Study of an Automated Negotiation Scheme to Solve Multiple Criterion Single Machine Scheduling Problems

René Ramacher, Lars Mönch

Department of Mathematics and Computer Science
University of Hagen
Universitätsstraße 1
58079 Hagen
Rene.Ramacher@fernuni-hagen.de
Lars.Moench@fernuni-hagen.de

Abstract: In this talk, an automated negotiation mechanism is presented to solve multiple criterion single machine scheduling problems. We consider a single machine that is utilized by two agents that have own jobs and private objectives. The objective of the first agent consists in minimizing the total weighted completion time of its jobs whereas the objective of the second agent is related to minimizing maximum lateness of its jobs. Scheduling problems of this type have recently attracted the interest of researchers in scheduling theory (cf. [BS03], [BF09]). The researched problem is also motivated by scheduling problems found in semiconductor manufacturing. The basic ingredient of the mechanism is a mediator that proposes contracts that are sent to the two agents as suggested by Fink [Fi06]. The contracts are proposed using a variable neighborhood search (VNS) technique. We study the behavior of greedy and cooperative agents. Furthermore, we study a hybrid strategy suggested by Klein et al. [KF03a], [KF03b] where the mediator accepts a deterioration of the two objective values with a certain probability. The performance of the suggested negotiation protocols is assessed using a large set of randomly generated problem instances. It turns out that the solutions determined by the automated negotiation mechanism are close to the Pareto frontier that can be derived when a centralized approach with full information is assumed. We use the NSGA-II algorithm [DP02] to determine the solutions on the Pareto frontier.

References

[BS03] Baker, K.; Smith, J. C.: A Multiple-Criterion Model for Machine Scheduling. Journal of Scheduling, 6, 2003, pp. 7-16.

[BF09] Balasubramanian, H.; Fowler, J.; Keha, A.; Pfund, M.: Scheduling Interfering Job Sets on Parallel Machines. European Journal of Operational Research, 199, 2009, pp. 55-67.

[DP02] Deb, K.; Pratap, A.; Agarwal, S.; Meyarivan, T.: A Fast and Elitist Genetic Algorithm: NSGA-II. IEEE Transactions on Evolutionary Computation, 6(2), 2002, pp. 182-197.

[Fi06] Fink, A.: Supply Chain Coordination by Means of Automated Negotiations Between Autonomous Agents. In: Multiagent based Supply Chain Management, Studies in Computational Intelligence, Vol. 28, 2006, pp. 351-372.

[KF03a] Klein, M.; Faratin, P.; Sayama, H.; Bar-Yam, Y.: Negotiating Complex Contracts. Group Decision and Negotiation, 12, 2003, pp. 111-125.
[KF03b] Klein, M.; Faratin, P.; Sayama, H.; Bar-Yam, Y.: Protocols for Negotiating Complex Contracts. IEEE Intelligent Systems, 18(6), 2003, pp. 32-38.

Berücksichtigung menschlicher Charaktereigenschaften in geschäftlichen Transaktionen

Herwig Unger, Coskun Akinalp

Fernuniversität in Hagen
Fakultät für Mathematik und Informatik
Universitätsstraße 27, PRG
D-58084 Hagen, Germany
Herwig.Unger@FernUni-Hagen.de, coskun@akinalp.com

Extended Abstract:

Obgleich ein großer Teil der Komponenten von Geschäftstransaktionen in bzw. mit Hilfe des Internets weitgehend automatisiert ablaufen können, bleibt der Mensch als Endverbraucher, Kunde oder Akteur in irgendeiner Position in den immer komplizierter werdenden Supply-Chain-Netzwerken eine Größe, dessen Aktivitäten schwer vorherzusagen sind. Die meisten bisherigen Ansätze versuchen, durch formal definierte Prozesse und Fristen sowie den Einsatz von BI-Lösungen den Einfluss des Zufalls weitgehend zu minimieren. Ein weiterer Ansatz ist das Studium der sozialen Beziehungen und Abhängigkeiten der Nutzer und der sich aus diesen ergebenden Konsequenzen mit formalen Mitteln, wie etwa die Modellierung und Analyse der gebildeten Netzwerke z.B. durch Graphenmodelle.

Im Marketingbereich geht Häusel [HH07] von der Nymphenburggruppe einen anderen, innovativeren Weg. Er beruft sich auf Ergebnisse psychologischer Untersuchungen, die den menschlichen Entscheidungsprozeß untersuchen und festegestellt haben ‚dass nur etwa 1/3 der Entscheidungen vom Bewusstsein getroffen werden, während 2/3 des Entscheidungsprozesses vom Unterbewusstsein vorbereitet bzw. durchgeführt werden. Interessant ist hierbei, dass es nicht beliebig viele Verhaltensmuster bei Menschen gibt. Vielmehr lassen sich in dem von den 3 wesentlichen quantifizierbaren, charakterbildenden Parametern Balance (Ausdruck der Stabilität und Sicherheit), Stimulanz (Ausdruck des Explorations- und Entdeckungsfreudigkeit) und Dominanz (Ausdruck des Konkurrenz- und des Verdrängungsverhaltes) aufgespannten Raum eindeutig 8 Cluster identifizieren, die 8 grundlegenden Charaktertypen von Menschen zugeordnet werden können. Häusel zeigt in weiteren Arbeiten, dass diese Charaktereigenschaften im wesentlichen das Kaufverhalten bis hin zur gezielten Markenauswahl beeinflussen können. Bislang sind diese Ergebnisse jedoch nicht angewendet worden, um

1. den Charakter eines Nutzers aus seinem Verhalten in Geschäftsprozessen zu bestimmen,
2. eine Vorhersage des Ressourcenbedarfs eines so klassifizierten Nutzers zu verbessern oder
3. den Einfluss des Charakters auf den Ausgang einer Geschäfts- bzw. Spielsituation mit Wettbewerbscharakter

zu untersuchen. Genau an dem wesentlichen letzten Punkt setzen die aktuellen Arbeiten der Autoren an.

Um die eigentlichen Effekte unabhängig untersuchen zu können, wird in diesem Fall auf die relative einfache Umgebung von Minoritätenspielen zurückgegriffen, die zur Modellierung bestimmter Eigenschaften von Marktprozessen entworfen wurden. Im speziellen wird ein in ein Nullsummenspiel abgewandeltes El-Farol-Bar Problem [AB94] verwendet. Im Unterschied zu dem klassischen Problem können die Spieler Ihre Aktivitäten nicht nur im Rahmen von Strategien planen sondern diese Strategien je nach Erfolg im Spiel (Höhe des Payoffs) auch ändern. Die jeweiligen Änderungen hängen dabei mit ihrer Intensität und Häufigkeit von dem jeweiligen limbischen Spielercharakter ab. Somit wird der bislang rein zufällige Spieler bzw. der rein rationale Spieler aus spieltheoretischen Untersuchungen wesentlich feiner modelliert.

Für die verschiedenen Simulationen können ferner Spielerpopulationen mit unterschiedlichen Charakterzusammensetzungen erzeugt werden. So können entweder nur Spieler 2er verschiedener Charakterklassen zusammenspielen oder aber eine Charakterverteilung gemäß der von [BC08] bekannten Verteilung der limbischen Charaktere in der deutschen Bevölkerung erzeugt werden.

Die im Vortrag präsentierten Simulationsergebnisse zeigen eindeutig, das unterschiedliche Verhalten der Charaktere und insbesondere, dass je nach Spielsituationen (insbesondere je nach Festlegung der Spielparameter wie z.B. dem Schwellwert) einzelner Charaktergruppen eindeutig Vorteile haben, d.h. einen über viele Spielrunden statistisch signifikant höhere Gewinnsumme einfahren.

Dieses erste wichtige Ergebnis gibt die Grundlage, über eine Verallgemeinerung und Anwendung der limbischen Charaktere in der Modellierung von Geschäftsprozessen detailliert nachzudenken, was als nächste Schritte ein Betrachtung der inobgen Punkten 1 und 2 erfordert.

Referenzen:

[AB94] Arthur W.B., (1994) Amer. Econ Rev. 84:406-411
[BC08] Burda Community Network Gmbh (2008), Typologie der Wünsche, Offenburg
[HH07] Häusel, H. G. (2007), Think Limbic, Rudof Haufe Verlag Gmbh

Finite system composition and interaction

Johannes Reich

Gerbersruhstraße 147
69168 Wiesloch
Johannes.Reich@sophoscape.de

Abstract: In this article, it is proven for finite systems that if by reciprocal interaction, one finite system determines the action of another finite system, then both systems become subsystems of a larger supersystem.

To achieve this result, the notion of a finite system is formalized and the rules for sequential and parallel system composition are provided. The reciprocal interaction is captured by the protocol concept.

Being part of a larger supersystem is shown not to be a property which can be attributed to the system itself but depends on the context of its interaction, namely whether its interactions determine its behavior or not.

The result seems to be especially relevant in Information Systems and eCommerce, as it raises concerns about what end-to-end for example in a security context in a system theoretic sense really means. It also demonstrates the tight connection between our system and our function notion and thereby contributes to a better understanding, why approaches that rest mainly on the function notion struggle so much with networklike interacting systems and the necessary ”loose coupling” in the sense of a sensible interaction with only little information about the internal state of the other actors.

Realization of Business Processes

Andreas Speck, Andreas Rusnjak, Marcel Schulz

aspe@informatik.uni-kiel.de
aru@informatik.uni-kiel.de
marcel.schulz@intershop.de

Abstract: Business process models are the fundamental models of commercial systems. Therefore the business processes are issue of optimization as well as verification. When a business process has been identified as comparatively good (or optimal) it is of interest if the realization of the system still realizes this process. In this paper we present an approach how to verify such realizations. As application system we use the e-commerce system Intershop Enfinity which provides an executable process model: the *Pipeline Model*. These *Pipeline Models* have to be transferred to formal automata models which then may be checked. As tools to check we use model checkers.

1 Introduction

A main concern for the optimization of business processes is the realization of business processes. This means that it is desirable that a comparatively good (or optimal) model of a business process is also realized by the implementation of the system. The other way round it may also be of interest to know if an implementation does not meet the specified process model. In this case it may be analyzed if there is at all a chance to realize the optimized business process model.

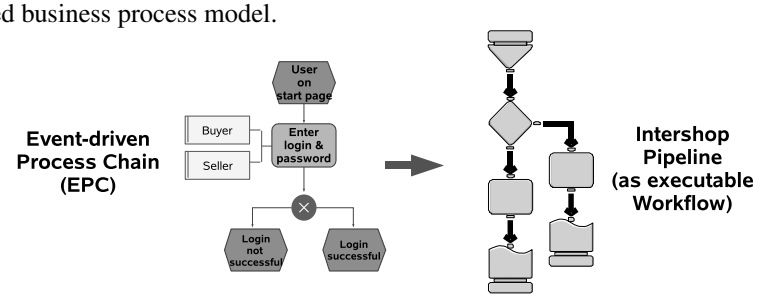


Figure 1: EPC Model and *Pipeline Model*

Due to the lack of standardized rules there are almost infinite possibilities for the realization of business processes. Therefore, we need to select a specific methodology or technology to implement business processes. In our case we focus on executable models, or to be more precise on executables workflow models. A well-known example for such an executable workflow model may BPEL4WS modeling the execution order of web services or the proprietary *Pipeline Model* of Intershop Enfinity [Bre02].

Similarely to BPEL4WS the *Pipeline Model* connects executable services (here named *Pipelets*). The *Pipelets* as nodes in the *Pipeline Model* may trigger the execution of comparatively large service (realized as EJBs). The differences between the *Pipeline Model* a BPEL4WS are: *Pipeline Models* do not model the interaction of distributed services like web services. Usually *Pipeline Models* focus on the interaction within one server. *Pipeline Models* are proprietary models limited to the e-commerce system Intershop Enfinity. The *Pipeline Models* are executed by a modified version of the Tomcat application server being part of the Intershop Enfinity system. The *Pipeline Models* were introduced before the web services and BPEL4WS were invented. *Pipeline Models* are part of the ARIS [Sch98] profile *ARIS for Enfinity* [Bre02].

2 EPC Model and Pipeline Model

The basic elements of an EPC model are shown in figure 1 (on the left): The control flow is symbolized by a sequence of events (magenta hexagons) and functions (green rectangles with rounded edges) which are connected by arrows representing the control flow. Branches in the control flow are defined by the Boolean logic operators: AND, OR and XOR. EPC models may be applied to express requirements of systems, specifically regarding the temporal behavior and they are also used in the design phase in order to meet the requirements. EPC models themselves may also be issue of verification [FSRK05].

The proprietary Intershop *Pipeline Models* (c.f. figure 1 on the right) describe actions of an e-commerce system and the paths linking these actions. The actions are represented by *Pipelets* which are the rounded rectangles. *Pipelets* consist of a declaration (in XML) and a definition in Java code which is then executed at run-time. This makes the model executable. The links between the *Pipelets* are the control flow represented by arrows and branching elements (as depicted in figure 1) or loop elements. The *Pipelines* are specifically used for web-based systems which mean that the *Pipelines* describe the activities between web pages presented to the clients. The *Pipelines* model and realize the interaction between two web pages. Therefore, the *Pipeline Models* provide modeling elements (Start element at the top of the *Pipeline Model*) for an ingoing interaction initiated by a clients action on a web page (e.g. activating a link or initiating a login action) and the outgoing interaction (Template Interaction, the two model elements at the bottom) which the feed the web page presenting the *Pipelines* results (in case of the example in figure 1: successful or unsuccessful login).

3 Formalization and Checking of Pipelines

In general the transformation of the Intershop *Pipeline Models* may be transformed straight forward into automata representation. Figure 2 shows these two formats. The example used is a price alert function. The function checks if a specific price goes beyond a defined threshold. If this condition is true (`PriceAlert = true` and additionally `PriceError = false`) then a mail is sent. This checking procedure is usually performed in a loop. In case the price doesn't change this is considered as `PriceError`

which means that the condition `PriceError = true` is set. This transformation into an automaton representation allows to use checking technology to verify automatically the model. The technology we have chosen is model checking [CGP01]. Model checking uses temporal logic in order to express the temporal sequence of states (the *Pipeline Model* elements are transformed to these states). Temporal logic provides besides the known Boolean logic operators temporal operators to express the temporal order. In our case we use the *Computational Tree Logic* (CTL) [BBF+01].

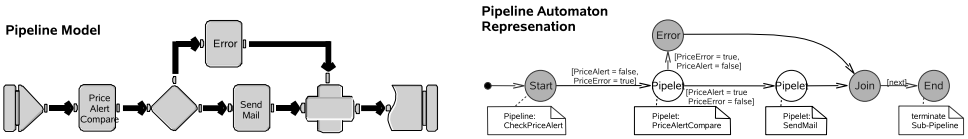


Figure 2: Example *Pipeline Price Alert* as *Pipeline Model* and Automaton Representation

The functionality of the *Pipeline Price Alert* from above may be expressed with CTL as follows: $A\,[(PriceAlert = \bot)\ W\ SendMail]$.

The formula expresses that always as long as the *Price Alert* is not true the *Pipelet SendMail* will not be sent.

4 Modeling and Checking of Real Systems

The examples up to now are quite small. Real e-commerce systems derived from the Intershop Enfinity base consist usually of several thousand *Pipelines* and *Sub-Pipelines* which are build of ten to hundred thousands of *Pipelets*.

Nevertheless these systems realize the functionality modeled in business processes. It is highly desirable to find an aid for (semi-) automated checking of these systems.

In order to illustrate the number of *Pipelines* and *Pipelets* we still concentrate of a small aspect of an e-commerce system: the presentation of products. This is one of the main functionality and usually most present to the customers using a web shop system. Usually this presentation functionality is always the way to present products the customers got arbitrarily, as opening presentation or as a result of concrete search.

Figure 3 shows the sample presentation in an ordinary web browser (Microsoft Internet Explorer in this case). Actually this is not a real web shop but a research system to examine the functionality.

The automaton in figure 3 consists of more than 50 nodes. Moreover, not all *Sub-Pipelines* are considered in this automaton. What is considered additionally is to the control flow of the web page template. This template is transformed by a specific processor to a simple HTML-page according to the number of the products to be displayed. Actually the automaton of figure 3 is only a sub-automaton of the large automaton representing the entire system. When we have a look at this sub-automaton it becomes clear that the checking of

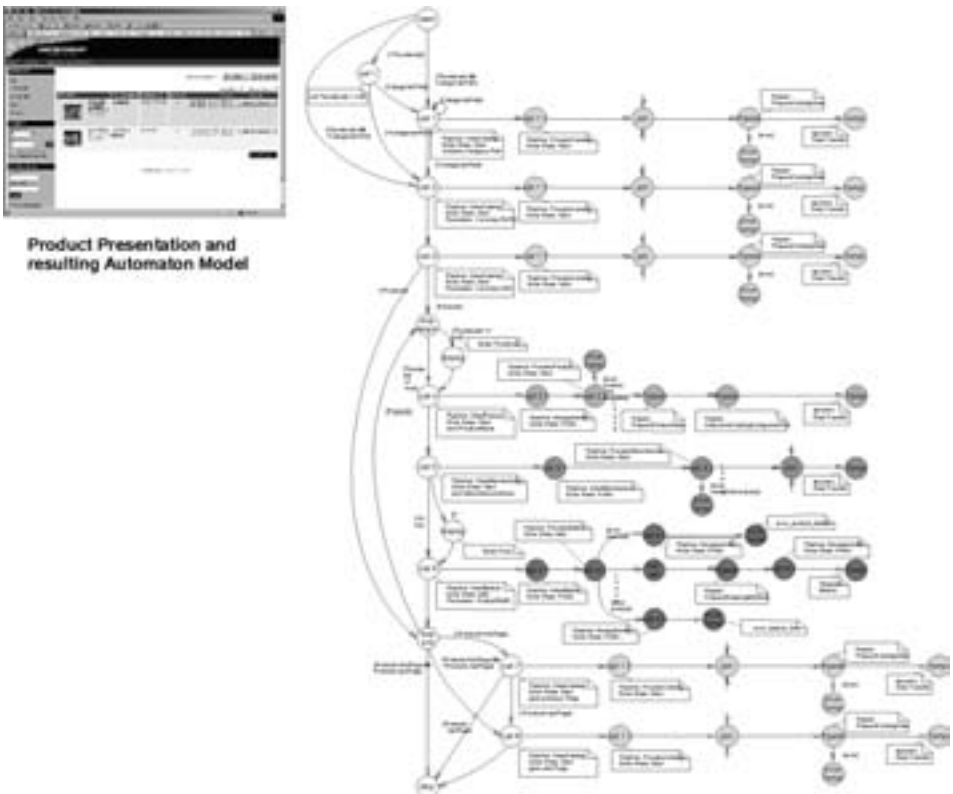


Figure 3: Product Presentation Web Page of Intershop Enfinity

such an automaton by hand is not that easy than the first examples. And if we take into account the number of several ten-thousand nodes of the complete automaton of a (still small) realization of a web shop system the need for automated support for the checking of such a system is obvious.

5 Related Work

The verification approach we use is based on a push-button formal technique to verify the fulfillment of automata-based specifications which is model checking [CGP01]. It uses (temporal) logics as specification language and checks these (temporal) requirements against a model. As opposed to other formal approaches (e.g. theorem provers) it is more restricted in what can be verified leading, on the other hand, to the advantage of a developer-friendly extensive automatization. The verification of software models has been a key issue in the model checking research for a long time (as in [EC80] or [McM93]).

In [FPR06] the verification of workflow-based systems is presented. This paper intro-

duces some of the basic principles we rely on in our paper. Further research on the verification of business process systems is the evaluation of different technologies in the area of model checking and other formal techniques in order to solve specific problems. Such problem-specific solutions have been elaborated. A further alternative approach is to apply constraint satisfaction presented in [Run09]. This paper concentrates on the configuration during design time. The rules for the checking and specifically the reuse of these rules are important issues. [Ger04] discusses the possibilities for reusing the specification rules.

An overview about the application of the proposed checking technology may be found at [FSRK05]. Here a first approach is presented. However, this approach focuses on the verification of the business process models (EPCs) itself.

Other approaches for the verification of business process systems are based on Petri nets (e.g. [DBS08] using BPMN). With Petri nets the business processes are mapped to the Petri net elements similar to our approach of direct mapping. In the Petri net based approaches the verification is often based on bi-simulation and algebraic solutions (e.g. [Mor08]).

6 Conclusion and Future Work

The presented approach allows verifying executable workflow models. As example we use the proprietary Intershop *Pipeline Models*. The rules applied are derived from best practices or business rules as they exist for almost any commercial system. Moreover, these rules for the executable model may be the optimized business processes themselves.

The paper demonstrates a straight-forward approach to transform executable workflow models (Intershop *Pipeline Models*) into automata which are the base of a formal verification. As verification technology we apply model checking. Our proposal allows checking comparatively large models. Comparatively means large from a human perspective. For the model checking tool the number of states is no real problem as we do not express critical items like numbers in the state space. The automated checking is provides the advantages of any automated process: It reduces the efforts for the human quality assurance teams while the number of errors is decreased.

The current state of the approach still uses the temporal logic CTL. Although we empirically learned that the application of this kind of logic seems to be very easy especially for quality assurance personal this will be issue of an improvement by presenting a more intuitive graphical notation. We may take the example of [FF08]. Here a graphical representation of CTL is introduced.

As we already stated the straight-forward transform results in simple but still models of a size processable by a model checker. A more elaborated concept is the differentiation of the element types as proposed in [Pul09] which may be adapted to our work.

This reduces the state space of the checking and allows to get a better feed-back when we analyze the checking results.

References

[BBF+01] B. Bérard, M. Bidoit, A. Finkel, F. Laroussinie, A. Petit, L. Petrucci, and P. Schnoebelen. *Systems and Software Verification – Model-Checking Techniques and Tools*. Springer, Berlin, Germany, 2001.

[Bre02] M. Breitling. Business Consulting, Service Packages & Benefits. Technical report, Intershop Customer Services, Jena, 2002.

[CGP01] E. M. Clarke, O. Grumberg, and D. A. Peled. *Model Checking*. The MIT Press, Cambridge, Massachusetts; London, England, 3 edition, 2001.

[DBS08] M. De Backer and M. Snoeck. Business Process Verification: a Petri Net Approach. Technical report, Catholic University of Leuven, Belgium, 2008.

[EC80] E. A. Emerson and E. M. Clarke. Characterizing Correctness Properties of Parallel Programs Using Fixpoints. In *ICALP 1980, Automata, Languages and Programming, 7th Colloquium*, pages 169–181. Springer LNCS 85, 1980.

[FF08] S. Feja and D. Fötsch. Model Checking with Graphical Validation Rules. In *15th IEEE International Conference on the Engineering of Computer-Based Systems (ECBS 2008), Belfast, NI, GB*, pages 117–125. IEEE Computer Society, April 2008.

[FPR06] D. Fötsch, E. Pulvermüller, and W. Rossak. Modeling and Verifying Workflow-based Regulations. In *Proceedings of the Workshop on Regulations Modelling and their Validation & Verification (REMO2V)*. Namur University Press, June 2006.

[FSRK05] D. Fötsch, A. Speck, W. Rossak, and J. Krumbiegel. A Concept of Modelling and Validation of Web based Presentation Templates. In *17. Internationale Tagung Wirtschaftsinformatik 2005 (WI2005)*, pages 391–406. Physika Verlag, February 2005.

[Ger04] R. Gerrits. Business rules, can they be re-used? *Business Rules Journal*, 5, 2004.

[McM93] K. L. McMillan. *Symbolic Model Checking*. Kluwer Academic Publishers, 1993.

[Mor08] S. Morimoto. A Survey of Formal Verification for Business Process Modeling. In *ICCS 2008, 8th International Conference*, pages 514–522. Springer LNCS 5102, 2008.

[Pul09] E. Pulvermüller. Reducing the Gap between Verification Models and Software Development Models. In *The 8th International Conference on Software Methodologies, Tools and Techniques (SoMeT 2009)*, pages 297–313. IOS Press, 2009.

[Run09] W. Runte. Modelling and Solving Configuration Problems on Business Processes Using a Multi-Level Constraint Satisfaction Approach. In *The Young Researchers Workshop on Modeling and Management of Business Processes (YRW-MBP 2009)*, pages 237–238. GI LNI 147, 2009.

[Sch98] A.-W. Scheer. *ARIS - Modellierungsmethoden, Metamodelle, Awendungen.* Springer, Berlin, Germany, 1998.

Formalizing Dependencies in Business Process Models Using Constraint Satisfaction

Wolfgang Runte
Institute of Computer Science
University of Osnabrueck
D-49069 Osnabrueck, Germany
woru@informatik.uni-osnabrueck.de

Abstract: Compliance management tackles issues related to both modelling and enforcement of business constraints in enterprises. In the context of business process management [Sch00], we propose and describe the use of constraint satisfaction problems [Dec03] as a formal mean for representing these dependencies [Run09, REK09]. We describe different scenarios using constraint satisfaction for the modelling of business processes and propose different constraint categories to distinguish between them. In order to illustrate this we give an example for each of these constraint categories. In the following we discuss which category of constraints is useful modelling different problem classes. Furthermore we discuss in this paper in which cases it is adequate to model a dependency as a constraint in general.

References

[Dec03] Rina Dechter. *Constraint Processing*. The Morgan Kaufmann Series in Artificial Intelligence. Morgan Kaufmann Publishers, San Francisco, California, USA, 2003.

[REK09] Wolfgang Runte and Marwane El Kharbili. Constraint Checking for Business Process Management. In Stefan Fischer, Erik Maehle, and Rüdiger Reischuk, editors, *INFORMATIK 2009: Im Focus das Leben. Beiträge der 38. Jahrestagung der Gesellschaft für Informatik e.V. (GI), 28.9. - 2.10.2009, Lübeck, Proceedings (Workshop on Business Process Modeling and Realization, BPMR-GI 2009)*, number P-154 in GI-Edition - Lecture Notes in Informatics (LNI), pages 4093–4103, Bonn, Germany, September 2009. Gesellschaft für Informatik.

[Run09] Wolfgang Runte. Modelling and Solving Configuration Problems on Business Processes Using a Multi-Level Constraint Satisfaction Approach. In Witold Abramowicz, Leszek Maciaszek, Ryszard Kowalczyk, and Andreas Speck, editors, *Business Process, Service Computing and Intelligent Service Management, BPSC 2009, ISM 2009, YRW-MBP 2009, March 23-25, 2009, Leipzig, Germany, Proceedings (The Young Researchers Workshop on Modeling and Management of Business Processes)*, number P-147 in GI-Edition - Lecture Notes in Informatics (LNI), pages 237–238, Bonn, Germany, March 2009. Gesellschaft für Informatik.

[Sch00] August-Wilhelm Scheer. *ARIS - Business Process Modeling*. Springer, Berlin, Heidelberg, New York, 2000.

Workshop

Wirtschaftlichkeit und Wertbeitrag von Informationssystemen

Stefan Strecker und Martin Kütz

Wirtschaftlichkeit und Wertbeitrag von Informationssystemen (WWI 2010)

Vorwort

Die Entwicklung, Einführung, Nutzung und Wartung betrieblicher Informationssysteme zielt auf eine möglichst effektive Unterstützung betrieblicher Entscheidungen, Handlungen und Abläufe. Der Einsatz betrieblicher Informationssysteme sollte die Wirtschaftlichkeit des Handelns in Unternehmen verbessern und die Wettbewerbsfähigkeit von Unternehmen befördern. Die Konzeptualisierung betrieblicher Informationssysteme als Mensch-Aufgabe-Technik-Systeme erfordert eine methodische Unterstützung, die den Spezifika betrieblicher Informationssysteme in Bezug auf ihre Wirkungen im unternehmerischen Kontext Rechnung trägt. Dazu zählt u. a. die Analyse des umgebenden organisatorischen Handlungssystems, in dem menschliche und maschinelle Aufgabenträger einen Wirkungsverbund bilden. In diesem Zusammenhang verspricht die Orientierung an Services einerseits Transparenz- und Effizienzgewinne, stellt jedoch andererseits neue Anforderungen an die Analyse und Bewertung ökonomisch relevanter Wirkungen von Informationssystemen.

Vor diesem Hintergrund beschäftigen sich die Beiträge des ersten Workshops „Wirtschaftlichkeit und Wertbeitrag von Informationssystemen“ (WWI 2010) mit der Entwicklung und dem Einsatz von Verfahren, Methoden und allgemein Konzeptionen, die darauf gerichtet sind, die Wirtschaftlichkeit von Informationssystemen und ihren Beitrag zur Zielerreichung etwa zur Verbesserung der Service-Qualität von Unternehmen zu erheben und/oder zu befördern.

Ziel des Workshops ist es, den Austausch zwischen Wissenschaft und Praxis zum Thema Wirtschaftlichkeit und Wertbeitrag von Informationssystemen zu fördern und zu intensivieren. Die Beiträge richten sich an Forscherinnen und Forscher sowie an Praktiker, die an den Themen Erfolgsmaßstäbe für und Erfolgsmessung von Informationssystemen, IT-Projektportfoliomanagement, IT-Wirtschaftlichkeitsanalyse und IT-Kostenmanagement interessiert sind. Der Workshop wird von der Fachgruppe IT-Controlling im Fachbereich Wirtschaftsinformatik der Gesellschaft für Informatik e.V. ausgerichtet.

Insgesamt wurden zwölf Beiträge zur Begutachtung eingereicht. Der Entstehungshintergrund von zehn Beiträgen ist dem akademischen Umfeld zu zurechnen. Zwei Einreichungen stammten aus der Praxis. Für jede Einreichung wurden durch das Programmkomitee drei Gutachten erstellt. Die fünf im Folgenden veröffentlichten Beiträge wurden nach Begutachtung durch das Programmkomitee zur Veröffentlichung und Präsentation am 30.09.2010 angenommen (Annahmequote 42%).

Der Beitrag *Praxisrelevanz des Modells von DeLone und McLean zur Erfolgsmessung von Informationssystemen* von A. Gemlik, M. Neumann, J. Sprenger und M. H. Breitner (Leibniz Universität Hannover) untersucht in Experteninterviews die Praxisrelevanz des seit langem diskutierten DeLone/McLean-Modells.

Ch. Meier und D. Kundisch (Universität Paderborn) setzen sich in ihrem Beitrag *Project interactions in value based IT project portfolio management* mit Abhängigkeiten und Interdependenzen zwischen IT-Projekten aus der Sicht des IT-Projektportfoliomanagements auseinander.

In ihrem Beitrag *Wirtschaftlichkeitsanalyse einer ausgelagerten Anwendungsentwicklung* beschreiben R. Martignoni und J. Stimmer (pliXos GmbH, München) exemplarisch ein Vorgehen zur Wirtschaftlichkeitsberechnung bei Auslagerungsvorhaben.

D. Heise (Universität Duisburg-Essen) zeigt in seinem Beitrag *Potentiale der Unternehmensmodellierung für das IT-Kostenmanagement (Kurzversion)* auf, welche Anforderungen an eine methodische Unterstützung für das IT-Kostenmanagement zu stellen sind und wie diesen durch einen unternehmensmodellbasierten Ansatz zu begegnen ist. Zu diesem Beitrag liegt eine Langfassung in elektronischer Form vor.

Der Beitrag *Ein Modell zur Einflussanalyse von Änderungen in Entwicklungsprozessen* von R. Koppe, S. Häusler und A. Hahn (OFFIS) entwickelt einen Ansatz zur Modellierung von Wirkzusammenhängen als Erweiterung des Software & Systems Process Engineering Metamodel (SPEM) der OMG.

Wir möchten den Mitgliedern des Programmkomitees und den Autoren der eingereichten Beiträge für ihren persönlichen Einsatz, die sehr angenehme Zusammenarbeit und den reibungslosen terminlichen Ablauf danken.

Stefan Strecker, *stefan.strecker@uni-due.de*

Martin Kütz, *martin.kuetz@inf.hs-anhalt.de*

3. Juli 2010

Essen und Köthen

Programmkomitee

Prof. Dr. Ulrike Baumöl, FernUniversität Hagen

Prof. Dr. Andreas Gadatsch, HS Bonn-Rhein-Sieg

Dr. Rüdiger Hannig, RAGH Dr. Rüdiger Hannig

Prof. Dr. Martin Kütz, HS Anhalt / TESYCON GMBH

Dr. Stefan Strecker, Universität Duisburg-Essen

Dr. Falk Übernickel, Universität St. Gallen

Praxisrelevanz des Modells von DeLone und McLean zur Erfolgsmessung von Informationssystemen

Arkadius Gemlik, Markus Neumann, Jon Sprenger, Michael H. Breitner

Institut für Wirtschaftsinformatik
Leibniz Universität Hannover
Königsworther Platz 1
30167 Hannover
{gemlik | neumann | sprenger | breitner}@iwi.uni-hannover.de

Abstract: Weitgehend unbestritten besteht ein wertschöpfendes Potenzial von Informationssystemen. Die Messung und Bestimmung von Wirkungszusammenhängen gilt jedoch weiterhin als Herausforderung. Das Modell zur Erfolgsmessung von Informationssystemen von DeLone und McLean leistet dabei einen Beitrag, ist in der Forschung verbreitet und wird dort umfassend diskutiert. Die Relevanz für die Praxis ist dagegen nur unzureichend untersucht. Die Zielsetzung des Beitrags ist somit die Prüfung der Praxisrelevanz. Kritische Aspekte hinsichtlich der Wichtigkeit, Zugänglichkeit und Tauglichkeit des Modells für die Praxis werden identifiziert und dargestellt. Eine Praxisrelevanz des Modells kann dabei nur bedingt nachgewiesen werden.

1 Einleitung und Problemstellung

Die Frage nach dem Wertbeitrag von Informationssystemen (IS) wird seit den Anfängen der angloamerikanischen „Information Systems Research“ bzw. der deutschsprachigen Schwesterdisziplin der „Wirtschaftsinformatik“ intensiv diskutiert. Auf verschiedenen Ebenen (u. a. Ökonomie, Industrie, Unternehmen, individuelle Anwendung) wird versucht, einen kausalen Zusammenhang zwischen IT-Ausgaben und Performance herzustellen – mit heterogenen Ergebnissen (u. a. [PBM08]). Obwohl das wertschöpfende Potenzial von Informationssystemen nicht mehr bestritten wird, bleibt zu klären, wie, wann und warum sich dieses im konkreten Einsatz realisieren lässt [KG08]. Regelmäßige Berichte über die Unzufriedenheit der Unternehmen mit den Ergebnissen ihrer IT-Projekte (u. a. [GRS07] [SG09]) verdeutlichen die Dringlichkeit dieser Problematik.

Die Forschung im Bereich der „IS-Erfolgsmessung“ leistet einen Beitrag zur Identifikation von Faktoren und Zusammenhängen, die einen erfolgreichen Einsatz von IS im Unternehmen determinieren. Mit Hilfe des in der Forschung verbreiteten DeLone und McLean IS-Erfolgsmodells [DM92][DM03] (IS-Erfolgsmodell) sind wesentliche Fortschritte auf diesem Gebiet erzielt worden. Das Modell wird intensiv diskutiert, empirisch geprüft und erweitert. Die wissenschaftliche „Rigorosität“ des Modells wurde bereits umfassend analysiert [PDM08]. Die Fragestellung der „praktischen Relevanz“ ist dagegen bisher nur unzureichend untersucht worden [USR09][RV08]. Zu diesem Zweck

empfehlen [RV08] die Methodik des „Applicability Check“, der eine Untersuchung von Wichtigkeit, Zugänglichkeit und Tauglichkeit wissenschaftlicher Artefakte umfasst.

Ziel dieses Beitrags ist eine Prüfung der Praxisrelevanz des IS-Erfolgsmodells. Im Rahmen einer zweistufigen empirischen Exploration erfolgt im ersten Schritt eine Untersuchung des bisherigen praktischen Einsatzes des Modells. Im zweiten Schritt werden mögliche Barrieren für den praktischen Einsatz aus der Praxis erarbeitet.

2 Kurzvorstellung des IS-Erfolgsmodells

Im Jahr 1992 organisieren, vergleichen und integrieren DeLone und McLean die existierende Literatur zum IS-Erfolg zu einem multidimensionalen Modell [DM92]. Aufbauend auf Masons Modifikationen der Kommunikationstheorie [Ma78] stellen sie fest, dass sich die meisten Erfolgsmessungen in nur sechs Kategorien einordnen lassen, die keine unabhängigen Erfolgskriterien, sondern zusammenhängende und voneinander abhängige Variablen darstellen: *Informationsqualität*, *Systemqualität*, *Nutzung*, *Nutzerzufriedenheit*, *individueller Einfluss* und *organisatorischer Einfluss*.

Nach der Veröffentlichung des originären Modells schlagen viele Forscher Modifizierungen vor (u. a. [PWK95][Se97]). Zehn Jahre später ergänzen DeLone und McLean deshalb ihr originäres Modell um das Konstrukt der *Servicequalität*, welches die Notwendigkeit von Service und Support bei modernen Informationssystemen widerspiegelt [DM03]. Um die Einstellung der Nutzer zur Nutzung des Systems zu messen, wird das Modell zudem um die „*beabsichtigte Nutzung*“ erweitert. Weiterhin fassen sie den individuellen und organisatorischen Einfluss zu einem einfacheren „*Net Benefits*“-Konstrukt zusammen. Abbildung 1 zeigt das aktualisierte IS-Erfolgsmodell.

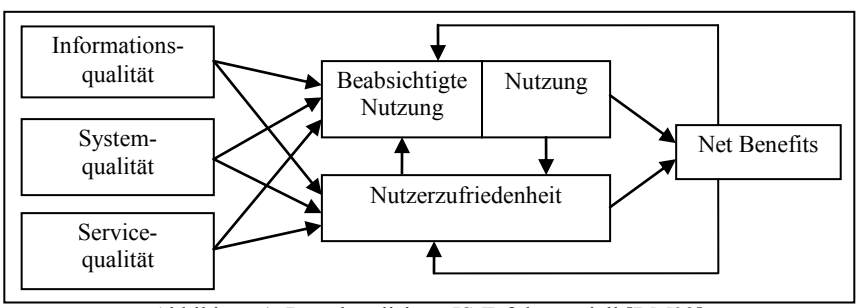


Abbildung 1: Das aktualisierte IS-Erfolgsmodell [DM03]

Das Originalmodell [DM92] ist das meistzitierte Modell in der IS-Forschung [GSC08]. Neben allgemeinen Herausforderungen der IS-Erfolgsmessung existiert spezifische Kritik am Modell wegen der unzureichenden Erklärung seiner theoretischen Fundierung und gemischter empirischer Untersuchungsergebnisse, die Bedenken hinsichtlich der Validität der vorgeschlagenen Verbindungen erzeugen [GSC08]. Gegenwärtig wird das Modell zudem in der Wirtschaftsinformatik intensiv diskutiert [Bu10].

3 Vorgehen der Analyse

Zur Überprüfung der praktischen Relevanz des IS-Erfolgsmodells ist aufgrund unzureichender wissenschaftlicher Ergebnisse [USR09][RV08] ein qualitatives Vorgehen gewählt worden. Bei der Planung und Durchführung der Exploration wurden allgemeine Richtlinien qualitativer Forschung (u. a. [KM99][Si05]) berücksichtigt. Die zweistufige Exploration (Tabelle 1) erfolgte im Zeitraum vom Juni 2009 bis März 2010 unter Verwendung des Instruments „Experteninterview" [GL09]. Für die Untersuchung wurden Experten aus verschiedenen Branchen und Tätigkeitsfeldern ausgewählt, die mindestens zehn Jahre Erfahrung in der IT-Praxis und darüber hinaus Kompetenzen im Bereich der IS-Erfolgsmessung aufweisen. Insgesamt erfolgten 50 Experteninterviews mit IT-Beratern (4), IT-Managern (24) und CIOs (22) weltweit tätiger, mittelständischer und großer Unternehmen.

Tabelle 1: Vorgehen der zweistufigen Exploration

Stufe	Zweck	Anzahl
1	Allgemeine Ermittlung der Verbreitung des IS-Erfolgsmodells in der Praxis	42
2	Spezifische Ermittlung der Praxisrelevanz des IS-Erfolgsmodells sowie der Barrieren einer Anwendbarkeit	8

Im Rahmen der ersten Stufe wurden 42 Experten allgemein hinsichtlich der Erfolgsmessung IT-basierter Investitionen befragt. Beabsichtigt war u. a. zu ermitteln, inwieweit das IS-Erfolgsmodell in der Praxis verwendet wird. Basierend auf den Ergebnissen dieser Befragung erfolgte in der zweiten Stufe eine spezifische, vertiefende Erhebung mit acht weiteren Experten hinsichtlich möglicher Barrieren der Anwendbarkeit des IS-Erfolgsmodells in der Praxis. Mittels einer qualitativen Inhaltsanalyse [GL09] sind die Interviewergebnisse zu den in dem folgenden Kapitel angeführten Aussagen verdichtet worden.

4 Ergebnisse der Analyse

Die Ergebnisse der ersten Stufe der empirischen Exploration lassen auf eine sehr geringe Bekanntheit des IS-Erfolgsmodells schließen. Die Schwerpunkte der derzeitigen Erfolgsmessung IT-basierter Investitionen lassen sich anhand von Investitionsphasen kategorisieren (Tabelle 2). Diese Darstellung deckt sich mit vorherigen Forschungsergebnissen, z. B. [SB08][Wa07].

Tabelle 2: Schwerpunkte der derzeitigen IS-Erfolgsmessung

Investitionsphasen	Schwerpunkte
Nutzen *identifizieren* und *planen*	Dominanz eindimensionaler Kostenziele, die nicht mit Projekt- und zugeordneten IS-Zielen verbunden sind
Nutzen *realisieren* und *evaluieren*	Vorrangige Evaluation von Zeit und Kosten bei der Projekt-Realisierung; Zwischenziele zur Nutzenrealisierung häufig nicht definiert und evaluiert; nachträgliche Erfolgsevaluation selten

In der *Identifizierungs*- und *Planungsphase* der Investitionen dominieren eindimensionale monetäre Erfolgs- bzw. Nutzenkennzahlen, die zur Investitionsbegründung herangezogen werden. Diese verlieren in der *Realisierungs*- und *Evaluierungsphase* an Bedeutung. In diesen Phasen wird lediglich der Projekterfolg im Sinne der drei Dimensionen Kosten, Zeit und Leistung beurteilt. Ein durchgehendes, mehrdimensionales Erfolgs- bzw. Nutzenmanagement, welches den Investitionsnutzen mit konkreten Zielsetzungen verbindet, findet nicht statt. Hierbei sehen die Experten eine der wichtigsten Verbesserungsmöglichkeiten. Dies gilt insbesondere für die Etablierung nachträglicher Evaluationen bzgl. der Nutzenrealisierung. Das IS-Erfolgsmodell ist der Mehrzahl der Experten (96 Prozent) nicht bekannt und wird folglich nicht als Hilfsmittel für die angestrebten Verbesserungen berücksichtigt. Wenige Experten (vier Prozent) wenden in der Praxis eine modifizierte Form des IS-Erfolgsmodells an. Der hohen Verbreitung des Modells in der Forschung steht eine geringe Bekanntheit in der Praxis entgegen. Dies deutet darauf hin, dass die *Zugänglichkeit* zum IS-Erfolgsmodell für Praktiker nicht gegeben ist.

Die Ergebnisse der zweiten Stufe der Exploration verdeutlichen, dass Barrieren bestehen. Die Experten betrachten Erkenntnisse zur IS-Erfolgsmessung allgemein als wichtig. Das IS-Erfolgsmodell liefert einen Beitrag zu diesem Themengebiet – folglich ist die *Wichtigkeit* gegeben. Probleme existieren dagegen bzgl. der *Tauglichkeit* des Modells für den praktischen Einsatz. So sind einige Experten an der Nutzung des Modells interessiert, merken aber an, dass es ein Modell sei, dass zwar in der Wissenschaft gut funktioniere, aber für die Praxis häufig nur bestimmte Kriterien unter Beachtung des Zusammenhanges übrig bleiben und statistische Auswertungen hinsichtlich der Kriterien aufgrund der Komplexität in der Praxis kaum umzusetzen sind. Des Weiteren berücksichtigt das IS-Erfolgsmodell die in der Praxis gängige Kategorisierung der Erfolgsmessung in einzelne Investitionsphasen nicht. Die kritisierten Aspekte der einzelnen Dimensionen des Modells (Barrieren einer Tauglichkeit) sind in Tabelle 3 zusammengefasst.

Tabelle 3: Tauglichkeit der Dimensionen des IS-Erfolgsmodells

Dimension	Barrieren
Informationsqualität	Nicht immer messbar/steuerbar
Systemqualität	Technisch gut messbar, aber im Gesamtzusammenhang nur untergeordnete Bedeutung; IS-Lebenszyklus nicht berücksichtigt
Servicequalität	Nicht immer auf die Art und Weise verstanden, wie von DeLone und McLean als Kriterium zur Messung des Service und Support vorgesehen, welcher den End-Nutzern des Systems geboten wird; Ausgestaltung der Lieferantenbeziehung nicht berücksichtigt
(Beabsichtigte) Nutzung	Messung von Systemzugriffen rechtlich nur bedingt zulässig
Nutzerzufriedenheit	Steht teilweise gar nicht mit dem IS in Bezug, da andere Personen, die Grundzufriedenheit einer Person oder die Absicht der Messung das Ergebnis verfälschen; Ergebnisse durch selektive Befragung manipulierbar; Objektivität dieser Erfolgsdimension fraglich, kaum glaubwürdig und daher nicht steuerbar
Net Benefits	Projekterfolg, Prozessbetrachtung, Spannungsfeld „Kosten versus Nutzen“, Total Cost of Ownership, Ausgestaltung der Business Requirements, des Kundenverhältnisses und strategischer Wettbewerbsvorteile nicht berücksichtigt oder nur unzureichend für die Anforderungen der Praxis dargestellt

Die Messung des IS-Erfolgs ist bis heute in der Praxis noch nicht zufriedenstellend adressiert. Das IS-Erfolgsmodell liefert zwar geeignete Dimensionen, die von den Experten für den IS-Erfolg bestätigt wurden, jedoch erfüllt es nur bedingt die Sichtweise der Praxis. Die meisten Experten setzen den Schwerpunkt bei der Erfolgsmessung auf den Projektverlauf. Dabei werden primär die Dimensionen Kosten, Zeit und Leistung berücksichtigt. Der eigentliche, nutzenstiftende Einsatz des Systems in den Geschäftsprozessen obliegt zumeist keiner Erfolgsmessung. Die Unzulänglichkeit dieser eingeengten Perspektive auf den IS-Erfolg wird jedoch auch von den Experten selbst wahrgenommen. Der IS-Erfolg dürfe ihrer Ansicht nach nicht isoliert betrachtet werden. Ihrer Meinung nach stellt der Projekterfolg keine umfassende Erfolgsdimension dar. Vielmehr müsse eine geschäftsprozessbezogene Top-Down-Perspektive aus Unternehmenssicht eingenommen werden, da die Bottom-Up-Betrachtung sehr schwierig sei. Einige Experten legen daher den Schwerpunkt auf das Erbringen des Service im Sinne einer „End-to-End-Betrachtung“, die ihrer Ansicht nach weitgehend im IS-Erfolgsmodell fehlt.

5 Zusammenfassung und Ausblick

Die Untersuchung diente generell dem Austausch zwischen Wissenschaft und Praxis. Dabei wurden relevante Konstrukte der IS-Erfolgsmessung mit den Experten diskutiert. Ziel war eine Prüfung der Praxisrelevanz des IS-Erfolgsmodells nach den Dimensionen *Wichtigkeit*, *Zugänglichkeit* und *Tauglichkeit*. Anhand der Ergebnisse der empirischen Untersuchung konnte in der Dimension Wichtigkeit bestätigt werden, dass die IS-Erfolgsmessung, und damit auch das IS-Erfolgsmodell, in der Praxis als wichtig angesehen wird. Dagegen wurden in den Dimensionen Zugänglichkeit und Tauglichkeit erhebliche Barrieren identifiziert. Letztlich konnte damit anhand der vorliegenden Untersuchung nur bedingt eine praktische Relevanz des Modells nachgewiesen werden. Zukünftige Forschung zum IS-Erfolg steht vor der Herausforderung, den Projekt-/Systemlebenszyklus, eine Prozessorientierung sowie verschiedene Projektkriterien in das Modell zu integrieren, um nach Ansicht der Experten eine höhere Praxisrelevanz zu erreichen. Des Weiteren ist die Verwendung des Konstrukts „Servicequalität“ im bisherigen Modell zu überprüfen (vgl. [GSC08]).

Der Beitrag generiert einen Mehrwert sowohl für die Theorie als auch für die Praxis, unterliegt aber dennoch einigen Limitationen. Aufgrund des explorativen Ansatzes und des damit einhergehenden hohen zeitlichen Aufwands wurden in der zweiten Phase lediglich acht Experten berücksichtigt. Weiterhin war wegen der generellen Komplexität wissenschaftlicher Artefakte keine tiefgehende Auseinandersetzung mit dem Modell möglich. So konnte bspw. keine umfassende Beurteilung bzgl. des Zusammenwirkens der einzelnen Konstrukte des Modells erfolgen. In Zukunft gilt es daher Wege zu finden, das Forschungsobjekt detaillierter und effektiver vorzustellen, ohne die Experten von einer Teilnahme abzuschrecken.

Literaturverzeichnis

[Bu10] Buhl, H.U.; Mertens, P.; Schumann, M.; Urbach, N.; Smolnik, S.; Riempp, G.: Leserbrief: Stellungnahme zum Beitrag von Urbach et al. aus Heft 4/2009. In: Wirtschaftsinformatik, 52(2), 2010; S. 109-114.

[DM92] DeLone, W.H.; McLean, E.R.: Information systems success: the quest for the dependent variable. In: Information Systems Research, 3(1), 1992; S. 60-95.

[DM03] DeLone, W.H.; McLean, E.R.: The DeLone and McLean Model of Information Systems Success: A Ten-Year Update. In: Journal of Management Information Systems, 19(4), 2003; S. 9-30.

[GL09] Gläser, J.; Laudel, G.: Experteninterviews und qualitative Inhaltsanalyse. 3., überarbeitete Auflage, VS Verlag für Sozialwissenschaften, Wiesbaden, 2009.

[GRS07] Gemino, A.C.; Reich, B.H.; Sauer, C.: Beyond chaos - Examining IT project performance. In: eProc. 2nd Int. Research Workshop on Information Technology Project Management (IRWITPM), Montreal, 2007; S. 29-38.

[GSC08] Gable, G.G.; Sedera, D.; Chan, T.: Re-conceptualizing information system success: the IS-Impact Measurement Model. In: Journal of the Association for Information Systems, 9(7), 2008; S. 377-408.

[KG08] Kohli, R.; Grover, V.: Business Value of IT: An Essay on Expanding Research Directions to Keep up with the Times. In: Journal of the Association for Information Systems, 9(1), 2008; S. 23-39.

[KM99] Klein, H.K.; Myers, M.D.: A Set of Principles for Conducting and Evaluating Interpretive Field Studies in Information Systems. In: MIS Quarterly, 23(1), 1999; S. 67-94.

[Ma78] Mason, R.O.: Measuring information output: a communication systems approach. In: Information and Management, Band 1(4), 1978; S. 219-234.

[PBM08] Pare, G.; Bourdeau, S.; Marsan, J.; Nach, H; Shuraida, S.: Re-examining the causal structure of information technology impact research. In: European Journal of Information Systems, 17(4), 2008; S. 403-416.

[PDM08] Petter, S.; DeLone, W.; McLean, E.: Measuring Information Systems Success: Models, Dimensions, Measures, and Interrelationships. In: European Journal of Information Systems, 17(3), 2008; S. 236-263.

[PWK95] Pitt, L.F.; Watson, R.T.; Kavan, C.B.: Service quality: a measure of information systems effectiveness. In: MIS quarterly, 19(2), 1995; S. 173-187.

[RV08] Rosemann, M.; Vessey, I.: Toward improving the relevance of information systems research to practice: The role of applicability checks. In: MIS Quarterly, 32(1), 2008; S. 1-22.

[SB08] Schwabe, G.; Bänninger, P.: IT-Benefits-Management in the Swiss Financial Sector. In: Proc. 41st Hawaii Int. Conf. on System Sciences, Hawaii, USA, 2008.

[Se97] Seddon, P.B.: A respecification and extension of the DeLone and McLean model of IS success. In: Information Systems Research, 8(3), 1997; S. 240-253.

[SG09] Standish Group Int.: The Chaos Report 2009, (www.standishgroup.com).

[Si05] Silverman, D.: Doing Qualitative Research. Sage Publications, London, 2005.

[USR09] Urbach, N.; Smolnik, S.; Riempp, G.: Der Stand der Forschung zur Erfolgsmessung von Informationssystemen - Eine Analyse vorhandener mehrdimensionaler Ansätze. In: Wirtschaftsinformatik, 51(4); S. 363-375.

[Wa07] Ward, J.; De Hertogh, S.; Viaene, S.: Managing Benefits from IS/IT Investments: an Empirical Investigation into Current Praxis. In: Proc. 40st Hawaii Int. Conf. on System Sciences, Hawaii, USA, 2007.

Project interactions in value based IT project portfolio management

Christian Meier, Dennis Kundisch

Lehrstuhl für Wirtschaftsinformatik
Information Management & E-Finance
Universität Paderborn
Warburger Straße 100
D-33098 Paderborn
Christian.Meier@wiwi.uni-paderborn.de
Dennis.Kundisch@wiwi.uni-paderborn.de

Abstract: Adequately considering interactions among IT projects in the process of constructing an IT project portfolio is a necessary condition in value-based IT project portfolio management (PPM). A lot of articles already deal with such interactions, but the literature lacks a common terminology and a structured perspective on the manifold types of interactions and their effects. In this article we present a framework that provides a structured perspective on interactions and thereby supports decision makers in the identification of possible interactions and the selection of appropriate methods for value-based IT PPM.

1 Introduction

For the value-based management of IT projects it is necessary that, along with other requirements [ZI08], IT project portfolio management (IT PPM) has to account for interactions among IT projects. We speak of an *interaction*, if resources consumed or outputs generated by a project influence the use of resources or outputs generated by one or several other projects. If, for example, the same piece of hardware is needed in more than one project and each project only temporarily needs this hardware, this certain piece of hardware may be shared between the projects and thus has to be procured only once. In the literature many state-of-the-art approaches can be found that consider interactions to some extent. Nevertheless, the literature lacks a framework that provides a structure to identify the manifold types of interactions and their effects on the value of the overall portfolio. In this article, based on a literature review and conceptual considerations, we identify possible types of interactions among IT projects and propose a framework to structure these types of interactions and the effects they inhere on the value of IT project portfolios. This framework can help decision makers to facilitate a more general understanding of project interactions and to identify appropriate approaches for value-based IT PPM from the literature.

2 Project interactions in the literature

While there seems to be consensus that the consideration of interactions is an important issue in the selection process of project portfolios, the perspective on project interactions and the degree of detail in which they are considered vary greatly in the literature. In a literature review, we examined the databases of Ebsco, Google Scholar, Science Direct, and Informs for journals featuring articles which contain the keywords "project selection" or "portfolio selection" in combination with "interaction" or "interdependency". We identified *Management Science*, *European Journal of Operational Research*, *Decision Sciences, Journal of Management Information Systems* and *International Journal of Project Management* to be highly relevant outlets for the addressed topic. Subsequently, we focused our literature review on these journals. From the resulting 230 articles of these journals we were able to exclude 159 by an abstract analysis, because they do not address the subject of our research sufficiently. In the following, we present an excerpt of 12 ([GE07], [FO84], [NE86], [DO06], [LI08], [SA96], [ME07], [BA04], [KL02], [MA94], [EI06], [LE01]) of – what we feel are – key-contributions from the remaining 71 articles. Many of the presented approaches consider certain types of project interactions in a detailed way, but it seems that there exists little consistency about the terminologies used and many of the interactions considered result from the special fields of applications for which the approaches have been developed. To our best knowledge no contribution exists, which classifies the types of interactions discussed in the literature and which provides a unified terminology and structure for project interactions. This will be the focus of the next section. To unify the terminology in this paper, in the following we will consistently refer to relationships or (inter-) dependencies among projects as *interactions*. We further denote the economic impact an interaction causes as *interaction effect*. If an interaction limits the space of candidate solutions for the portfolio we denote this impact *constraint effect*.

3 Structuring interactions in IT project portfolio management

IT projects can be seen as a transformation process in which certain inputs are transformed into outputs. The inputs – or resources (including e.g. technologies, workforce, and equipment) – needed to conduct a project in general induce monetary costs, whereas the outputs produced can be interpreted as services (e.g. infrastructure services, a webshop functionality, a new reporting system) that can deliver direct monetary benefits (by e.g. selling them), indirect benefits (by e.g. granting competitiveness or improving business process efficiency), or provide a basis and become resources for other projects. Therefore, we distinguish between the *transformation* level and the *economic effect* level (see figure 3.1).
In the following, we present a framework which provides a structured perspective on interactions among IT projects, and which classifies the types of interactions identified in the literature with respect to the transformation level and the economic effect level. Interactions and constraint effects can only occur on the transformation level, whereas interaction effects purely take effect on the economic effect level. Among resources and outputs, three types of interactions can occur.

Resource-Resource interactions arise solely between the resources, whereas *Output-Output* interactions occur just between the projects outputs. *Output-Resource* interactions occur among outputs and resources. To keep it simple in a first step, on the economic effect level we distinguish just between (monetary) costs and (monetary) benefits. Further, we assume that all parameters of interest (e.g. benefits resulting from a project's output) are deterministic and known at the time the portfolio is planned. So called *intertemporal* interactions – interactions that may influence the decision-making today based on potential follow-up projects in future phases of project portfolio planning – are not considered in the first step. Thus, in this article we focus on interactions that just affect the planning decision of the actual portfolio, which are referred to as *intratemporal* interactions. The consideration of uncertainties in resource availability, project outputs, benefits, costs, and interaction effects are as well subject to further research.

In the following, along with a description for each of the different types of interactions, we provide a short example and discuss the specific forms this particular type of interaction can adopt, as well as the effect this interaction is expected to have. Further, we provide an assignment of the contributions found in the literature to the different types of interactions, if the considered interaction in a contribution in substance corresponds to our definition.

3.1 Resource-Resource interactions

a. *Competitive resource utilization interactions*
   Description: Projects require the same resource and therefore the amount of resource required for the joint implementation of the related projects is *greater* than the sum of the resources required if the projects would have been implemented separately.
   Example: A staff member shared among different projects may need some time to mentally switch between the projects. This may result in set-up costs which could have been saved if the staff member would only be employed in one project at a time.
   Forms of appearance: This interaction affects all related projects in some way, which we denote as *symmetric*.
   Interaction effect: Costs increase. Due to diseconomies of scale in the resource utilization, additional resources may have to be procured to conduct the related projects.
   Considered by: [FO84], [GE07], [SA96].

b. *Complementary resource utilization interactions*
   Description: Projects require the same resource and therefore the amount of resource required for the joint implementation of the related projects is *less* than the sum of the resources required if the projects would have been implemented separately.
   Example: A staff member shared among different projects may benefit from his knowledge of a specific programming language required in more than one project. This may reduce the effort and working time (e.g. due to learning effects) needed by this member.
   Forms of appearance: This interaction affects all related projects in some way, which we denote as *symmetric*.

Interaction effect: Costs decrease due to economies of scale.
Considered by: [FO84], [KL02], [SA96], [ME07], [MA94], [NE86], [EI06], [LE01] .

Apparently, the types of interaction presented above result in similar interaction effects, merely affecting the costs in different directions (positive and negative). For further modeling purposes they can be subsumed by the term *resource utilization interactions*.

3.2 Output-Output interactions

a. *Competitive output interactions*
Description: The benefit of the outputs generated through the joint implementation of related projects is *smaller* than the benefit of the outputs generated if the projects would have been implemented separately.
Example: The benefit of the implementation of an Enterprise Resource Planning (ERP) system will be substantially diminished by the parallel implementation of a second ERP system.
Forms of appearance: Can be either *symmetric*, so that all projects in this relationship are affected, or *asymmetric*, so that a project influences other projects, but is not influenced by the other projects itself. As a special symmetric form of this interaction, projects can become *mutual exclusive*[1] (as in the example above).
Interaction effect: Benefits decrease (in the symmetric or asymmetric case).
Constraint effect: Restricts the solution space (for the mutual exclusive case)
Considered by: [ME07], [DO06], [LI07], [FO84], [SA96], [ME07], [NE86], [EI06].

b. *Complementary output interactions*
Description: The benefit of the outputs generated through the joint implementation of related projects is *greater* than the benefit of the outputs generated if the projects would have been implemented separately.
Example: The joint development of a portable multimedia player and a compatible software-tool to administrate and arrange music and playlists for that device can have synergetic effects on sales of both of the products.
Forms of appearance: Can be either *symmetric*, so that all projects in this relationship are affected (as in the example above), or *asymmetric*, so that a project influences other projects, but is not influenced by other projects itself.
Interaction effect: Benefits increase due to economies of scope.
Considered by: [ME07], [KL02], [DO06], [FO84], [SA96], [NE86], [EI06], [LE01].

3.3 Output-Resource interactions

a. *Binary contingency interaction*
Description: A project cannot stand alone and requires the outputs of other projects as mandatory resources.
Example: The implementation of an ERP system may require the installation of computer hardware to be completed, whereas the hardware can be installed without the ERP system.

[1] The simultaneous conduction of two or more interrelated projects may lead to a situation where the projects technically could be conducted in parallel, but become "economically" mutual exclusive. For modeling purposes it still seems favorable to consider this being a constraint effect.

Forms of appearance: Is *asymmetric*, so that a project's output is required as a mandatory resource by other projects, but is not influenced by other projects itself.
Constraint effect: Necessitates the selection of distinct projects if related projects are selected.
Considered by: [KL02],[SA96].

b. *Continuous competitive contingency interactions*
Description: An influenced project may stand alone, but the outputs of related projects deteriorate the resource requirements/utilization of the influenced project.
Example: A project implements new reporting guidelines for projects resulting in increased reporting efforts per project and thereby reduced available working time for project team members.
Forms of appearance: Is *asymmetric*, so that a project influences other projects, but is not influenced by the other projects itself.
Interaction effect: Costs increase.
Considered by: [BA04], [FO84], [NE86], [LE01].

c. *Continuous complementary contingency interaction*
Description: An influenced project may stand alone, but the outputs of projects with interactions to the influenced project improve the resource requirements/utilization of the influenced project.
Example: A project implements new reporting guidelines for projects that provide more transparency in the staffing of projects. This results in a more cost efficient assignment of team members to projects.
Forms of appearance: Is *asymmetric*, so that a project influences other projects, but is not influenced by the other projects itself.
Interaction effect: Costs decrease.
Considered by: [BA04], [FO84], [NE86], [LE01].

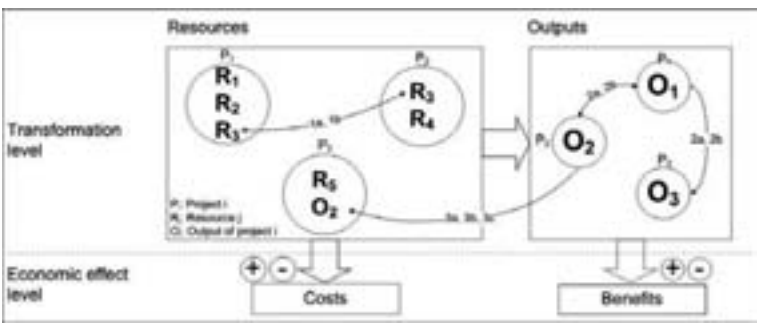


Fig. 3.1: Interactions and their effects in IT portfolios

3.4 Degree of considered interactions

For the assessment of the different types of interactions in our framework, the number of involved projects has to be considered. Each type of interaction basically can be assessed in a "bottom-up" approach for two projects at a time and for subsets of more than two projects up to the whole number of project proposals. Here, an appropriate trade-off between estimation effort and the desired degree of accuracy has to be found.

4. Summary and Outlook

Based on a literature review we identified the need for a unified terminology and structure of interactions among IT projects. We presented a framework that structures these interactions and thereby provides valuable insights for the decision makers to select appropriate approaches for value-based IT PPM from the literature. At this time the framework only accounts for deterministic parameters and does not consider uncertainty and the potential for risk diversification, which will be the subject of further research. In addition, the detailed assessment of each of the identified interactions can become very expensive. Therefore, the question of which interactions have to be assessed in greater detail and which can be neglected for specific project portfolio selection problems should be also addressed in prospective research. Finally, temporal considerations have to be included into the framework.

Literaturverzeichnis

[BA04] Bardhan, I.; Bagchi, S.; Sougstad, R.: Prioritizing a Portfolio of Information Technology Investment Projects. Journal of Management Information Systems, 2004, 21:33-60.

[DO06] Doerner, K. F.; Gutjahr, W. J.; Hartl, R. F.; Strauss, C.; Stummer, C.: Pareto ant colony optimization with ILP preprocessing in multiobjective project portfolio selection. European Journal of Operational Research, 2006, 171:830-841.

[EI06] Eilat, H.; Golany, B.; Shtub, A.: Constructing and evaluating balanced portfolios of R&D projects with interactions: A DEA based methodology. European Journal of Operational Research, 2006, 172:1018-1039.

[FO84] Fox, G. E.; Baker, N. R.; Bryant, J. L.: Economic Models for R and D Project Selection in the Presence of Project Interactions. Management Science, 1984, 30:890-902.

[GE07] Gear, T. E.; Cowie, G. C.: A note on modeling project interdependencies in research and development. Decision Sciences, 2007, 11:738-748.

[KL02] Klapka, J.; Pinos, P.: Decision support system for multicriterial R&D and information systems projects selection. European Journal of Operational Research, 2002, 140:434-446.

[LE01] Lee, J. W.; Kim, S. H.: An integrated approach for interdependent information system project selection. International Journal of Project Management, 2001, 19:111-118.

[LI08] Liesio, J.; Mild, P.; Salo, A.: Robust portfolio modeling with incomplete cost information and project interdependencies. European Journal of Operational Research, 2007, 190:679-695.

[MA94] De Maio, A.; Verganti, R.; Corso, M.: A multi-project management framework for new product development. European Journal of Operational Research, 1994, 78:178-191.

[ME07] Medaglia A. L.; Graves, S. B.; Ringuest, J. L.: A multiobjective evolutionary approach for linearly constrained project selection under uncertainty. European Journal of Operational Research, 2007, 179:869-894.

[NE86] Nelson, C. A.: A scoring model for flexible manufacturing systems project selection. European Journal of Operational Research, 1986, 24:346-359.

[SA96] Santhanam, R.; Kyparisis, G. J.: A decision model for interdependent information system project selection. European Journal of Operational Research, 1996, 89:380-399.

[ZI08] Zimmermann, S.: IT-Portfoliomanagement: Ein Konzept zur Bewertung und Gestaltung von IT. Informatik Spektrum, 2008, 31:460-468.

Wirtschaftlichkeitsanalyse einer ausgelagerten Anwendungsentwicklung

Robert Martignoni, Jörg Stimmer

pliXos GmbH
Agnes-Pockels-Bogen 1
80992 München
robert.martignoni@plixos.com
joerg.stimmer@plixos.com

Abstract: Die Wettbewerbsfähigkeit von Unternehmen hängt zunehmend von der Leistungsfähigkeit der IT ab. Die IT eines Unternehmens unterliegt, ebenso wie alle anderen Unternehmensbereiche, Wirtschaftlichkeitsanforderungen. Die Wirtschaftlichkeitsberechnung von Auslagerungsvorhaben ist sehr komplex, da sowohl unternehmensinterne als auch externe Faktoren berücksichtigt werden müssen. Kernpunkt dieses Beitrages ist die Beschreibung eines exemplarischen Vorgehens zur Wirtschaftlichkeitsberechnung.

1 Hintergrund

Die Basis einer jeden Auslagerungsentscheidung in der IT sollte eine Wirtschaftlichkeitsberechnung sein. Deshalb ist die Wirtschaftlichkeitsbetrachtung auch ein wichtiger Teil der DIN SPEC 1041 (Outsourcing technologieorientierter wissensintensiver Dienstleistungen) [1]. Die Praxis sieht anders aus. Laut einer Studie der Warwick Business School haben weniger als die Hälfte aller befragten CIOs versucht, die Wirtschaftlichkeit von IT Outsourcing zu bewerten. Nur ein Fünftel der CIOs, die überhaupt versucht haben die Vorteile zu quantifizieren, vertrauen dem Ergebnis [2]. Trotz der Tatsache, dass die Kosten und damit die Wirtschaftlichkeit Hauptmotivation einer Auslagerungsentscheidung ist, ist die Berücksichtigung der Kostenentwicklung eines Auslagerungsvorhabens schwierig [3]. Vor allem, da eine Abschätzung der Kosten ex ante erfolgen muss [4]. Des Weiteren kann die Wirtschaftlichkeit im Nachhinein nur schwerlich berechnet werden, da nur Daten für die gewählte Umsetzungsoption vorliegen.

Dieser Beitrag beschreibt exemplarisch die Herleitung einer Wirtschaftlichkeitsberechnung für Auslagerungsprojekte in der IT. Ziel ist ein Modell zur Berechnung der Wirtschaftlichkeit von IT Outsourcing Projekten.

Die Modellannahmen basieren auf ex post Analysen von in der Praxis durchgeführten IT Projekten und einer Literaturanalyse mit Fokus auf Diskussionen versteckter Kosten von IT Outsourcing (z.B. [5][6][7]). In einigen Beiträgen wird die Höhe dieser versteckten Kosten mit bis zu 65% des Auftragsvolumens angegeben [8]. Verdeckte Kosten sind z.B. administrative Kosten zur Auswahl und Steuerung des Dienstleisters, Kosten für Mitarbeiterabbau, Produktivitätsunterschiede. Das Modell beschränkt sich auf quantifizierbare Faktoren – qualitative Minderleistungen, Wissensverlust, und weitere Aspekte können nicht quantifiziert werden und finden daher keinen Eingang in die Wirtschaftlichkeitsberechnung. Das in diesem Beitrag diskutierte Kostenmodell berücksichtigt explizit diese Kosten.

2 Grundlagen der Wirtschaftlichkeitsberechnung

Die Berechnung der Wirtschaftlichkeit eines Projektes basiert auf dem Vergleich unterschiedlicher Szenarien. Im Fall der Wirtschaftlichkeitsberechnung eines Outsourcing Projektes muss ein Modell identifiziert werden, mit dem sich einerseits die Kosten für ein konventionelles internes Projekt berechnen lassen und andererseits auch die Kosten für die Auslagerung. Hierbei kann wiederum nach unterschiedlichen Auslagerungstypen unterschieden werden. Die Differenz der Ergebnisse der einzelnen Berechnungen zeigt dann die Einsparungen bzw. im ungünstige Fall auch die Mehrkosten für ein Outsourcing Vorhaben auf. Die Aussagekraft der Wirtschaftlichkeitsberechnung wird vor allem durch die Qualität des Kostenmodells und der Genauigkeit in der Schätzung der Einflussfaktoren bestimmt.

2.1 Kosten für die konventionelle Umsetzung

Die Kosten für die Umsetzung bestimmen sich maßgeblich durch das Projektvolumen. Das lässt sich am einfachsten durch die Anzahl der FTE (Full Time Equivalent bzw. Vollzeitäquivalent) bestimmen. In vielen Projekten sind Freie Mitarbeitern regulär beteiligt. Dieser signifikante Block muss daher dringend zusätzlich berücksichtigt werden, genau wie die Nebenkosten für die Projektdurchführung. Zusammengenommen sollte ein möglichst vollständiges Bild der internen Kostenstruktur entstehen, um eine sinnvolle Vergleichsbasis zur Verfügung zu haben.

2.2 Kosten für die externe Umsetzung (Auslagerung)

Die Berechnung der Kosten für die Auslagerung ist wesentlich komplexer als ein erster Blick vermuten lassen könnte. Dafür muss nicht nur die existierende Kostenstruktur identifiziert werden, sondern auch Erwartungen hinsichtlich zukünftiger Entwicklungen quantifiziert werden. Ein weiterer Aspekt der die Komplexität erhöht ist, dass auch weiterhin interne Kosten berücksichtigt werden müssen. Darüber hinaus sollen zumindestens die beiden Grundtypen des Outsourcing berücksichtigt werden: captive Outsourcing (eigener Betrieb eines ausgelagerten Standortes) und non-captive Outsourcing (Zusammenarbeit mit einem externen Dienstleister).

Als Basis zur strukturierten Identifikation relevanter Kosten wurde von Barthélemy vorgeschlagen, den Outsourcing Lebenszyklus als Basis zu nehmen [7]. Für die Ermittlung der Wirtschaftlichkeit wird deshalb folgendes Vorgehen bei Outsourcing Projekten als Basis genommen.

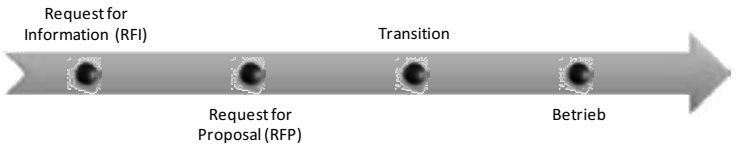


Abbildung 1 Vorgehen bei IT Outsourcing Projekten

In jeder dieser vier Phasen fallen unterschiedliche Kosten an, die bei der Wirtschaftlichkeitsberechnung berücksichtigt werden müssen. Eine Auswahl relevanter Kosten ist in der folgenden Tabelle aufgeführt:

RFI	RFP	Transition	Betrieb
• Interne Vorbereitung der RFI Phase	• Interne Vorbereitung der RFP Phase • Verhandlungskosten • Anwaltskosten	• Schulungskosten • Reisekosten • Abbau von Überkapazitäten	• Governance-Kosten • Reisekosten • Schulung

Tabelle 1 Auswahl der wichtigsten phasenabhängigen Kosten

Einige davon lassen sich sehr genau quantifizieren (da diese gesondert beauftragt werden bzw. sehr transparent sind), andere lassen sich eher schwer beurteilen.

2.3 Übersicht des Modells

Das Modell basiert auf der Berechnung und Allokation von Kostenblöcken. Die Kostenblöcke werden zuerst identifiziert, dann der entsprechenden Phase des Auslagerungsprojektes zugeordnet und zuletzt mathematisch modelliert.

Die folgende Tabelle zeigt eine Auswahl relevanter Kostenblöcke. Die Auswahl beschränkt sich auf die Faktoren, die der Erwartung nach den höchsten Einfluss auf das Ergebnis haben. Eine Auflistung und Erklärung aller Faktoren würde den Umfang dieses Beitrages sprengen.

Intern	Auslagerung	
	Captive	Non-captive
• Kosten für Mitarbeiter inkl. Nebenkosten • Arbeitsplatzkosten • Kostensteigerung	• Tagessatz eines Mitarbeiters • Aufbaukosten • Infrastrukturkosten • Kosten für Personalbeschaffung • Nebenkosten	• Tagessatz eines Mitarbeiters • Nebenkosten • Produktivität

Tabelle 2 Übersicht der Modellabhängigen Kosten (Auswahl)

Die Kosten für eine konventionelle (interne) Umsetzung wie auch für die beiden Outsourcing-Grundtypen werden vor allem durch die Aufwendungen für entsprechende Mitarbeiter beeinflusst. Diese werden wiederum indirekt durch die Produktivität und Arbeitsplatznebenkosten beeinflusst.

Alle Kostenblöcke werden in dem letzten Schritt mathematisch moduliert und in den Phasen des Projektablaufs dargestellt. Eine detaillierte Darstellung kann aufgrund des Umfangs des Modells an dieser Stelle nicht erfolgen.

3 Validierung der Wirtschaftlichkeitsberechnung

In der Literatur finden sich zahlreiche Methoden zur Verifzierung und Validierung von (Simulations-)-Modelle (siehe z.B. [9][10]). Die beste Validierungsmethode wäre ein Vergleich der Ergebnisse des Modells mit den in der Realität gemessenen Werten. Die Wirtschaftlichkeitsberechnung von Outsourcing Projekten beruht allerdings auf dem Vergleich verschiedener Umsetzungsszenarien. In der unternehmerischen Realität wird allerdings nur immer eine Option umgesetzt. Ein Vergleich verschiedener Umsetzungsoptionen kann daher nur theoretisch erfolgen, da in der Realität nur Werte für eine Option zu messen sind. Ein Vergleich verschiedener Projekte mit verschiedenen Umsetzungsszenarien, wäre aufgrund der fehlenden Vergleichbarkeit nicht sinnhaft.

Eine weitere, zuverlässige Methode ist die „Parameter Variability – Sensitivity Analysis“ [9]. Bei dieser Validierungsmethode werden die Eingangsparameter variiert und das veränderte Ergebnis beobachtet. Dadurch ist zwar das Modell nicht final validiert, aber das Vertrauen in das Modell kann erhöht werden. Am Beispiel ausgewählter Szenarien sollen nun die Ergebnisse der Wirtschaftlichkeitsberechnung aufgezeigt und validiert werden.

Die folgenden drei IT Outsourcing Projekte werden analysiert:

- Szenario 1: Ein langes Projekt (Laufzeit 3 Jahre) mit 80 FTEs zum Ende des Projektes. Gegenstand des Projektes ist eine Software-Individualentwicklung.

- Szenario 2: Ein kurzes Projekt (Laufzeit 2 Jahre) mit bis zu 50 FTEs im zweiten Jahr. Gegenstand des Projektes ist ebenfalls eine Software-Individualentwicklung.

- Szenario 3: Ein kurzes Projekt (Laufzeit 18 Monate) mit dem Ziel Standard-Komponenten zu integrieren, an dem durchschnittlich 25 FTEs beteiligt sind.

Für jedes dieser Szenarien wurde Schätzungen der Parameter basierend auf der langjährigen Erfahrung getroffen. Ein Teil der Parameter kann exakt angegeben werden (beispielsweise das Auftragsvolumen), während andere nur geschätzt werden können, wie z.B. die zukünftige Kostenentwicklung oder auch die Tagessätze, welche erst durch den Vertrag mit dem Dienstleister konkretisiert werden.

Um der Unsicherheit bei der Parameterschätzung gerecht zu werden, wurden dein Teil der Annahmen für die Parameter für jedes Basisszenario variiert (zufällig um bis zu +/- 20% des Basiswertes). Für jedes der drei Szenario wurden auf diesem Wege 2.000 Unterszenarien gebildet, die sich bezüglich 39 der insgesamt 70 Parameter unterschieden. Anschließend wurde für jedes der Unterszenarien die Kosteneinsparung der beiden Outsourcing-Modelle berechnet. Die Ergebnisse finden sich in dem folgenden Kapitel.

3.1 Interpretation der Ergebnisse

Die Varianz der Eingangsparameter führt zu einer entsprechenden Varianz der Höhe der Kosteneinsparung. Die folgende Abbildung zeigt die Varianz der Kosten sowie die Einsparungen verglichen mit einer konventionellen Umsetzung der einzelnen Szenarien.

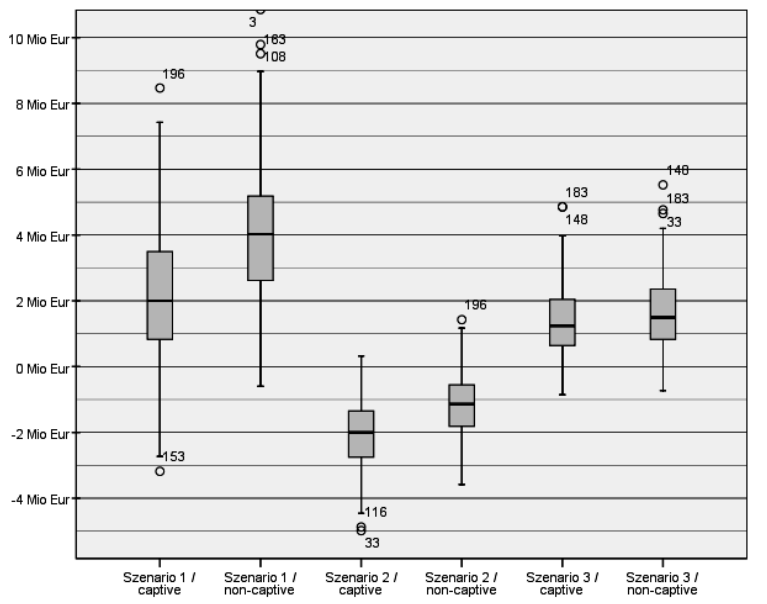


Abbildung 2 Varianz der berechneten Einsparungen für die 3 Szenarien

Das Ergebnis zeigt dass die Ergebnisse entsprechend der Varianz der Eingangsparameter in einem vertretbaren Bereich bewegen.

Die Ergebnisse der Wirtschaftlichkeitsuntersuchung zeigen, dass mit einer Auslagerung von Projekten signifikante Kosteneinsparungen möglich sind (Szenario 1 (non-captive) zeigt eine erwartete Kosteneinsparung von fast 40 %), durchaus aber auch Mehrkosten entstehen können (Szenario 2). Prinzipiell weisen die non-captive Modelle eine höhere Einsparung auf.

Im Sinne der Validierung zeigt das Modell ein nachvollziehbares Verhalten. Die Ergebnisse sind - trotz hoher Variabilität der Einflussfaktoren – relativ stabil. Nur wenige Ausreißer sind zu beobachten.

4 Zusammenfassung und Ausblick

Dieser Beitrag hat die Entwicklung eines Modells zur Berechnung der Wirtschaftlichkeit von Auslagerungsvorhaben beschrieben. Basis dieses Modells waren umfangreiche ex post Analyse von durchgeführten Projekten. Das resultierende Modell wurde abschließend validiert. Das Ergebnis waren konsistente Ergebnisse trotz sehr variabler Einflussfaktoren. Die Anwendung weiterer Validierungstechniken sollten folgen, und könnte das Vertrauen in das vorgeschlagene Modell noch weiter erhöhen.

Das wahrscheinlich wichtigere Ergebnis dieses Beitrages ist, dass Auslagerungsprojekte in der IT durchaus Kosteneinsparungspotenziale liefern können. Diese Analyse sollte vor Beginn des Projektes stehen, um so die Erwartungshaltung aller Stakeholder (inkl. des Dienstleisters) zu manifestieren, und darauf aufbauende Steuerungsmechanismen zu etablieren.

Literaturverzeichnis

[1] DIN SPEC 1041, *Outsourcing technologieorientierter wissensintensiver Dienstleistungen*, 2010.

[2] J. Kotlarsky, “Business leaders 'not sure' about contribution of outsourcing,” Apr. 2010.

[3] R. Gonzalez, J. Gasco, und J. Llopis, “Information systems outsourcing risks: a study of large firms,” *Industrial management & Data systems*, vol. 105, 2005, S. 45–62.

[4] A. Vining und others, “A conceptual framework for understanding the outsourcing decision,” *European Management Journal*, vol. 17, 1999, S. 645–654.

[5] R. Dué, “The real costs of Outsourcin,” *Information Systems Management*, vol. 9, 1992, S. 78-81.

[6] S. Overby, “The hidden costs of offshore outsourcing,” *CIO Magazine*, vol. 16, 2003, S. 60–66.

[7] J. Barthélemy, “The hidden costs of IT Outsourcing,” *MIT Sloan management review*, vol. 42, 2001, S. 60-69.

[8] G. Szatvanyi, “The Real Cost of Outsourcing,” *EzineArticles.com*, Sep. 2008.

[9] R.G. Sargent, “Verification and validation of simulation models,” *Proceedings of the 37th conference on Winter simulation*, Orlando, Florida, USA: 2005, S. 130 - 143.

[10] J.P. Kleijnen, “Verification and validation of simulation models,” *European Journal of Operational Research*, vol. 82, 1995, S. 145–162.

Potentiale der Unternehmensmodellierung für das IT-Kostenmanagement (Kurzversion)

David Heise
Lehrstuhl für Wirtschaftsinformatik und Unternehmensmodellierung
Universität Duisburg-Essen
david.heise@uni-duisburg-essen.de

Abstract: IT-Controller sehen sich gegenwärtig einem Paradigmenwechsel gegenüber – von einer nachlaufenden IT-Kostenkontrolle hin zu einem IT-Kostenmanagement, das einen erweiterten Kostenbegriff, eine aktive Kostensteuerung und verschiedene Perspektiven betont. In diesem Beitrag werden die Potentiale der Unternehmensmodellierung zur Unterstützung für ein IT-Kostenmanagement untersucht und Erweiterungen skizziert, die sich für eine aktive Kostensteuerung empfehlen. Eine entsprechend erweiterte Methode der Unternehmensmodellierung verspricht, eine höhere Transparenz über IT-Einsatz, damit verbundenen Kostenwirkungen und zugehörigen Gestaltungsoptionen zu schaffen sowie die Kommunikation zwischen betriebswirtschaftlichen und technischen Perspektiven zu verbessern.

1 Motivation

Die Sicherstellung der Wirtschaftlichkeit des laufenden IT-Betriebs bildet ein zentrales Aufgabenfeld des IT-Controllers in der Praxis [BFM06]. Im Vordergrund steht dabei der nachvollziehbare und differenzierte Ausweis des IT-Einsatzes, der damit assoziierten IT-Kosten sowie die Analyse und Bewertung der Einflussmöglichkeiten auf diese Kostenstrukturen. Die IT-Controller sehen sich bei der Umsetzung eines IT-Kostenmanagements allerdings der erheblichen Komplexität heutiger Unternehmens-IT gegenüber. Zum einen müssen sie der Komplexität begegnen, die aus der Vielfalt der Steuerungsobjekte wie Hardware, Software, Informationssysteme und IT-Services sowie ihren Abhängigkeiten untereinander entsteht. Zum anderen bedingt die enge Verflechtung von IT-Services, Geschäftsprozessen und verfolgten Zielen, Faktoren auf sämtlichen Unternehmensebenen – von der IT- über die Geschäftsprozess- zur strategischen Ebene – mit einzubeziehen und betrifft damit Akteure verschiedener Perspektiven wie IT-Experten, IT-Manager und Geschäftsprozessverantwortliche. Sprachbarrieren erschweren Kommunikation und Kooperation zwischen diesen Gruppen. Zugleich wird, nicht zuletzt im Zuge der Debatte um den Wertbeitrag der IT, zunehmend eine wert- und unternehmenszielorientierte Analyse und Gestaltung der IT gefordert [Bau08]. In der Folge ist das IT-Controlling dazu angehalten, für das IT-Kostenmanagement einen „erweiterten“ Kostenbegriff zu entwickeln, der diese Wertorientierung angemessen widerspiegelt. Vor diesem Hintergrund beleuchtet der Beitrag die Potentiale der Unternehmensmodellierung zur Unterstützung für ein IT-Kostenmanagement und skizziert Erweiterungen, die für eine aktive Kostensteuerung notwendig sind.

2 Anforderungen an eine Methode für das IT-Kostenmanagement

Das IT-Kostenmanagement ist eine Funktion des IT-Controllings und zielt auf die Analyse und Gestaltung der Kostenstrukturen der IT-Organisation. Im Gegensatz zur Kostenrechnung basiert das Kostenmanagement *nicht* auf der Prämisse gegebener Kostenstrukturen, sondern fokussiert deren aktive Beeinflussung [SK08]. Grundlage hierfür ist Transparenz in der Betrachtung des IT-Einsatzes. Eine methodische Unterstützung für das IT-Kostenmanagement sollte daher die Komplexität der Unternehmens-IT soweit reduzieren, dass die im jeweiligen Anwendungskontext notwendigen Informationen nachvollziehbar dargestellt werden. Von Relevanz sind dabei insbesondere die Kostenstrukturen der IT-Ressourcen sowie ihre jeweiligen kostenrelevanten Wirkungsbeziehungen zu den IT-Services und den Geschäftsprozessen. Die Betrachtung von Kostengrößen ist für Fragen des IT-Kostenmanagement jedoch keineswegs hinreichend: Zum einen sind Kosten für eine angemessene Interpretation und Bewertung zu ergänzen um den technischen und organisatorischen Kontext ihrer Bezugsobjekte, z. B. Kapazitäten, Bindungsdauern und Vertragsinformationen [May98]. Gefordert ist somit eine Möglichkeit einer komplexitätsreduzierenden, zugleich aber auch differenzierten Abbildung von IT-Ressourcen, ihren Beziehungen und Kostenwirkungen untereinander sowie zu den Geschäftsprozessen. Zum anderen ist, vor allem mit Blick auf die Wertorientierung, der IT-Einsatz stets vor dem Hintergrund des umgebenden Handlungssystems, z. B. Geschäftsprozessen, Zielen und Strategien, zu bewerten. Der Wertorientierung liegt dabei ein erweitertes Wirtschaftlichkeitsverständnis zu Grunde, dass neben monetären Größen auch nicht bzw. schwer quantifizierbare (z. B. „Risiken“) sowie positive Wirkungen („Nutzen“) des IT-Einsatzes einschließt [Str09]. Schließlich bedingt dies die Einbindung von Akteuren technischer und betriebswirtschaftlicher Perspektiven, die z. T. durch unterschiedliche Zielvorstellungen und Wahrnehmungen geprägt sind. Die Methode sollte unterschiedlichen Akteursgruppen zugänglich sein und eine Kommunikation zwischen ihnen fördern; um Zielkonflikte aufzudecken bietet es sich darüber hinaus an, die Annahmen der verschiedenen Beteiligten explizit zu machen. Tabelle 1 fasst die Anforderungen zusammen.

Existierende Methoden für das IT-Kostenmanagement (z. B. [Ueb08]) bieten eine Grundlage. So forcieren sie eine erste Entzerrung des Beziehungsgeflechts zwischen IT-Services und Geschäftsprozessen, indem sie ihre Kalkulationsverfahren entlang der Abstraktionsebenen des IT-Betriebs (Hardware/Software→IT-Services→Geschäftsprozesse) ausrichten und damit eine Strukturierung der IT-Organisation vornehmen. Mit Blick auf eine aktive Steuerung der IT-Kosten weisen diese Verfahren dennoch vielfach bemängelte metho-

Anforderung
A1 – Reduktion der Komplexität in der Betrachtung des IT-Einsatzes
A2 – Ergänzung um technischen / organisatorischen Kontext
A3 – Berücksichtigung weiterer Wirkungen neben Kosten (z. B. Risiken, Nutzen)
A4 – Förderung der Kommunikation zwischen Akteuren unterschiedlicher Perspektiven
A5 – Offenlegung von Zielvorstellungen und Annahmen

Tabelle 1: Anforderungen an eine methodische Unterstüzung für das IT-Kostenmanagement

dische Defizite auf (z. B. [May98, SK09]): Sie abstrahieren von den Steuerungsobjekten (IT-Services etc.) und ihrem Kontext (A1,A2); sie betrachten lediglich Kostenwirkungen im traditionellen Sinne (A3); und ihre Kalkulationen basieren auf vielfachen Annahmen, Mechanismen zur Offenlegung sind nicht vorgesehen (A5). Vor diesem Hintergrund basiert der Beitrag auf der Annahme, dass sowohl Bedarf als auch Potential existieren, das methodische Angebot für das IT-Kostenmanagement zu erweitern.

3 Potentiale der Unternehmensmodellierung

Die Unternehmensmodellierung zielt auf die Strukturierung von Unternehmen mittels zielgerichteter Abstraktionen von IT *und* umgebendem Handlungssystem [Fra02]. Hierfür bietet sie verschiedene konzeptuelle Modelle, die unterschiedlichen Akteuren eine dedizierte Sicht auf das Unternehmen ermöglichen, etwa Modelle der Geschäftsprozesse oder des IT-Betriebs. Da die Modelle miteinander integriert sind, ermöglichen sie umfassende Analysen über den IT-Einsatz und unterstützen eine Perspektiven-übergreifende Kommunikation. Die Entwicklung der zu Grunde liegenden Modellierungssprachen erfolgt über die *Rekonstruktion* der zentralen Konzepte der Domäne. Auf diese Weise profitiert die entstehende Modellierungssprache von der etablierten und fundierten Fachsprache der Domäne; und es macht sie zugänglich für Akteure der Domäne, da die Konzepte eng mit den Begriffen der Domäne korrespondieren. Die graphische Notation ermöglicht darüber hinaus einen intuitiven Zugang und fördert die Kommunikation der Betrachter untereinander.

Abbildung 1 illustriert einen Auszug der Anwendung der Unternehmensmodellierungsmethode „Multi-Perspective Enterprise Modelling“ (MEMO, [Fra02]). Er zeigt ein Modell einer IT-Landschaft, das die Leistungserstellung der IT-Organisation strukturiert und darstellt (modelliert mit der ITML, [Kir08]). Das IT-Modell ist mit einem Modell der Prozesslandschaft integriert (ORGML, [Fra02]). Dies ermöglicht, neben der Strukturierung des IT-Einsatzes, Software-unterstütze Analysen z. B. über Abhängigkeiten von IT-Services oder IT-Ressourcen. Die in Abb. 1 annotierten Balken zeigen typische Akteure

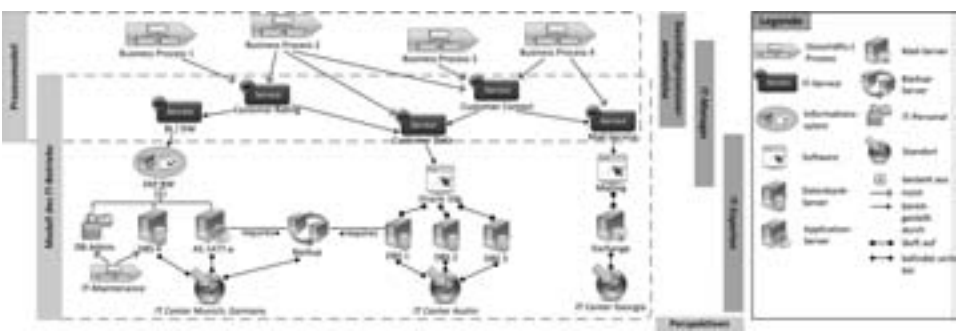


Abbildung 1: Anwendungsbeispiel der Unternehmensmodellierung mit integriertem Prozess-/IT-Modell und verschiedenen Perspektiven

des IT-Kostenmanagements und (auf horizontaler Ebene der Balken) ihre Perspektive auf das Unternehmen mit jeweils relevanten Konzepten.

Für das IT-Kostenmanagement bieten Unternehmensmodelle – z. B. auf Basis der ORGML und ITML – eine fundierte und vielversprechende Grundlage. Einerseits schaffen sie Transparenz (A1), indem sie die Komplexität in der Betrachtung des IT-Einsatzes durch gezielte Abstraktion(en) reduzieren; der Fokus der Modelle auf Typen (damit: die Abstraktion über konkrete Instanzen) trägt zu der Komplexitätsreduktion bei. Zugleich ermöglichen sie, bei Bedarf technische oder organisatorische Details z. B. zu IT-Ressourcen einzublenden (A2). Die Integration der Modelle erlaubt darüber hinaus Perspektiven-übergreifende Analysen von z. B. Wirkungsbeziehungen und fördert, zusammen mit der graphischen Notation, die Kommunikation zwischen den verschiedenen Akteuren (A4).

4 Spezifische Erweiterungen der Unternehmensmodellierung für das IT-Kostenmanagement – Beispiel und Forschungsagenda

Jenseits der grundsätzlichen Potentiale der Unternehmensmodellierung ist jedoch offensichtlich, dass für eine gezielte Unterstüzung des IT-Kostenmanagements das bisherige Sprachangebot gängiger Ansätze der Unternehmensmodellierung um spezifische Konzepte zu erweitern ist, die dedizierte Analysen der Kostenstrukturen und zugehöriger Gestaltungsoptionen ermöglichen. Anhand des skizzierten Anwendungsbeispiels aus Abbildung 1 werden die Erweiterungen beispielhaft illustriert und Forschungspotentiale für Unternehmensmodellierung und IT-Kostenmanagement aufgezeigt.

Zum einen ist es notwendig, Kosten zu Bezugsobjekten annotieren zu können (siehe Abb. 2, ①). Zum anderen werden spezifische Assoziationen benötigt (②, rote Kanten), die eine Verrechnung von Kosten über die verschiedenen Beziehungsgeflechte auf IT-Services und die Geschäftsprozesse ermöglichen (③). Auf diesen grundlegenden Sprachmitteln kann ein IT-Kostenmanagement aufbauen. Der IT-Controller kann beispielsweise IT-Kosten auf verschiedenen Abstraktionsebenen in ihre Bestandteile differenzieren, indem die modellierten Kostenallokationen zurückverfolgt werden (④). Dabei sind die Annahmen, die der Ermittlung der Kostengrößen zu Grunde liegen, aufzuzeigen (⑤). Um eine angemessene Interpretation der Kostenwirkungen zu unterstützen, ist eine Abbildung des technischen und organisatorischen Kontextes zu einzelnen Kostengrößen notwendig. Hierfür empfiehlt sich, Konzepte für z. B. Verträge und Kapazitäten anzubieten, welche die Kosteninformationen von IT-Ressourcen und IT-Services ergänzen (④, ⑥). Durch eine Verknüpfung der Unternehmensmodelle mit operativen Informationssystemen können diese um Instanzinformationen, etwa die Kapazitätsauslastung von Instanzen eines bestimmten IT-Ressourcentypen, angereichert werden (⑥). Schließlich ist das Modell um Wirkungen und Informationen zu ergänzen, welche den IT-Controller bei der Bewertung der Kostenstrukturen und zugehöriger Gestaltungsoptionen unterstützen. Zu denken ist u. a. an die Offenlegung der Zielvorstellungen, die z. B. mit der Anschaffung einer IT-Ressource verbunden waren, Unternehmens-externer Begründungen des IT-Einsatzes (z. B. die Erhaltung von Zertifizierungen) oder assoziierter Nutzeffekte und Risiken (⑦).

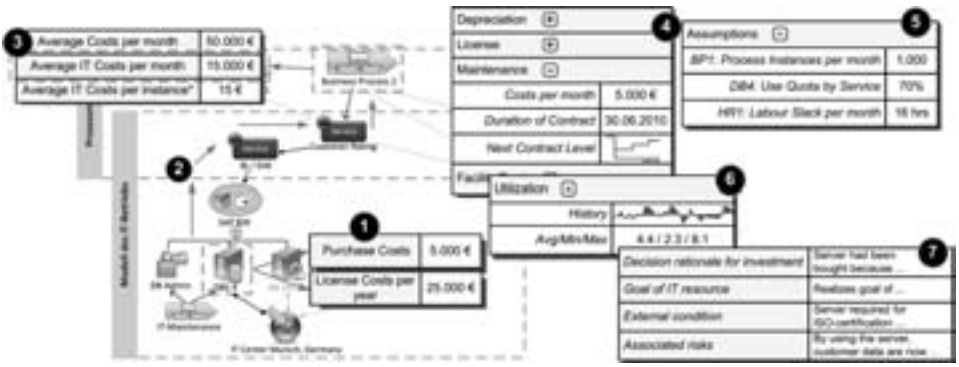


Abbildung 2: Anwendungsbeispiel eines um spezifische Konzepte für das IT-Kostenmanagement angereicherten Unternehmensmodells

Mit einem entsprechend erweiterten Unternehmensmodellierungsansatz hat der IT-Controller die Möglichkeit, auf unterschiedlichen Abstraktionsebenen jeweils relevante Wirkungen und Informationen einzublenden und dadurch gezielt mit einzelnen Akteuren in Diskurs zu treten. Die Visualisierung im Unternehmensmodell und die Annotation zu den jeweiligen Bezugsobjekten (z. B. IT-Services) verspricht, nicht nur den IT-Controller bei seinem Analyse- und Entscheidungsprozess im IT-Kostenmanagement, sondern auch die Kommunikation an Akteure mit spezifischen Perspektiven zu unterstützen.

Die Entwicklung entsprechender Erweiterungen erfordert die Rekonstruktion der skizzierten Konzepte als Sprachmittel und ist mit reizvollen Fragestellungen für die Unternehmensmodellierungsforschung und für die Praxis des IT-Controllings verbunden:

- Die Unternehmensmodellierung fokussiert i. d. R. auf die Abbildung von Typen. Bislang ist unklar, wie zeitpunkt-bezogene Informationen auf Typ-Ebene abgebildet werden können (z. B. Zeitpunkte bei Verträgen oder degressive Abschreibung).
- Vor dem Hintergrund verschiedener Anwendungsszenarien und beteiligter Perspektiven existiert kein einheitlicher Kostenbegriff für das IT-Kostenmanagement. Fragen nach beispielsweise der Relativität von Einzel- oder Fixkosten (z. B. [May98]) oder der Harmonisierung verschiedener Kostenbegriffe (z. B. interne/externe Unternehmensrechnung, [SK08]) werden zwar stetig diskutiert, bislang jedoch meist pragmatisch – von Fall zu Fall – gelöst; für das IT-Kostenmanagement ist ein flexibles Kostenkonzept zu entwickeln, dass diese Szenarien und Begriffsverständnisse integriert.
- Es ist zu klären, welche Informationen IT-Controller für Fragen des IT-Kostenmanagements (noch) benötigen. Insbesondere die Konzeption eines sowohl von technischen wie betriebswirtschaftlichen Perspektiven verstandenen und konsensierten „erweiterten“ Kosten- bzw. Wirtschaftlichkeitsbegriffs befindet sich erst am Anfang [Str09].

Für einige der skizzierten Konzepte existieren erste, wiederverwendbare Vorschläge etwa zur Abbildung von Kosten, Nutzen oder Risiken (z. B. [HSFJ08, SHF]). Andere Konzepte, etwa Kapazitäten, Verträge oder Unternehmens-externe Vorgaben, sind im Kontext der Unternehmensmodellierung bislang noch unerforscht.

5 Diskussion

Ziel des Beitrags ist die Analyse der Potentiale der Unternehmensmodellierung für das IT-Kostenmanagement. Hierfür wurden die grundsätzlichen Potentiale herausgearbeitet und notwendige Spracherweiterungen skizziert. Die Rekonstruktion der Konzepte aus der Fachsprache und die Metamodell-basierte Spezifikation stellen dabei in Aussicht, existierende Ansätze des IT-Kostenmanagements zu integrieren – und dadurch deren methodischen Defizite auszugleichen – sowie die angereicherten Modelle als Grundlage für die Entwicklung eines Informationssystems für das IT-Kostenmanagement nutzen zu können. Den Potentialen der skizzierten Methode steht aus Sicht der Unternehmen allerdings ihr erheblicher Aufwand gegenüber. Dieser sollte jedoch sinken, wenn (1) Synergiepotentiale in Erstellung und Nutzung der Unternehmensmodelle, z. B. mit dem Geschäftsprozess- oder dem IT-Management, realisiert werden und (2) ein Modellierungswerkzeug verfügbar ist, das – im besten Fall – mit den operativen Informationssystemen integriert ist.

Literatur

[Bau08] Ulrike Baumöl. IT-Controlling – Stand und Herausforderungen. *Controlling*, 20(12):649–654, 2008.

[BFM06] Wolfgang Becker, Stefan Fischer und Sascha Mika. Implementierungsstand des IT-Controlling. Bericht, Otto-Friedrich-Universität, Bamberg, 2006.

[Fra02] Ulrich Frank. Multi-Perspective Enterprise Modeling (MEMO): Conceptual Framework and Modeling Languages. In *Proc. of the* 35^{th} *Hawaii International Conference on System Sciences*, Honolulu, 2002.

[HSFJ08] David Heise, Stefan Strecker, Ulrich Frank und Jürgen Jung. Erweiterung einer Unternehmensmodellierungsmethode zur Unterstützung des IT-Controllings. In Bichler et al, Hrsg., *MKWI 2008*, Seiten 1017–1028, Berlin, 2008. GITO.

[Kir08] Lutz Kirchner. *Eine Methode zur Unterstützung des IT-Managements im Rahmen der Unternehmensmodellierung*. Logos, Berlin, 2008.

[May98] Reinhold Mayer. *Kapazitätskostenrechnung*. Vahlen, München, 1998.

[SHF] Stefan Strecker, David Heise und Ulrich Frank. RiskM: A multi-perspective modeling method for IT risk assessment. *Information Systems Frontiers*. accepted for publication.

[SK08] Marcell Schweitzer und Hans-Ulrich Küpper. *Systeme der Kosten- und Erlösrechnung*. Vahlen, München, 9. Auflage, 2008.

[SK09] Stefan Strecker und Herbert Kargl. Integrationsdefizite des IT-Controllings – Historischer Hintergrund, Analyse von Integrationspotentialen und Methodenintegration. *WIRTSCHAFTSINFORMATIK*, 51(3), 2009.

[Str09] Stefan Strecker. Wertorientierung des Informationsmanagements. *HMD – Praxis der Wirtschaftsinformatik*, 269:27–33, 2009.

[Ueb08] Falk Uebernickel. *IT-Produktkostenrechnung*. D-Druck-Spescha, St. Gallen, 2008.

Ein Modell zur Einflussanalyse von Änderungen in Entwicklungsprozessen

R. Koppe, S. Häusler, A. Hahn

OFFIS Institut für Informatik
{koppe, haeusler, hahn}@offis.de

Abstract: Wie ist der Einfluss einer neuen Methode auf die Produktivität eines Produktentwicklungsprozesses? Die Beantwortung dieser Frage ist eine der wichtigsten Erfolgsfaktoren für Unternehmen. Die quantitative Bewertung von neuen Methoden und Werkzeugen für existierende Entwicklungsprozesse unterstützt die Entscheidungsbasis von Managern für Prozessoptimierungen signifikant. Dieser Beitrag stellt ein Modell zur Einflussanalyse von Prozessänderungen vor, welches eine Bewertung neuer Methoden im Kontext von Entwicklungsprozessen gegen die Projektdimensionen von Zeit, Kosten und Produktqualität unterstützt.

1 Einleitung

Reifegradmodelle wie das Capability Maturity Model Integration (CMMI) [CMU06] definieren Ebenen der Reife einer Organisation. Ziel ist die Entwicklung von Prozessen zu solchen, die ständig optimiert werden. In diesem Zusammenhang versprechen neue Methoden und Werkzeuge große Potentiale für Prozess- und Produktverbesserungen. Doch ob sich eine neue Methode oder ein neues Werkzeug in einem Entwicklungsprozess wirklich lohnt, ist nicht leicht zu beantworten. Eine vom restlichen Entwicklungskontext isolierte Evaluierung neuer Methoden oder Werkzeuge ist oft nicht ausreichend. Als Alternative erlauben Pilotprojekte einen Einblick in das Prozessverhalten. Allerdings gestatten Pilotprojekte nur eine begrenzte Kontrolle, liefern keine statistisch signifikanten Ergebnisse [MPR05] und sind oftmals mit hohem (monetärerem) Aufwand verbunden. Zudem fehlt auch für die Bewertung von Pilotprojekten ein methodisches Gerüst.

Dieser Beitrag beschreibt ein Modell zur Einflussanalyse von Änderungen in Produktentwicklungsprozessen komplexer Systeme. Eingebunden ist das Modell in eine Methodik [Ko10], die es Prozessplanern und Entscheidern erlaubt, auf Basis identifizierter Wirkzusammenhänge (z.B. Expertenwissen), bereits vor Investitionen in neue Methoden (vgl. Investitionsrechnung z.B. [Go06]) Abschätzungen über deren Mehrwert durch Prozess-Simulation vorzunehmen. Methodik und Modell eignen sich dabei für beliebige Entwicklungsprozesse von Produkten (Software, Hardware, Computer Aided Design) und werden in aktuellen Fallstudien angewandt und evaluiert.

Im Folgenden gibt Abschnitt 2 begrenzt durch die Länge des Beitrags einen kurzen Überblick über verwandte Arbeiten. Abschnitt 3 beschreibt das Modell zur Einflussanalyse, welches in Abschnitt 4 exemplarisch angewendet wird. Schließlich liefert Abschnitt 5 einen Ausblick auf weitere Arbeiten.

2 Verwandte Arbeiten

Verwandte Arbeiten motivieren einen klaren Handlungsbedarf zur Analyse von Prozessänderungen durch neue Methoden, auch wenn ihr Fokus allgemeiner [Ba04] oder detaillierter [Kr08] als hier gesetzt wird. Es existieren zahlreiche Arbeiten zur Einflussanalyse und Einflussbewertung im Bereich des Software Engineering (SE) [Ke99], deren Vorgehen und Ansätze auch auf andere Bereiche der Produktentwicklungen anwendbar sind. Software Process Simulation (SPS) ist ein verbreiteter Ansatz, der durch Abdel-Hamid und Madnick [AM91] geprägt wurde. SPS wird unter anderem dazu verwendet Performanz-Analysen und „was-wäre-wenn"-Analysen durchzuführen und den Einfluss von Prozessänderungen vorherzusagen. Für einen Überblick verweisen wir auf [MP08] und [ZKP08]. Die grundlegenden Modelle werden in der Regel durch empirisch gewonnen Daten kalibriert, um das Verhalten der realen Welt abzubilden [Ra99]. Dieses führt zum Empirical SE, mit dem Ziel Wirkzusammenhänge zwischen verschiedenen Prozesselementen zu beschreiben. Der Detaillierungsgrad modellierter Wirkzusammenhänge ist proportional abhängig von ihrem Einfluss auf kritische Faktoren wie Zeit, Kosten und Qualität. Andere Ansätze simulieren stochastische Prozesse [LK91] und damit die Folge von Aktivitäten, die durch Transitionen verbunden sind. Eine Transition besitzt dabei eine Übergangswahrscheinlichkeit, welche die Folge von Aktivitäten und damit auch den Prozessaufwand bestimmt. Existierende Arbeiten bieten kein standardkonformes generisches Modell zur Beschreibung von Prozessen und inhärenten Wirkzusammenhängen. Sie abstrahieren entweder sehr von diesen ([CE01], [DP07]), oder beschreiben Produkte und deren Charakteristika, welche den Prozessfortschritt wesentlich bestimmen, nicht adäquat ([Ma96], vgl. [Ru02]).

3 Beschreibung des Modells

Wir beschreiben zunächst die Grundidee zur Modellierung von Entwicklungsprozessen und Wirkzusammenhängen. Dann erläutern wir die Kernkonzepte des Modells und zeigen die Anwendung des Modells an einem Beispiel aus der Hardware Entwicklung.

Das Modell basiert auf der Idee, ein probabilistisches Modell zur Analyse des ökonomischen Mehrwerts von Prozessalternativen zu verwenden [DP07]. Ein solches Modell abstrahiert von der realen Welt und deren inhärentem Verhalten. Die Berechtigung hierfür liegt im immensen Aufwand, der notwendig wäre, um „vollständige" Pilotprojekte oder empirische Studien durchzuführen.



Abbildung 1: Einfacher Prozess mit Wahrscheinlichkeiten an Transitionen

Abbildung 1 zeigt einen vereinfachten Prozess zur Entwicklung eines Chips [Ko10] (vgl. Abbildung 3). Aufgaben (Blöcke) sind durch Transitionen verbunden. Die Analyse-Aufgabe überprüft die Ergebnisse der vorherigen Schritte. Durch Expertenbefragung wird die Notwendigkeit eines Rücksprungs auf 20% festgelegt, um die Ergebnisse des

Register Transfer Level (RTL) Designs zu überarbeiten. Analysen erlauben nun die Abschätzung des Aufwands eines solchen Prozesses, indem jede Aufgabe einen bestimmten Aufwand besitzt. Diese Modellierung abstrahiert allerdings sehr weit von der Wirklichkeit, in der unterschiedliche Einflüsse (Komplexität des Produkts, Genauigkeit des Analyse-Werkzeugs, Erfahrung des Entwicklers, …) etwa auf den Aufwand und den Erfolg einer Aufgabe wirken. Um diesem Problem Rechnung zu tragen, beschreiben wir relevante Wirkzusammenhänge explizit durch sogenannte Verhaltensmodelle in einem detaillierteren probabilistischen Modell.

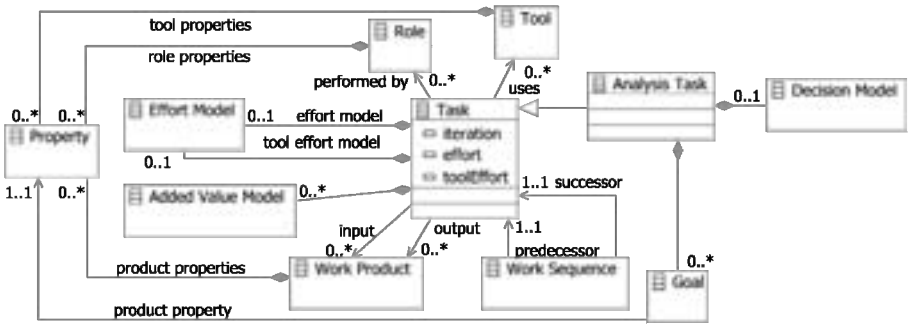


Abbildung 2: Basis des konzeptionellen Modells

Abbildung 2 zeigt einen Ausschnitt des konzeptionellen Modells, als Erweiterung des Software & Systems Process Engineering Meta-model (SPEM 2.0), einem Standard der OMG [OMG08] zur methodischen Beschreibung von Entwicklungsprozessen. Methodische Beschreibungen können für konkrete Prozesse wiederverwendet werden und dienen somit dem Aufbau einer Art Wissensdatenbank, deren praktische Vorteile wir auch bei unserem Ansatz nutzen. Im Gegensatz zu anderen Modellen (z.B. [DP07]) werden explizit Produkte (Work Product) als Ein- und Ausgaben von Aufgaben (Task) modelliert. Eigenschaften (Property) charakterisieren Produkte sowie auch Rollen (Role) und Werkzeuge (Tool), die einen nicht zu vernachlässigenden Einfluss auf das Prozessverhalten besitzen. Analyse-Aufgaben (Analysis Task) verfügen über Ziele (Goal), welche die Validierung von Produkteigenschaften (z.B. Qualität) beschreiben. Dynamische Verhaltensmodelle beschreiben das Prozessverhalten auf Basis der genannten Konzepte und detaillieren einfache Übergangswahrscheinlichkeiten (etwa von Markov-Ketten) in parametrisierbare Wahrscheinlichkeitsverteilungen. Das Modell verbindet also Methoden- / Prozessmodellierung mit analysierbaren Prozessverhalten. Wir unterscheiden drei Arten von solchen dynamischen probabilistischen Verhaltensmodellen: Added Value Model, Effort Model und Decision Model, die wir im Folgenden erläutern.

Added Value Model beschreibt den Wertbeitrag einer Aufgabe zu Eigenschaften von Produkten (z.B. Qualität und Umfang) oder Rollen (z.B. Lerneffekte). Solche Veränderungen werden als lineare, nicht-lineare oder durch parametrisierte Wahrscheinlichkeitsverteilungen beschrieben. Formal ist *Added Value* eine Funktion $i \times R \times T \times P_{in} \rightarrow p_{out}$ mit i als Zähler der aktuellen Iteration (Fortschritt, Lernen) der Aufgabe, R als Teil der Eigenschaften beteiligter Rollen, T als Teil der Eigenschaften verwendeter Werkzeuge und P_{in} als Teil der eingehenden Produkteigenschaften. Das Ergebnis p_{out} ist z.B. eine berechnete Produkteigenschaft (funktionale Korrektheit) eines ausgehenden Artefakts und beeinflusst somit alle Verhaltensmodelle nachfolgender Entwicklungsschritte.

Effort Model dient zur Berechnung des Aufwands einer Aufgabe. Der Aufwand unterteilt sich in menschliche Arbeit und Werkzeuglaufzeit, um so Kosten für Mensch und Werkzeug genau zu unterscheiden. Ein *Effort Model* ist eine Funktion i x R x T x P_{in} → effort ∈ {effort, toolEffort} und analog zu Added Value definiert. Das Ergebnis ist ein konkreter Wert in Stunden für den menschlichen Aufwand bzw. die Werkzeuglaufzeit.

Decision Model beschreibt die Entscheidung für eine folgende Aufgabe auf Grundlage bestimmter Ziele, die an Produkteigenschaften gestellt werden. Ein *Decision Model* ist eine Funktion i x R x T x P_{in} x G → ws mit den oben beschriebenen Parametern. G ist eine Menge von Zielen die für die Analyse-Aufgabe definiert wurden. Ein Entscheidungsmodell wird damit durch den einfachen Vergleich von Soll und Ist oder als komplexer (boolescher) Ausdruck definiert, der durch R, T und P_{in} beeinflusst wird. Das Ergebnis ws ist eine ausgehende Transition, welche auf die folgende Aufgabe verweist.

4 Anwendung des Modells zur Einflussanalyse

Wir zeigen nun die exemplarische Anwendung des Modells an einem Entwicklungsprozess aus der Hardware Entwicklung [Ko10]. Ziel der Prozessänderung ist die Reduktion des notwendigen Aufwands. Es sind zwei Prozessalternativen zu vergleichen:

- im bisherigen Entwicklungsprozess wird aus einem SystemC Modell durch einen Entwickler manuell ein Very High Speed Integrated Circuit Hardware Description Language (VHDL) Modell abgeleitet,
- der Prozess wird um ein neues Werkzeug zur automatischen Transformation eines SystemC Modells in ein entsprechendes VHDL Modell erweitert.

Es sei angemerkt, dass der grundlegende Ansatz auch auf andere Entwicklungsprozesse (System-, oder Softwareentwicklung) anwendbar ist. Lediglich die Wirkzusammenhänge sind andere. Außerdem ist der Ansatz nicht auf Bewertungen von Werkzeugänderungen oder auf Neuentwicklungen beschränkt. So erlaubt das Modell auch die Abbildung anderer Entwicklungsmethoden und -vorgehen oder auch Mitarbeiterkonstellationen.

Wir führen Einflussanalysen von Prozessänderungen durch, indem wir zunächst den Ist-Prozess und die Prozessänderung mit ihren Wirkzusammenhängen, die einen Einfluss auf den Prozessfortschritt besitzen (z.B. die aktuelle Qualität des Produkts) modellieren. Abbildung 3 zeigt dazu den Entwicklungsprozess mit Produkten, Eigenschaften, einem Ziel und einem Entscheidungsmodell. Die Prozessänderung ist umrahmt. Das Behavioral Design erzeugt ein *SystemC* Modell mit der Eigenschaft *functional correctness*. Diese Eigenschaft wird durch ein Added Value Model bestimmt, welches *Behavioral Design* zugeordnet ist. Das *SystemC* stellt eine Eingabe sowohl für die neue Aufgabe *Design Transformation* als auch für das *RTL Design* dar. Beide Aufgaben überführen das *SystemC* in ein *VHDL* Artefakt, welches auch eine Eigenschaft *functional correctness* besitzt. Das modellierte Verhalten der beiden Prozesse unterscheidet sich also in den Added Value Modellen der beiden Aufgaben, welche den Wertbeitrag der verschiedenen Aktivitäten zur funktionalen Korrektheit der Artefakte beschreiben. Erfüllt die *VHDL functional correctness* nicht das definierte Ziel (z.B. ist die verifizierte funktionale Korrektheit kleiner 100%), so ist ein Rückschritt zu *RTL Design* notwendig.

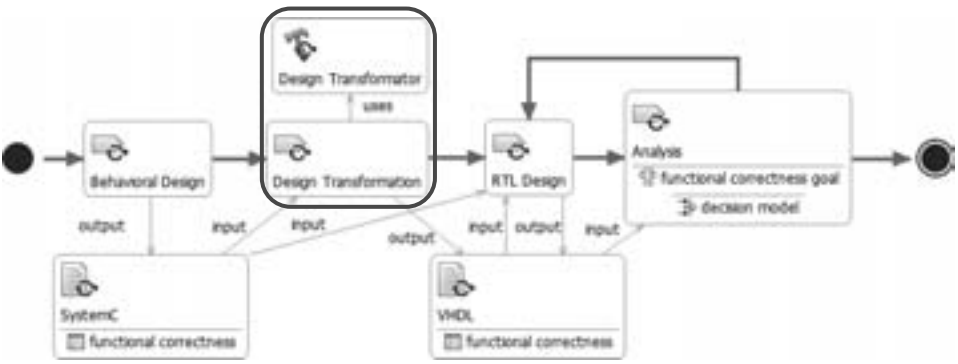


Abbildung 3: Erweiterung des Prozesses um Produkte, Ziel und Entscheidungsmodell

Die Added Value Modelle von *Design Transformation* und *RTL Design* werden als parametrisierte Wahrscheinlichkeitsverteilungen modelliert. Die Prozessänderung soll nun diese Verteilung der *functional correctness* deutlich nach rechts verschieben (z.B. sichert der Hersteller des Werkzeugs zu, aus einem korrekten SystemC immer korrektes VHDL zu generieren), um den notwendigen Aufwand durch Iterationen deutlich zu reduzieren, wie in Abbildung 4 gezeigt.



Abbildung 4: Wertbeitrag zur funktionalen Korrektheit des VHDL

Für die Analyse interessiert uns nun jeweils der notwendige Prozessaufwand der beiden Prozessalternativen. Monte Carlo Simulation wird für die Exploration der Verhaltensmodelle und Wahrscheinlichkeitsverteilungen verwendet. Die Ergebnisse werden dann statistisch aufbereitet (z.B. Minimum, Maximum, Mittelwert, Median und Quantile).

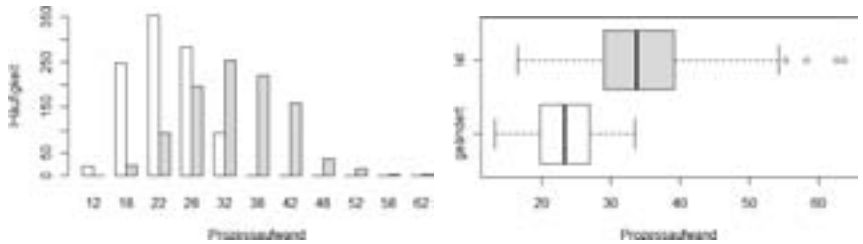


Abbildung 5: Ergebnisse des Prozessaufwands bei 1000-facher Simulation des Ist-Prozesses (dunkel) gegen den geänderten Prozesses (hell), links Verteilung des Aufwands als Histogramm, rechts aggregierte Lagemaße des Prozessaufwands als Box Plot

Abbildung 5 zeigt die Ergebnisse der beiden Alternativen grafisch als Histogramm und Box Plot. Es ist leicht zu erkennen, dass der geänderte Prozess (hell) vor dem aktuellen Prozess (dunkel) liegt. Der Wertbeitrag der neuen Methode wird als Differenz der jeweiligen Prozessaufwände etwa durch Mittelwert- oder Quantil-Vergleich beschrieben und eignet sich dann für strategische Break-even Point bzw. Pay-off Analysen.

5 Zusammenfassung und Ausblick

Dieser Beitrag präsentiert ein Modell zur Einflussanalyse von Methodenänderungen in Entwicklungsprozessen. Das Modell erweitert den OMG Standard SPEM 2.0 [OMG08] zur Beschreibung von Prozessen um Konzepte zur Modellierung von Wirkzusammenhängen zwischen Produkt- und Prozesseigenschaften. Eingebunden ist das Modell in eine Methodik [Ko10], welche die Modellierung von Ist- und Soll-Prozessen beschreibt und mittels Simulation einen Vergleich der Entwicklungsdauer und -kosten der Prozesse ermöglicht. Wir entwickeln derzeit Werkzeuge zur Modellierung und Analyse von Entwicklungsprozessen auf Basis des hier vorgestellten Modells. Weitere Werkzeuge unterstützen die Sammlung und Analyse von operativen Daten zur Modellierung von Wirkzusammenhängen. Bei unserer weiteren Arbeit werden wir unseren Ansatz und Werkzeuge evaluieren und deren Ergebnisse zur weiteren Verbesserung dieser nutzen.

Literaturverzeichnis

[AM91] Abdel-Hamid, T.K.; Madnick, S.E.: Software project dynamics: an integrated approach. Englewood Cliffs, Prentice Hall, 1991.

[Ba04] Balazova, M.: Methode zur Leistungsbewertung und Leistungssteigerung der Mechatronikentwicklung. Dissertation, Universität Paderborn, 2004.

[CE01] Cho, S.; Eppinger, S.: Product Development Process Modeling Using Advanced Simulation. In: Design Engineering Technical Conferences and Computers and Information in Engineering Conference, Pittsburgh, 2001. ASME 2001.

[CMU06] Carnegie Mellon University: CMMI for Development, Version 1.2. 2006.

[DP07] Deissenboeck, F; Pizka, M.: The Economic Impact of Software Process Variations. In: Software Process Dynamics and Agility. Springer, Berlin, 2007; S. 259-271.

[Go06] Götze, U.: Investitionsrechnung: Modelle und Analysen zur Beurteilung von Investitionsvorhaben. Springer, Berlin, 2006.

[HE08] Hassine, A.; Barke, E.: On Modeling and Simulating Chip Design Processes: The RS Model. In: Int. Engineering Management Conference. IEEE, Estoril, 2008; S. 81-85.

[Ke99] Kellner, M. et.al.: Software Process Simulation Modeling: Why? What? How?. In: Journal of Systems and Software, Band 46, Nr 2-3. Elsevier, Amsterdam. S. 91-105

[Ko10] Koppe, R. et.al.: Process Model Based Methodology for Impact Analysis of New Design Methods. In: Proceedings of the 1st International Conference on Modelling and Management of Engineering Processes, Cambridge 2010. Springer, London, 2010.

[Kr08] Kröll, M.: Methode zur Technologiebewertung für eine ergebnisorientierte Produktentwicklung. Dissertation, Universität Stuttgart, 2008.

[Ma96] Madachy, R.: System dynamics modeling of an inspection-based process. In: 18th Int. Conference on Software Engineering. IEEE, Berlin, 1996. S. 376-386.

[MP08] Müller, M; Pfahl, D.: Simulation Methods. In: Guide to Advanced Empirical Software Engineering. Springer, London, 2008.

[MPR05] Münch, J.; Pfahl, D.; Rus, I.: Virtual Software Engineering Laboratories in Support of Trade-off Analyses. In: Software Quality Journal, Band 13, Nr 4. Springer, Netherlands, 2005; S. 407-428

[OMG08] OMG: Software & Systems Process Engineering Meta-model Specification, 2008.

[Ra99] Raffo, D. et.al.: Empirical Studies Applied to Software Process Models. In: Empirical Software Engineering, Band 4, Nr 4. Springer, Netherlands, 1999; S.353-369.

[Ru02] Rus et.al.: Systematically combining process simulation and empirical data in support of decision analysis in software development. In: 14th Int. Conference on Software Engineering and Knowledge Engineering. ACM, Ischia, 2002. S. 827-833.

[ZKP08] Zhang, H.; Kitchenham, B.; Pfahl, D.: Reflections on 10 Years of Software Process Simulation Modeling: A Systematic Review. In: Making Globally Distributed Development a Success Story. Springer, Berlin, 2008; S. 345-356.

Workshop

3. Workshop Informatik und Nachhaltigkeitsmanagement

Burkhardt Funk, Peter Niemeyer

Nachhaltigkeitsinformatik

Burkhardt Funk, Peter Niemeyer
Leuphana Universität Lüneburg
Scharnhorststraße 1, 21335 Lüneburg
{funk|niemeyer}@uni.leuphana.de

1 Einleitung

Das betriebliche Nachhaltigkeitsmanagement berücksichtigt die ökonomische, soziale und ökologische Dimension unternehmerischen Handelns. Der Wirtschaftsinformatik und allgemein der angewandten Informatik kommt dabei eine Schlüsselrolle zu. Sie unterstützt einerseits die informatorische Absicherung unternehmerischer Entscheidungen, beispielsweise durch den Einsatz betrieblicher Umweltinformationssysteme, und ermöglicht andererseits erst die Realisierung von Optimierungspotentialen im Nachhaltigkeitsmanagement. Dies schließt die aktuell in der Diskussion befindliche Green IT im Sinne der Bestrebungen nach erhöhter Energieeffizienz und damit einhergehend der Reduzierung der CO_2-Emissionen ein, geht aber im zuvor beschriebenen Sinne darüber hinaus. Der Workshop konzentriert sich auf die IT-Unterstützung der ökonomischen und ökologischen Dimensionen des Nachhaltigkeitsmanagements und erörtert Forschungsfragen, Perspektiven und Anwendungen in diesem Feld.

Im Rahmen des Review-Prozesses wurden vier Beiträge zur Veröffentlichung angenommen, die sich überwiegend mit Konzeption und Evaluation von IT-Systemen zur Unterstützung spezifischer Aspekte des Nachhaltigkeitsmanagements befassen. Die Beiträge werden im Folgenden kurz vorgestellt.

2 Beiträge

Vor dem Hintergrund zunehmender regulativer Maßnahmen im Bereich der Energiepolitik beschreibt Niggemann Anforderungen an IT-Systeme zur Unterstüzung des Energiecontrollings, sogenannter Energiecontrollingsysteme (ECS). Solche IT-Lösungen sollen Unternehmen dabei unterstützen, über eine verursachungsgerechte Zuordnung von Energieverbräuchen Transparenz in die energierelevanten Prozesse zu bringen und darüber die Optimierung der Energieeffizienz zu ermöglichen. Der Beitrag erläutert an einem Fallbeispiel in einem KMU die durch den Einsatz eines ECS identifizierten Effizienzsteigerungspotentiale.

Der Beitrag von Bachmann beschäftigt sich mit der IT-seitigen Abbildung von Compliance-

Anforderungen wie REACh und RoHs, deren Umsetzung eine genau Kenntnis der stofflichen Zusammensetzung von Produkten erfordert. Bei der Beschaffung entsprechender Substanzdaten kommte den Lieferanten eine wesentliche Rolle zu. Die Anforderungen, die in diesem Zusammenhang an die betrieblichen Informationssysteme der betroffenen Unternehmen abzuleiten sind, erläutert Bachmann am Beispiel einer auf SAP Netweaver basierenden IT-Lösung.

Wohlgemuth, Krehahn und Ziep untersuchen den Einsatz von Methoden des Mobile Computings zur Vereinfachung der Erfassung von Umweltdaten. Der hohe manuelle Aufwand bei der Erfassung von Umweltdaten ist mit dafür verantwortlich, dass die Verbreitung von betrieblichen Umweltinformationssystemen gerade in dem Segment kleiner und mittlerer Unternehmen vergleichsweise gering ist. Eine Erfassung von Abfalldaten, Zählerdaten und weiteren umweltrelevanten Informationen über mobile Endgeräte kann diesen Aufwand reduzieren. In ihrem Beitrag beschreiben und evaluieren Wohlgemuth et al. eine prototypische Umsetzung im Rahmen eines Open Source Projektes.

Die Vernetzung nationaler Studienprogramme im Bereich der bertrieblichen Umweltinformatik betrachten Giesen, Haak und Marx Gómez in ihrem Beitrag. Am Beispiel eines binationalen Promotionsprogramm zwischen der Universität Oldenburg und der Technischen Universität Havanna, sowie gemeinsamer Studiengangsangebote deutscher und lateinamerikanischer Universitäten werden die Potentiale der internationalen wissenschaftlichen Kooperation im Bereich Betrieblicher Umweltinformationssysteme dargestellt.

3 Danksagung

Unseren Kollegen Lorenz Hilty (Uni Zürich), Jorge Marx Gómez (Uni Oldenburg), Andreas Möller (Uni Lüneburg), Frank Teuteberg (Uni Osnabrück) und Volker Wohlgemuth (HTW Berlin) danken wir für die engagierte Mitarbeit im Programmkomitee.

Nachhaltigkeit durch Energiecontrolling

Stephan Niggemann

econ solutions GmbH
Heinrich-Hertz-Straße 25
75334 Straubenhardt
Stephan.Niggemann@econ-solutions.de

Abstract: Nachhaltigkeit in ihrer ökonomischen, ökologischen und sozialen Dimension kann durch unterschiedliche Maßnahmen erreicht werden. Maßnahmen zur Energieeffizienz sind ein wesentlicher Bestandteil, dessen Potentiale von ausgewiesenen Experten als überdurchschnittlich eingestuft werden. Um Maßnahmen ergreifen zu können, bedarf es jedoch der Transparenz bei energierelevanten Größen. Auch wenn der Return on Investment für ein IT-System zum Energiecontrolling im Vorhinein schwer zu berechnen ist, zeigt das Fallbeispiel, wie sich mit geringem Aufwand signifikante Effizienzpotentiale heben lassen.

1 Nachhaltigkeit und Energiemanagement - nur Trends?

Der Begriff der Nachhaltigkeit (engl. Sustainability) ist seit geraumer Zeit in aller Munde. Nachhaltiges Handeln ist dabei immer durch das Gleichgewicht bzw. die Berücksichtigung der ökonomischen, ökologischen und sozialen Auswirkungen geprägt. Gesellschaftlich genießt nachhaltiges Wirtschaften den höchsten Stellenwert, da die Umwelt und Ressourcen geschont werden, Arbeitsplätze erhalten bleiben und gleichzeitig wirtschaftliches Wachstum erzielt wird. Eberhard Jochem vom Fraunhofer-Institut für System- und Innovationsforschung ist der Meinung, dass Klima- und Umweltschutz mit wirtschaftlichem Wachstum sich nicht widersprechen, sondern mit entsprechendem Einsatz und Know-How möglich ist.

Der Informations- und Kommunikationstechnik kommt in diesem Zusammenhang eine bedeutende Rolle zu. Dem Begriff des Informationssystems folgend (vgl. [FS06], S. 1, [HN04], S. 84) stellt ein IT-System für das Thema „Energie“ Funktionen zur Erfassung, Übertragung, Transformation, Speicherung und Bereitstellung von kontextbezogenen Daten bereit. Da Energie in nahezu allen Unternehmensprozessen eine Rolle spielt, ist eine integrierende Funktion eines solchen Informationssystems für den zunehmend zum Wettbewerbsfaktor werdenden Themenbereichs „Energie“ notwendig. Die nachfolgende Abbildung enthält einen Überblick der energierelevanten Aspekte in Anlehnung an die grundlegenden Unternehmensprozesse nach Porters Definition der Wertschöpfungs- bzw. Wertkette (vgl. [Po85]).

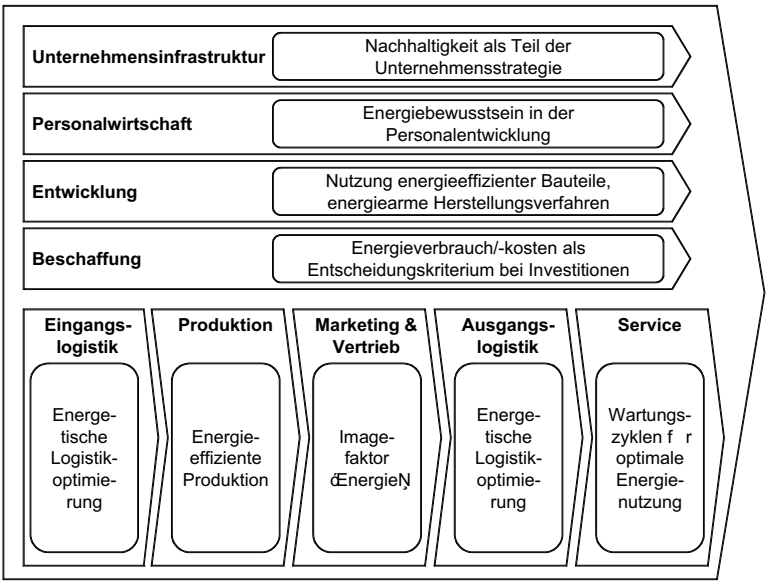


Abbildung 1 Energieaspekte am Beispiel der Wertschöpfungskette

Aufgrund des noch jungen Fokus in der Anwendung lässt sich aktuell noch keine allgemein gebräuchliche Definition zu Begriffen wie Energiemanagement und Energiecontrolling finden. Die im Jahr 2009 verabschiedete Norm DIN EN 16001:2009 versucht sich in der Definition des Begriffs „Energiemanagementsystem" mit folgendem Wortlaut: „Gesamtheit von miteinander zusammenha ngenden oder in Wechselwirkung zueinander stehenden Elementen einer Organisation zur Erstellung einer Energiepolitik sowie strategischer Ziele und zur Erreichung dieser Ziele" [DIN09]. Dem Energiecontrolling kommt in diesem Zusammenhang die Rolle des Überwachens und Monitorings zu. In Anlehnung an die Feststellung von Horváth zu den anwendungsbezogenen Aufgaben des Controllings (vgl. [Ho09], S. 130) stellt Energiecontrolling die „Ausrichtung an den energetischen Unternehmenszielen durch Koordination von Planung und Informationsversorgung mittels Kennzahlen im gesamten Energiemanagementsystem unter Sicherstellung der Rationalität" dar.

Die Bedeutung des Themas „Energie“ spiegelt sich auf politischer und gesellschaftlicher Ebene wieder. Die Vorgabe der EU mit der EU-Richtlinie 2006/32/EG, die in nationales Recht überführt werden muss, mündet in den aktuellen Entwurf des Energie-Dienstleistungs-Gesetzes [EDLG10]. Des Weiteren lässt sich auf Ebene der Parteien feststellen, dass Umweltbewusstsein und Ressourcenschonung, in früheren Jahren das politische Betätigungsfeld der Partei Bündnis90/Die Grünen, mittlerweile in allen Parteiprogrammen einen wesentlichen Stellenwert einnimmt. Damit wird, ähnlich wie in der Wirtschaft, dem Imagefaktor dieses Themas Rechnung getragen. Dieser stark wachsende und von unterschiedlichen Seiten und mit unterschiedlichen Intentionen behandelte Interessenbereich verlangt eine schnelle, strukturierte Einführung von IT-Systemen zur Beherrschung. Hierbei ist besonders interessant, dass sowohl Datenmaterial als auch notwendige technische Infrastruktur, um sich mit dem Thema Energie eingehender zu beschäftigen in der Industrie zwar vorliegt, jedoch aufgrund der geringen Beachtung brach lag. Nachfolgend wird aufgezeigt, wie eine Integration dieser Aspekte mit dem Ziel des Energiecontrollings im ersten Schritt und dem Ausbau zu einem ganzheitlichen Energiemanagement im zweiten Schritt gestaltet werden kann sowie welche Erfolge hieraus bereits in kurzer Zeit realisiert werden können.

2 Rahmenbedingungen und Defizite in der Industrie

Energiemanagement ganzheitlich zu betreiben wird zunehmend von äußeren Einflussfaktoren bestimmt oder auch gefordert. Nachhaltigkeit als expliziter Bestandteil der Unternehmensstrategie haben sich viele Unternehmen unter teilweise anders lautenden Begriffen wie Klima- und Umweltschutz oder Verantwortung schon lange auf die Fahnen geschrieben. Der Hauptantrieb war und ist das Image des Unternehmens. Gesetzliche Regelungen wie das Erneuerbare-Energien-Gesetz ([EEG08]) oder der Entwurf des Energie-Dienstleistungs-Gesetzes ([EDL10]) sowie damit verbundene politische Anreize erhöhen die Notwendigkeit zur verstärkten Behandlung dieses Themas in Unternehmen. Weitere Aspekte in diesem Zusammenhang sind branchenspezifische Auflagen wie der angedachte CO2-Footprint in der Automobilindustrie[1], aber auch technische Neu- und Weiterentwicklungen durch den Fokus auf optimierte Energieausnutzung, z.B. im Bereich von energieeffizienten Elektromotoren oder der Wärmerückgewinnung.

[1] Mit dem CO2-Footprint sollen zukünftig alle Neuwagen versehen werden. Diese Kennzahl soll darstellen, wie viel CO2 bei der Produktion des Fahrzeugs ausgestoßen wurde.

Gerade im Bereich der Energieeffizienz in der Industrie sind staatliche Fördertöpfe sowohl für Investitionen als auch für Beratungsleistungen vorhanden. Die durch die KfW geförderte Energieeffizienzberatung, die sich aus den Bestandteilen der Initial- und der Detailberatung zusammensetzt, soll Unternehmen bei der Identifikation von Einsparpotentialen unterstützen. Im Rahmen dieser Beratung mangelt es jedoch vielfach an einer belastbaren Datenbasis sowie einem signifikanten Detailgrad der Messwerte. Empfehlungen können oftmals nur auf dieser Basis von Schätzung und Berechnung durch die Berater abgegeben werden. Dieses Defizit ist gleich doppelt nachteilig, da einerseits, der berechnete energetische Sachverhalt durch Messwerte nicht bestätigt und andererseits die Wirksamkeit der Maßnahme nicht validiert und verifiziert werden kann. Die notwendige Transparenz über den Energieverbrauch, um die eigene Erwartungshaltung zum Energieverbrauch zu überprüfen sowie Maßnahmen zu definieren, ist in der Regel nicht gegeben. Dafür stand das Thema Energie und dessen Einsatz aus unterschiedlichen Gründen[2] nicht im Fokus der Industrie.

Das Potential, das durch Energieeffizienzmaßnahmen gehoben werden kann, wird von Studien des Fraunhofer Instituts für System- und Innovationsforschung ([SWB09]) und McKinsey ([Mc09]) auf 15-35% branchenabhängig geschätzt. Die Internationale Energieagentur weist den notwendigen Beitrag durch Energieeffizienzmaßnahmen bis zum Jahr 2030 mit 57% als größten notwendigen Bestandteil aus, um damit der Erderwärmung durch Reduzierung des CO2-Ausstosses Einhalt zu gebieten ([IEA08]).

Energiecontrollingsysteme sind für dieses Gebiet kein neues Produkt. Die Anbieter weisen sehr unterschiedliche Hintergründe auf, so dass das Spektrum der Produkte vielfältig ist. Angefangen bei reinen Software-Produkten finden sich Hersteller von Energiezählern, die durch weitere Hardwarekomponenten wie Datenlogger und entsprechende Software ihre Produktpalette erweitern.

3 Konzept eines IT-Systems für Energiecontrolling

Auf Basis der aufgeführten Rahmenbedingungen und Defizite wurde bei der econ solutions GmbH ein Konzept für ein ganzheitliches (IT-)System zum Energiecontrolling entwickelt. Unter Energiecontrolling wurde bei der Entwicklung die oben aufgeführte Definition zu Grunde gelegt. Ein Energiecontrollingsystem stellt darauf aufbauend ein Informationssystem dar, das die energierelevanten Aspekte des Unternehmens umfasst. Der Anforderungskatalog für die Entwicklung wies die nachfolgend dargestellten Aspekte auf.

[2] Hierzu zählen einerseits ökonomische Aspekte wie niedrige Preise für den Energieverbrauch sowie der geringe gesellschaftliche Fokus auf Themen wie Ressourcenschonung und Umweltbewusstsein.

Die Einführung eines neuen IT-Systems erfordert oftmals die Anschaffung von Hardware. Bei einem Energiecontrolling-System ist gemäß der oben aufgeführten Definition eine Infrastruktur zur Erfassung der Messwerte notwendig, ein Messsystem. Vorhandene Zähler oder Sensoren sollten in das System über entsprechende Schnittstellen eingebunden werden können. Als Schnittstellen-Standards für Messpunkte sind die folgenden in Europa am meisten verbreitet. Der Impulsausgang nach DIN 43864 ist eine gängige Schnittstelle für Verbrauchszähler, z.B. bei Gas, Wasser, Strom oder Fernwärme. Analoge Sensoren mit integrierten Wandlerelektroniken messen Zustände wie Temperaturen, Drücke oder auch Konzentrationen und geben analoge Strom- (z.B. 4 bis 20 mA) oder Spannungssignale (z.B. 0 bis 10V) aus, die zur proportionalen Umrechnung in die entsprechenden Einheiten verwendet werden können. Als relevante Standards aus dem Bereich der Gebäudeleittechnik sind LON[3], EIB[4] oder KNX[5] zu nennen. Der M-Bus ist als ebenfalls genormtes Kommunikationsprotokoll nach DIN 13757 für die Übermittlung von Zählerständen relevant [DIN05]. Eine RS485- und RS232-Schnittstelle zur Übermittlung von Daten in einem leistungsfähigen Bus-System gehörte ebenfalls zu den Anforderungen.

Zur Kommunikation von Daten oberhalb der Erfassungsebene stellte das TCP/IP-Protokoll den einzig sinnvollen Ansatz dar. Die Verarbeitung von zu erwartenden Datenmengen sowie der hohe Verbreitungsgrad von IP-basierten Netzwerken an nahezu allen Orten im Unternehmen weisen hier den Weg. Ein weiterer Vorteil dieses Protokolls stellt der vielfältig gestaltbare Zugriff auf das System dar, einerseits im Netzwerk selbst (Intranet) als auch von außen für Partner (Extranet) oder die Öffentlichkeit (Internet) (vgl. [LLS10], S. 341).

Da Sicherheitsaspekte in Unternehmen im Bereich der IT an Bedeutung gewinnen, werden die Konfigurationsmöglichkeiten der Arbeitsplatzrechner durch die Nutzer zunehmend eingeschränkt. Dies hat zur Folge, dass proprietäre Software nur durch Administratoren eingerichtet werden kann. Proprietäre Software ist oftmals auch Betriebssystem bezogen, so dass eine einwandfreie Integration auf unterschiedlichen Einsatzrechnern nicht zwingend möglich ist. Web-Anwendungen stellen hier eine adäquate Lösung bereit, so dass dieser Aspekt für die Entwicklung wesentlich war. Durch die Lauffähigkeit der Anwendung im Browser ist keine Installation auf den Clients erforderlich. Zudem werden die Wartungsarbeiten durch nur ein einmaliges Update auf dem Server minimiert.

Neben den auf Hardwareseite zu implementierenden Schnittstellen ist der Zugriff auf Datenbanken zur Extraktion von energierelevanten Daten ebenfalls notwendig. Der Einsatz einer relationalen Datenbank zur Nutzung von SQL sowohl zum Import von Daten als auch ggf. zum Export in andere Anwendungen stellt hier den verwendeten Standard dar.

[3] Das Local Operating Network, kurz LON, ist ein Feldbus und findet vornehmlich in der Gebäudeautomation Anwendung.
[4] Der Europäische Installationsbus (EIB) ist nach EN 50090 standardisiert und wird in der Gebäudeautomation eingesetzt.
[5] KNX ist der Nachfolger diverser Feldbusse, wie z.B. EIB, BatiBus. Er ist nach ISO 14543-3 standardisiert.

Energierelevante Daten werden nicht nur in einem Messsystem erhoben. Je nach Anwendungs- und Verarbeitungszweck ist die Anreicherung der Messdaten durch Daten aus anderen Unternehmenssystemen sinnvoll oder notwendig. In umgekehrter Richtung können andere Systeme auch Werte aus einem Energiecontrolling-System für die eigene Verarbeitung integrieren. In ERP-Systemen können beispielsweise die verrechneten Werte des Energieverbrauchs auf Kostenstellen automatisiert geladen werden. In Systemen der Leittechnik (z.B. Gebäudeleittechnik, Prozessleittechnik) werden teilweise auch energierelevante Daten erhoben, die jedoch oftmals ungenutzt bleiben oder nur Ausschnitte des Unternehmens abbilden. Für das Energiecontrolling sind jeweils die Daten relevant, die in unmittelbaren Zusammenhang zum Themenfeld Energie stehen. Diese galt es über entsprechende Softwareschnittstellen zu integrieren. Die Betriebsdatenerfassung weist Informationen zu Stückzahlen, Laufzeiten usw. auf. Durch die Integration dieser Daten sind Aussagen zu Energiekosten pro Stück oder energiekostenoptimalen Herstellungsverfahren möglich.

Das auf dieser Basis entwickelte IT-System übernimmt die komplette Konfiguration der Systemkomponenten zur Erfassung von Messdaten sowie die Bereitstellung der Daten für Auswertungen und Berichte in Form einer Webanwendung. Als Hardware-Komponenten kommen dezentrale Datenlogger zum Einsatz, die mit einem zentralen Server über das TCP/IP-Protokoll für die Konfiguration und den Datenaustausch kommunizieren. Vergleiche von Messdaten können ex post durchgeführt werden. Eine Simulation von potentiellen Verbrauchseinsparungen ist aufgrund der oftmals komplexen Stellhebel der Maschinen- und Anlagensteuerungen zum aktuellen Zeitpunkt nicht integriert.

4 Anwendungsbeispiel bei einem mittelständischen Unternehmen

Der Einsatz des Energiecontrolling-Systems der econ solutions GmbH wurde im Rahmen einer Installation bei einem mittelständischen Unternehmen aus dem Bereich des Kunststoffspritzgießens über vier Monate untersucht. (andere Systeme, Innovation, Neuentwicklung) Die Rahmenbedingungen, die bei diesem Unternehmen vorgefunden wurden, sowie die Erkenntnisse aus der Nutzung des Systems werden nachfolgend dargestellt.

Neben den Produktionsprozessen von Kunststoffbauteilen sind Querschnittsprozesse, deren Output in allen Produktionsprozessen notwendig ist, wesentlicher Bestandteil. Ein wesentliches Charakteristikum der Querschnittsprozesse ist die Erwartungshaltung der Mitarbeiter zur entsprechenden Performance dieser. Eine direkte Verantwortung und Betreuung dieser Prozesse lag nicht vor. Deshalb wurde im Untersuchungszeitraum eine detaillierte Erfassung der Prozesse „Druckluft“ und „Kunststoffgranulat-Trocknung“ vorgenommen. Besonderes Interesse galt den Verbrauchs- und Lastprofilen in Produktions- und produktionsfreien Zeiten.

Die Druckluft-Anlage verfügte über zwei Kompressoren mit einer maximalen Leistungsaufnahme von jeweils 25 kW. Die Analyse der Leistungsaufnahme sowie des Verbrauchs gerade in produktionsfreien Zeiten (in diesem Fall an einem Wochenende) zeigte eine regelmäßige Leistungsaufnahme von ca. 11 kW, was mehr als 40% der maximal möglichen Leistungsaufnahme beträgt. Eine vollständige Abschaltung des Druckluftkreislaufs während dieser Zeit war nicht möglich, da andere Querschnittsfunktionen wie z.B. eine Kühlungsanlage über Druckluft ihre Ventile öffnen und schließen ließ. Die Abbildung unten zeigt das Leistungsprofil des Druckluftsystems an einem produktionsfreien Tag.

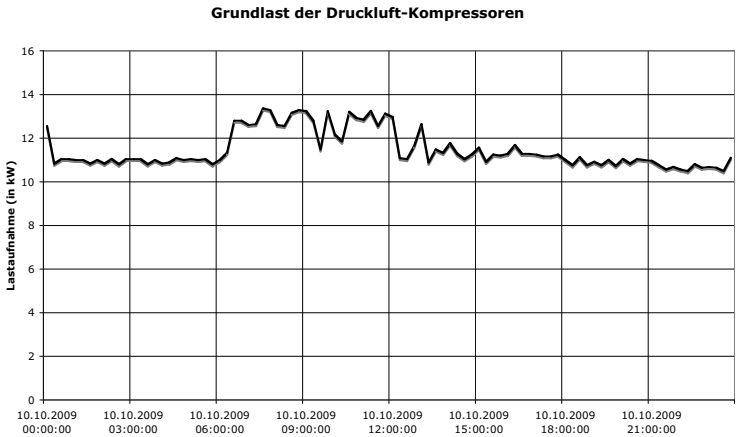


Abbildung 2 Lastaufnahme der Druckluft-Kompressoren in produktionsfreier Zeit (1 Tag)

Bevor eine Investitionsentscheidung zur Optimierung des Systems getroffen wurde, konnten in Zusammenarbeit mit dem Wartungspartner Versuche im Bereich der Netzoptimierung durchgeführt werden. Dazu zählte unter anderem die Simulation einer Teilabschaltung im Druckluftnetz, wodurch das Einsparpotential durch Vermeidung von Leckageverlusten offensichtlich wurde. Die Abbildung unten zeigt die Simulation dieser Teilabschaltung.

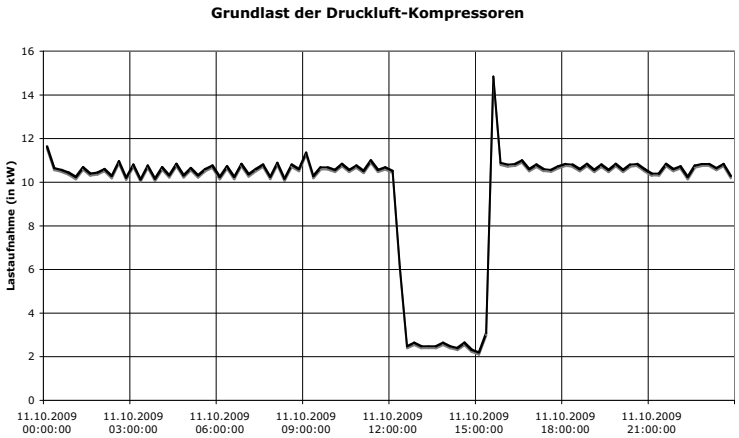


Abbildung 3 Simulation einer Teilabschaltung in produktionsfreier Zeit

Weitere Maßnahmen in Bezug auf die Steuerung sowie die Auslegung des Druckluftnetzes wurden ergriffen. Die Investitionen beliefen sich inkl. Montage- und Umbauaufwand auf ca. €5.000. Die monatlichen Einsparungen liegen aktuell bei knapp €800, so dass mit einer Amortisationszeit von etwas mehr als 6 Monaten gerechnet werden kann.

Im Bereich der Trocknungsanlage für Kunststoffgranulat wurden Analysen im Produktionsbetrieb gefahren. Die ersten Auswertungen offenbarten, dass die Steuerung der Trocknungsanlage nicht im energieeffizienten Modus der taupunktgesteuerten Trocknung gefahren wurde. Die Umstellung der Steuerung war mit einem Zeitaufwand von knapp fünf Minuten realisiert. Im Rahmen der Fehlersuche zu dieser Fehlkonfiguration wurde deutlich, dass die Verantwortung hierfür nicht beim Unternehmen selbst, sondern beim Wartungspartner der Anlage zu suchen war. Dieser hatte im Rahmen der Wartung technisch bedingt eine Umstellung auf eine zeitgesteuerte Trocknung in regelmäßigen Intervallen vorgenommen. Die Abbildung unten zeigt die signifikanten Unterschiede bei den Leistungskurven sowie den Tagesverbrauchsdaten. Die Einsparungen aus der Maßnahme beliefen sich auf ca. 10% pro Tag. Neben den reinen Verbrauchseinsparungen durch die energieeffiziente Steuerung verlängern sich die Wartungsintervalle, da sich die Betriebsstunden im Volllastbereich reduzieren.

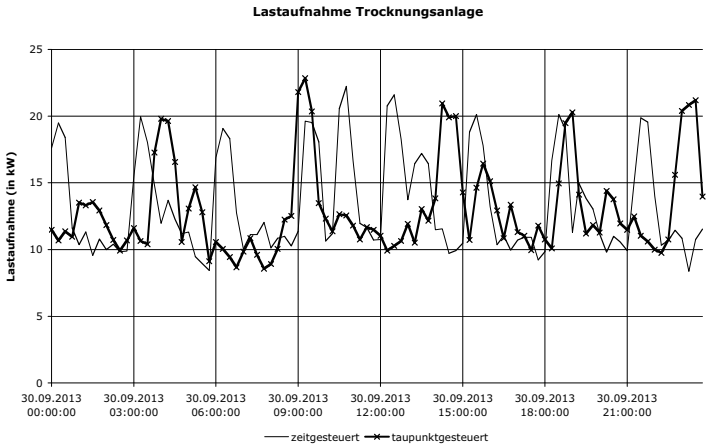


Abbildung 4 Lastaufnahme der Trocknungsanlage - zeit- und taupunktgesteuert

Die dargestellten Eingriffe in diese beiden Unterstützungsprozesse wurden erst durch die transparente Darstellung der Lastaufnahmen erkennbar und für die Verantwortlichen für notwendig erachtet. Die Erwartungshaltung der Verantwortlichen vor Durchführung der Maßnahmen entsprach hingegen stets dem bereits optimierten Zustand. Durch das Aufzeigen dieser signifikanten Unterschiede zwischen der Erwartungshaltung und der Realität wurde das Bewusstsein dahin gehend geschärft, dass nun weitere Messungen durchgeführt werden und die kontinuierliche Überwachung der Maschinen, Anlagen und Prozesse Einzug erhält.

5 Fazit und Ausblick

Die Fallstudie zeigt beispielhaft, wie durch gewonnene Transparenz beim Energieverbrauch Maßnahmen ergriffen werden können, die zu signifikanten Steigerungen der Energieeffizienz führen. Hierfür war die Einführung eines Energiemesssystems notwendig, das die Daten für ein durchgängiges Energiecontrolling zur Verfügung gestellt und die Planung, Steuerung und Kontrolle des Energieverbrauchs ermöglich hat. Um unternehmensweit Energiemanagement betreiben zu können, stellt das Energiecontrolling einen wesentlichen Bestandteil des Energiemanagementsystems dar, da zur Optimierung des Energieverbrauchs die kontinuierliche Überwachung und Verbesserung unabdingbar ist. Energieeffizienzmaßnahmen liefern auf den ersten Blick in jedem Fall Einsparungen im Bereich der Kosten durch die Verbrauchsreduzierungen. Gerade für KMUs werden Emissionen zukünftig jedoch auch eine entscheidende Rolle spielen, sobald der Emissionshandel auch auf sie ausgeweitet wird. Die erreichten Emissionseinsparungen spielen dann neben ihrem ohnehin vorhandenen ökologischen Beitrag auch eine weitere ökonomische Rolle. Durch optimierte Prozessauslegung, wie in der Fallstudie anhand von zwei Prozessen gezeigt, reduzieren sich weitere Kostenbestandteile wie Wartungs- und Verschleißkosten. Energieeffizienzmaßnahmen können somit als einzige sozialverträgliche Kostensparmaßnahme benannt werden, die Nachhaltigkeit in der ökonomischen, ökologischen und sozialen Dimension entsprechen.

Das dargestellte Fallbeispiel zeigt, dass in Unternehmen Potentiale aufgrund von nicht vorhandenem Wissen über die Produktions- und Querschnittsprozesse brach liegen. Transparenz spielt in diesem Zusammenhang eine wesentliche Rolle, da durch sie das Wissen über die eigenen Prozesse wie auch die Verschwendung erweitert wird. Die neu gewonnene Transparenz bietet zudem Einblicke in die Querschnitts- und Unterstützungstechnologien, denen es oftmals an der Zuteilung eines Verantwortlichen mangelt und die deshalb nicht im Fokus der Betrachtung stehen. Die Produktionsprozesse werden als die reinen wertschöpfenden Prozesse angesehen, so dass der Fokus der Betrachtung, sei es in Bezug auf Qualität oder auch Energie, sich dort konzentriert. Es ist jedoch nicht selten, dass die Unterstützungsprozesse der Produktion einen wesentlichen Anteil an den Energiekosten aufweisen, da diese im Dauerbetrieb laufen.

Zu Beginn einer jeden Investition in ein neues IT-System stellt sich die Frage nach dem Return on Investment. Dieser ist in der Regel bei IT-Systemen schwer zu beziffern und auch zu messen. Energiecontrolling-Systeme haben es in diesem Zusammenhang leichter, da sie anhand von belastbarem Datenmaterial ihre Wirksamkeit unter Beweis stellen können. Die Energieagentur NRW rechnet im Bereich der Kommunen vor, dass sich der Einsatz solcher Systeme innerhalb kurzer Zeit rechnet und sogar gegebenenfalls neu aufgebaute Humanressourcen mitfinanziert (vgl. [EA07]). Trotz dieser Erfolgsaussichten bleibt festzuhalten, dass die Anfangshürde immer die Erstinvestition in ein neues System ist. Diese ist verbunden mit Kompetenzzuteilungen und teilweise auch organisatorischen Änderungen. Bei der Systemauswahl kommt es dabei auf eine modulare Struktur an, die einen kontinuierlichen Ausbau ermöglicht und diesen bestenfalls aus den realisierten Einsparungen stemmen kann.

Literaturverzeichnis

[DIN05] Deutsches Institut für Normung: DIN EN 13575, Kommunikationssysteme für Zähler und deren Fernablesung, 2005.

[DIN09] Deutsches Institut für Normung: DIN EN 16001, Energiemanagementsysteme - Anforderungen mit Anleitung zur Anwendung, 2009.

[EA07] EnergieAgentur.NRW: Kommunales Energiemanagement - Grundlagen und Organisation, 2007.

[EDL10] Bundesministerium für Wirtschaft und Technologie: Entwurf eines Gesetzes zur Umsetzung der Richtlinie des Europa ischen Parlaments und des Rates u ber Endenergieeffizienz und Energiedienstleistungen, 2010.

[EEG08] Gesetz fu r den Vorrang Erneuerbarer Energien (Erneuerbare-Energien-Gesetz - EEG), Ausfertigungsdatum: 25.10.2008, 2008.

[FS06] Ferstl, K.; Sinz, J.: Grundlagen der Wirtschaftsinformatik, Oldenbourg Verlag, München, 2006.

[HN04] Hansen, H. R.; Neumann, G.: Wirtschaftsinformatik 1, Lucius & Lucius, Stuttgart, 2004.

[Ho09] Horváth, P.: Controlling. Vahlen Verlag, München, 2009.

[IEA09] International Energy Agency: World Energy Outlook 2009, 2009.

[LLS10] Laudon, K. C.; Laudon, J. P.; Schoder, D.: Wirtschaftsinformatik, Eine Einführung, Pearson Studium, München, 2010.

[Mc09] McKinsey & Company, Inc.: Wettbewerbsfaktor Energie - Neue Chancen für die deutsche Wirtschaft, Frankfurt, 2010.

[Po85] Porter, M. E.: Competitive Advantage: creating and sustaining superior performance. The Free Press, New York, 1985.

[SWB09] Schröter, M.; Weißfloch, U.; Buschak, D.: Energieeffizienz in der Produktion - Wunsch oder Wirklichkeit? Mitteilung aus der ISI-Erhebung. PI-Mitteilung Nr. 51, 2009.

REACH, RoHS, LCA… – Managing several complex material compliance requirements and evaluation of the environmental effects of a product during the whole lifecycle in SAP

Erik Bachman

Corporate Portfolio Management
TechniData AG
Hessenring 13 b
64546 Mörfelden - Walldorf, Germany
Tel. +49 (61 05) 2 82-212

Abstract: Companies in the automotive and high-tech sectors are required by EU legislation such as REACh, ELV, RoHS, and WEEE to provide product-related compliance information. This product data relates to regulated substances used in products. As the level of outsourced production and development of assemblies and individual parts rises, it becomes increasingly necessary to involve suppliers in providing compliance data. The challenge then is to find a declaration model which provides maximum information on substances with only a minimum of queries to the supplier. To secure comprehensive support for compliance management processes TechniData and SAP have developed the SAP REACh Compliance Solution based on the SAP NetWeaver technology. This underpins companies' internal processes and collaborative processes with customers and suppliers.

1 Introduction

The European directives REACh (Registration, Evaluation and Authorization of Chemicals), WEEE, EU- RoHS and the China-RoHS (Pollution Control of Electronic Information Products) build a basis for manufacturer`s requirements in the Electronic Industry. Together with other national and international regulations and laws as well as client-specific requirements, long lists of banned substances have to be managed, controlled and fulfilled by the Supplier Industry.

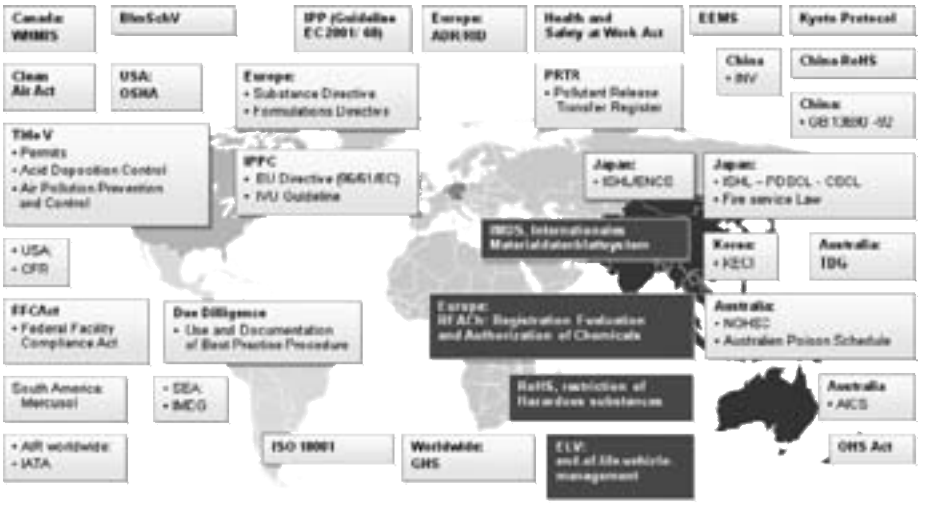


Abbildung 1: The environmental legislation challenges are various [Ba08]

Therefore, automated organisational security between business partners and between company departments becomes essential. In order to reduce process costs, the handling of compliance data has to be transferred from island solutions into companies' main business solution. Data about "not-to-use-substances" are required where product decisions are made daily, i.e. in Laboratory, Purchase and Engineering department but also in Production and Controlling/Finance. The centralised availability of data within companies combined with effective tools to link supplier data as well as customer requirements together is a main challenge for IT-tools today. Based on SAP's business suite, the following text provides a solution map to meet today's business requirements in the area of compliance, environment and safety.

2 Integrated Environmental Compliance Management

National and international regulations, laws, and recommendations are confronting all industries with more and more tasks and obligations. And economic competition is not getting less intense. In nearly all industries new regulations are coming up – especially the REACh directive has an impact on various sectors of the economy and obliges companies to take a cross-section look at their business. These European Directives build the basis for manufacturer`s requirements in the electronic industry. Long lists of banned substances have to be managed, controlled and fulfilled by the supplier industry. Therefore, under economical and time consuming aspects, not only the Engineering and Research departments are placing their focus more and more on these challenges.

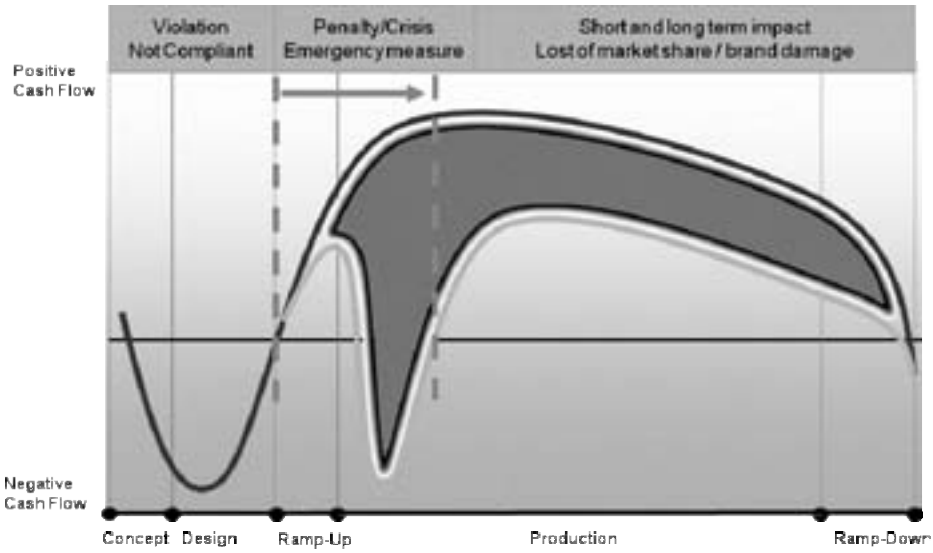


Abbildung 2: Economic aspects: costs of non-compliance [Fo06]

One example shows the resulting consequences for enterprises: In 2007 non-compliance of a product line forced a big toy manufacturer to carry out a number of product recalls. Media coverage did not mention where exactly the problem originated from but still the economical consequences where quite considerable: The incident destroyed 42 million Euro of sales volume. T his shows that product design in line with environmental requirements is no longer a mere subject of internal company policy. It is a substantial factor which is decisive not only for revenue losses, but also for market chances.

Among all named Directives, REACh, EU-RoHS, China-RoHS and Korea-RoHS are of great significance. Companies ignoring the impact on construction imposed by the directive run the risk of non-compliance. The resulting consequences are:

- loss of revenue
- denied markets
- penalties and enforced fees
- competitive regional disadvantage

Manufacturers have financial responsibility for all hazardous substances within products from cradle to grave. REACh compliance affects the entire value chain including the supplier qualification process, and in many cases, fundamental reengineering will be required. According to the business role that is assigned to the participants on the

environmental supply chain, the participants have a special compliance management profile.

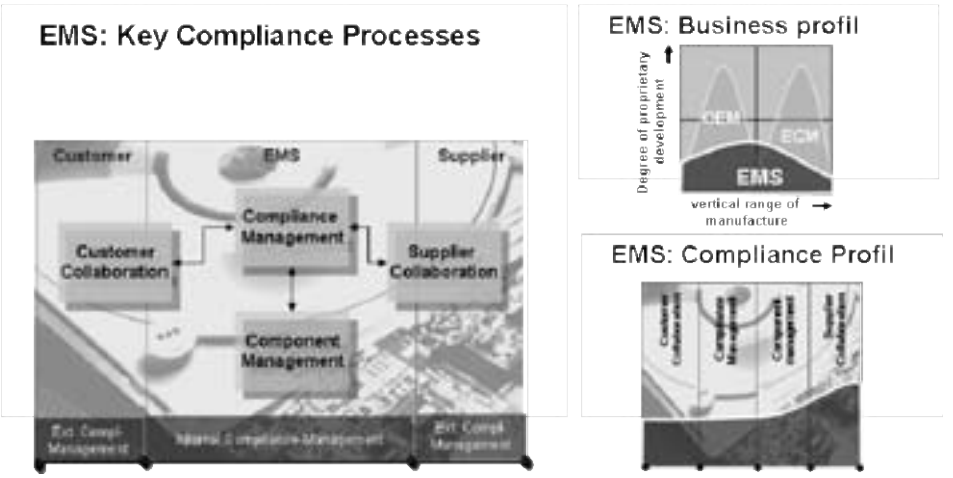


Abbildung 3: Compliance management profile of a Contract manufacturer [Kr06]

Automated organisational security across businesses between departments and between business partners becomes not only necessary but also essential. In or-der to reduce process costs, a transfer of compliance data has to be effected from island solutions into companies' main business solution. Data about "not-to-use-substances" is required where product decisions are made daily.

3 Company-wide declaration objectives are the key to success

When it comes to efficiently meeting all product-related complience requirements the assessment of a declaration objective is the silver bullet. It is vital to decide on a number of fixed criteria for compilation mode and transmission of a product declaration. The relevancy of declaration objectives becomes even more evident when we consider the possibility to choose suppliers from different parts of the world: Gobal choice of business partners implicates a global perspective of compliance standards. In other words: Global sourcing and regional compliance do not necessarily go together. In the process of defining declaration objectives various aspects have to be considered:

3.1 Declaration Objective

The assessment of a declaration objectives can be accomplished via three different approaches:

- Level 1 – qualitative propositions concerning the compliance of a product, e.g. yes/no.
- Level 2 – quantitative information on a defined list of substances, e.g. the possible contingent of a substance in a certain item.
- Level 3 – 100 percent complete declaration about composition and structure of a product.

Today, most companies tend to adhere to Level 2 in an effort to strike the balance between depth of data, frequency of queries and sustainable information value. Quite a number of enterprises have started out with Level 1 for EU-RoHs. However, when China-RoHs came up they were faced with the problem that existing data could not be reused. On the other hand companies which aspire towards implementing Level 3 have to deal with a dilemma: So far the supply chain does not support the need for detailed information to the full extend.



Abbildung 4: Example of a maximal list in DCT

3.2 Maximal List of Declarable Substances

To commit oneself to one of these Levels means to make a decision on how to gain compliance-relevant data. In this context the so-called “maximal list” is a key component. This list is utilized to address an important question: “Is our supplier able to provide us with all the information we need to answer the different legal and client-specific enquiries about our products?” The maximal list is the foundation of every query to the supplier which means that it sets the limits for possible propositions about the composition of a product. Therefore it is important to design the list in a way that matches the intersection of legal and client-specific demands. Additionally, the basis of valuation of each directive must be taken into account. The target must be to harmonise the aspired validation context of the diverse requirements. Thus it should be possible to send a request to the supplier and to get back the comprehensive information needed to decide whether a product matches legal or client-specific requirements. TechniData supports the compilation of intra-corporate maximal lists and the transmission of these lists to the supplier with its Data Collection Tool (DCT), an integral part of the SAP REACh Compliance Solution. The output of this process is a harmonised list of declarable substances which can be utilized company-wide whenever it is necessary to send a request for information to a supplier. The supplier on his part is in a position to answer queries about a product quickly and comprehensively. Importing these lists into an inhouse compliance management system such as the SAP / TechniData solution allows a multiple use of the generated data to answer different queries by legal institutions or by the customer. The effect: Fewer enquiries produce better data with less effort.

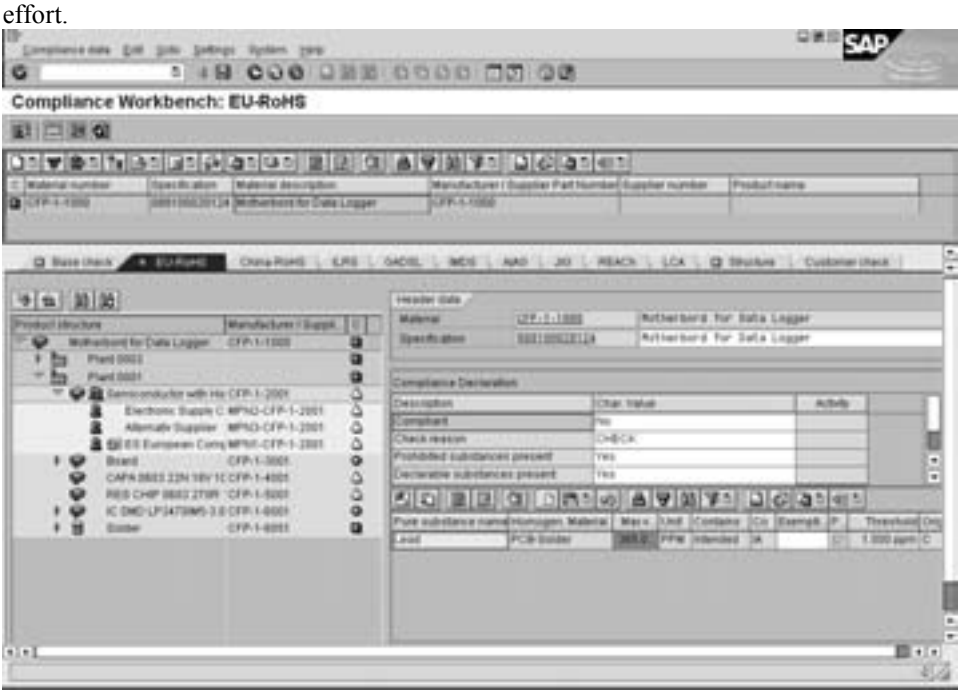


Abbildung 5: One maximal list to address multiple enquiries about the compliance of a product.

The release of the „List of Candidates“ in the context of the REACh directive considerably expands the list of substances which have an impact on product design in the high-tech industry. Additionally, this list must be perceived as an ever changing work in progress – another good reason to assemble company-specific lists of substances and to establish a well-defined communication process with suppliers.

3.3 Influencing factors

The declaration objective is influenced by a number of factors [Ab09]. Different surveys indicate that whenever the amount of notifiable substances increases the rate of return decreases and the quality of the data provided gets poorer. In the process of defining the declaration objective enterprises should weigh different factors against one another:

- Contract conditions – will existing contracts with suppliers correspond with the employment of a declaration objective?
- External data sources – are these both accessible and compatible with the declaration objective (e.g. IMDS)?
- Standards – are there declaration objectives which are commonly used and widely supported within the branch?
- Formats – do existing exchanges formats support the declaration objects (e.g. AIAG, IEEE, IPC1752)?
- Position – does our enterprise occupy a sufficiently strong position within the supply chain to support the requirements?

Mixed declarations – they do not lead to the desired results because the poorest quality decides about how the product is rated as a whole

According to experience companies often misconceive the impact of these influencing factors as being more positive than they really are. Eventually this can prompt companies to define their declaration objective in a much too comprehensive way. This in turn might put a drag on the accomplishment of the project, demanding increased internal and external time and effort. A number of examples show that the declaration objective could not be reached because the targets were set too high from the beginning which in fact stalled the support both from the supply chain and the inhouse staff. At the end of the day running the risk of loosing internal and external support has a much more serious impact than not being able to provide a 100 per cent perfect product declaration.

3.4 Best Practices

Backed by the experience of more than 30 implementation projects which TechniData has accomplished we recommend the following procedure to secure successful and effective compliance management:

- Identification of legal implications in all top target markets
- Identification of high priority customer requirements (primarily A-level customers)
- Identification of internal requirements
- Harmonisation of all these three aspects
- Checkup on the conditions stipulated in contracts with A-level suppliers
- Deduction of the company's declaration objective
- Verification of this objective with regard to the A-products
- Compilation of the company-specific maximal list
- Assessment and evaluation of standards and external data sources
- Selection of preferred exchange format
- Instruction/training of A-level suppliers
- Review of accomplished achievements
- Roll-out of the maximal list to the suppliers
- Evaluation of the rate of return and the quality of the provided information
- Definition of an update cycle (recommended: every 12 months)

4 Learning from Experiences – Automotive Industry and IMDS

The automotive industry has already been confronted by similar tasks. Here – some years ahead of EU-and China-RoHS – the initial issue was to impose a ban on a number of heavy metals. This resulted in the concept of an integrated constituent material management which meanwhile has passed a comprehensive test in practice. Compared to the automotive industry where, due to a manageable number of OEMs, exchange of information can be solved in a relatively simple way by an Internet application, the electrical and electronics industry is faced by an enormous challenge because of its heterogeneous structure – with a large number of OEMs very different in size. The collaboration with customers and suppliers is of great importance; and this must be mapped by the respective software. The key for this is a global harmonized and standardized compliance data exchange format. The benefit of this model is evident in the fact that the automotive industry covers the requirements of REACh for Articles by a simple extension of the GADSL list of substances: Substances listed as “candidates” can be transferred to the GADSL list of substances. The process to exchange material data sheets can be re-used entirely.

5 Challenge for IT-Tools

The collaboration between Laboratory, Purchase De-partment and Engineering Department gives a trans-parent example of a built-in-process to release a new material within the design process. As each depart-ment will have its own k.o. criteria on when not to use the material under substantial aspects, only integrated solutions can streamline the process with overall workflows and collaboration between departments. System support for compliance management recommends a strong commitment and adherence to ISO standards and proved and audited transparency of risk potential, combined with advanced task management.Companies have to be aware that only accurate, reli-able and timely information, modelled in sustainable processes will guarantee and improve image and legal compliance within each step of the daily business. The area of compliance, environment, health and safety, is subject to many national and international regulations, laws and last but not least, customer-specific requirements. The centralised availability of data within companies combined with effective tools to link supplier data as well as customer requirements together is a main challenge for IT-tools today.

6 The Software Has to Cover All Parts of the Process Chain

An example: The use of cadmium is prohibited for an important regional market, but allowed in all other markets on certain conditions. The company has to make sure that none of the products delivered to the regional market contains cadmium. This has an effect on all central departments:

- R&D: Is there a substitute? Can their properties be compared?
- Purchasing: Are there any suppliers which deliver the substitute? How does the procurement of the substitute affect costs?
- Material Laboratory: Does the substitute meet all requirements?
- Production: Which alternatives does the company have?
    - The entire production will be changed to cadmium-free products.
    - Only that part of the production will be changed which is required for the re-gional market (e.g. 10%).
    - Only part of manufacturing will be changed (e.g. one plant, all other pro-duction facilities will continue to work as before).
- Finance/Controlling: Which decision is the most appropriate one under business aspects?
- Logistics: How can it be guaranteed that only cadmium-free products reach the regional market. This includes
- Generation of import documents („Products are cadmium-free as provided by local requirements“)
    - Generation of packaging labels. Here, the company can retrieve constituent material data via software.
    - Supplier auditing and documenting.

Insular solutions are not able to manage this complexity. Here, a solution is required which is completely integrated into the main business processes of a company. The SAP REACH is the SAP solution for Environmental product compliance and covers all requirements imposed by REACh (Chemicals and Non-Chemical Industries) Global-RoHS and material declaration aspects for the high-tech industry. The solution comprises fully integrated business processes from Design for Environment, Supplier Collaboration and Status Tracking for Supplier Components to Restriction Checks and Compliance Analytics. The comprehensive software has been developed by TechniData, which is the exclusive SAP development partner for the Compliance Management solutions SAP Environment, Health and Safety (SAP EH&S) und SAP Environmental Compliance (SAP EC), and as such has developed environmental and safety solutions for almost 20 years now.

7 Optimum Integration

The integration into the SAP environment – for example in purchasing – provides great advantages for the company. The existing ERP system with master data record and vendor assessment is merely extended by environmental aspects; there is no redundancy in collecting and processing data. This also allows the systematic management of hundreds of suppliers. Relevant information includes:

- Which supplier is compliant, which is non-compliant and in which areas. Which substances have been released as of which date and for which period, etc.

- The system collects data not only on prohibited substances, but also quantities under the aspect of adhering to absolute limit values for substances that have to be monitored and are not to be used. This is of decisive importance due to the cumulative effect of a large number of suppliers and components, respectively, for the final product.

- In addition, detailed information on environmental impacts can be collected for the classification as hazardous substance, dangerous goods, etc. Not only data on the recycled content and on disassembling of a product can be stored, but also serial numbers of components, etc.

The system provides an electronic interface for supplier data. This data is structured and then made available e.g. for the production department via a status network (Status Tracking) both on component and constituent material level. And, for example, answers given in supplier surveys with regard to constituent materials used can be imported and further used via document management.

8 Effective Restriction Check

The system allows companies to check the bill of ma-terial of a product quickly and systematically for compliance issues – in both directions, top down and vice versa. Calculations can be made for the entire product. Information on substances to be monitored can be aggregated both on product and component level. Especially, for proofs of use the system shows which products or components contain a certain substance. Since SAP-based software uses automated checking procedures, also product revisions will be considered. This is an advantage which is mainly important in ex-tensive configure-to-order manufacturing. Changes in the operational strategies will automatically be checked for product-relevant environmental and ma-terial data with regard to REACh, Global-RoHS, and other customer specific provisions.

9 Transparency Regarding Regions and Customers

Based on this high transparency, companies can view all substance information of its products vertically and horizontally:

- From a regional point of view considering country-specific requirements
- From a customer point of view, each with different company-specific requirements, re-striction lists, etc.

This shows very quickly whether a company is able to meet the requirements of its key customers, which action is to be taken or if there are new market chances. Thus, the system gives a clear picture of the situation and provides the company with a higher legal security with regard to the organisation of its markets. At the same time, the system is able to consider time lines in the background – for example, certain limit values which are to be met as of a certain date. Based on the customer master file of the SAP system the company always keeps track of to which company or where which deliveries have been made. The integration into Finance/Controlling is not only important for profitability calculations, but also for making provisions.

10 Future prospects

The increase in legal requirements goes hand in hand with a growing demand for corporate responsibility such as the demand for climate-neutrality. The first step towards fulfilling these expectations must be to identify the possible environmental impact of one's own products and plants. Secondly companies have to make optimum use of these findings: The task is to put the focus on measures which are suited to efficiently reduce negative impacts on the environment. The definition of the Carbon Footprint for

products and locations is already an integral part of our roadmap and has already been successfully tested in different pilot projects.



Abbildung 6: Life Cycle Assessment with TechniData CfP [Ab06]

References

[Ba08] E. Bachmann, “The environmental legislation challenges are various”, April 2008

[Fo06] J. Fox, “The Cost of Noncompliance: Calculating the Total Cost of an Environmental Compliance Failure”, page 4, 2006

[Kr06] F. Kröber, “Compliance management profile of a Contract manufacturer”, January 2006

[Ab09] AberdeenGroup, “Materials Compliance for Green Product Development: Balancing Social Responsibility with Profitability”, Dec. 2009

[Ab06] Abele, Eberhard ; Feickert, Stefan ; Kuhrke, Benjamin ; Clesle, Frank-Dieter: Environmental product lifecycle management - customizing the enterprise specific manufacturing processes. In: CIRP International on Life Cycle Engineering Conference, 13. 2006,. , pp. 651-656. ISBN 90-5682-712-X [Buchkapitel] , (2006)

Mobile Anwendungen als Datenquelle für das Stoffstrommanagement

Volker Wohlgemuth, Peter Krehahn, Tobias Ziep

Studiengang Betriebliche Umweltinformatik
HTW-Berlin
Wilheminenhofstraße 75 A
12459 Berlin
Volker.Wohlgemuth@htw-berlin.de
Peter.Krehahn@htw-berlin.de
Tobias.Ziep@htw-berlin.de

Abstract: Sustainable Development ist ein Leitbild, welches von immer mehr Unternehmen angenommen wird. Das Stoffstrommanagement spielt in diesem Zusammenhang eine wichtige Rolle in Bezug auf die effektive Umsetzung dieses Leitbildes. Jedoch stellt der mit dem Stoffstrommanagement einhergehende Aufwand, der insbesondere mit der Datenerfassung verbunden ist, eine Hemmschwelle für Unternehmen, besonders im Segment der kleinen und mittleren Unternehmen (KMU), dar. Zwar existieren Methoden, die den Prozess der Datenaufnahme im Rahmen des Stoffstrommanagement strukturieren und vereinfachen sollen, doch sind diese meist als analoge Verfahren konzipiert und erfordern eine weitere Erfassung bzw. Wiedereingabe in andere, weiterverarbeitende Informationssysteme. Insbesondere im Kontext des Stoffstrommanagements stellen sog. betriebliche Umweltinformationssysteme (BUIS) ein wichtiges Abnahmesystem dar. An dieser Stelle setzt das in diesem Beitrag beschriebene Vorhaben an, welches mobile Anwendungen zur Verfügung stellen will, welche eine digitale Erfassung stoffstromrelevanter Daten für verschiedene Aspekte des betrieblichen Umweltschutzes unterstützen sollen. Dabei wird neben der Standardisierung der Datenerfassung unter anderem auch auf eine direkte Weiterverwendung der gewonnenen Daten in BUIS fokussiert und die Wertschöpfung innerhalb einer Stoffstromanalyse erhöht.

1 Motivation und Problemstellung

Das Leitbild der **Nachhaltigen Entwicklung** (Sustainable Development) prägt in den letzten Jahren die Diskussionen im betrieblichen Umweltschutz, wenn es um umweltschutzbezogene Zukunftsbilder für die entwickelten Industriegesellschaften geht (vgl. [Mor06], S. 18ff). Das Ziel, nicht auf Kosten zukünftiger Generationen und somit nachhaltig zu wirtschaften und zu leben, ist mit dem oben genannten Begriff verknüpft. Insbesondere sind Betriebe aufgefordert, dieses Leitbild bei der Erstellung von Produkten und Dienstleistungen in ihr tägliches Handeln umzusetzen, da diese durch die Herstellung und Nutzung ihrer Produkte und Dienstleistungen direkt an der Umweltbelastung beteiligt sind. Eine Unterstützungshilfe für eine konsequente Umsetzung des Sustainable Developments innerhalb von Betrieben stellt das Stoffstrommanagement dar. Jedoch ist ein aktives Stoffstrommanagement auch immer mir einer Erfassung der Ist-Situation der von dem Betrieb verursachten umweltrelevanten Stoff- und Energieströme verbunden, teilweise insbesondere auch aufgrund gesetzlicher Vorschriften. Dies stellt auch die kleinen und mittleren Unternehmen vor die Herausforderung, umweltrelevante Daten zu erfassen, aufzubereiten und zu speichern und macht es notwendig, umweltrelevante Prozessketten und deren Teilprozesse vor Ort zu untersuchen. Aus Stoffstromanalysen ist bekannt, dass die Datenaufnahme insbesondere in KMU wie in Abbildung 1 gezeigt abläuft, nach der händischen Erfassung der Daten, z.B. während einer Begehung, werden die Daten digitalisiert und danach an die Fachabteilung weiterverteilt, bevor letztere die Daten aufbereitet und ggf. in Informationssysteme überträgt (vgl. [Woh081], S. 213ff). Gekennzeichnet ist dieser Ablauf durch einen großen Anteil an händischer Datenerfassung, bei der die einzelnen Teilschritte einen nicht zu verachtenden Kostenfaktor darstellen und das Risiko einer fehlerhaften Datenaufnahme bzw. –weitergabe bergen. Hier können mobile Anwendungen schon während der Datenaufnahme helfen, Fehler zu vermeiden, indem diese z.B. durch die Bereitstellung erfasster Daten aus vorherigen Zeitperioden dem Nutzer Hinweise über die gewonnenen Daten geben.

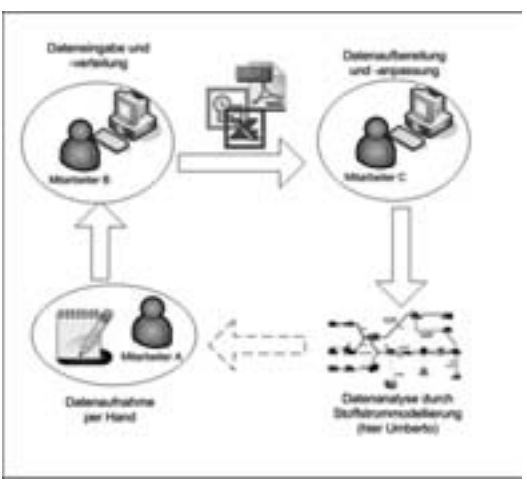


Abbildung 1: Ist-Stand der Datenaufnahme in KMU

2 Stoffstrommanagement

Das Stoffstrommanagement hilft, das Leitbild Sustainable Development konsequent umzusetzen und ermöglicht eine verursachungsgerechte Zuordnung der Stoff- und Energieströme für die betrieblichen Produktionsprozesse. Dabei ist die Auflösung des Unternehmens als Black-Box hin zur White-Box ein grundlegender systemtheoretischer Ansatz, der hier bei der stofflichen Analyse der Produktionsprozesse verfolgt wird (vgl. [WO05], S.140). Durch das Stoffstrommanagement soll eine ganzheitliche Beeinflussung von Stoff- und Energieströmen sowohl in ökologischer als auch ökonomischer Sicht erreicht werden, die zur Steigerung der Ressourcen- bzw. Materialeffizienz und Schaffung nachhaltiger Kreisläufe beiträgt. Das Thema Materialeffizienz und Ressourceneinsparung ist in den meisten KMU ein immer wichtigeres Thema geworden. Trotzdem scheuen sich Unternehmen vor den Investitionskosten, die mit einer Untersuchung der Prozesse verbunden sind (vgl. [Nov09], Interview mit der Geschäftsleitung). Das Instrument des Stoffstrommanagements hat sich aber grundsätzlich als ein erfolgreiches Managementinstrument zur Untersuchung der Unternehmensprozesse durchgesetzt (vgl. [LfU04], S.35).

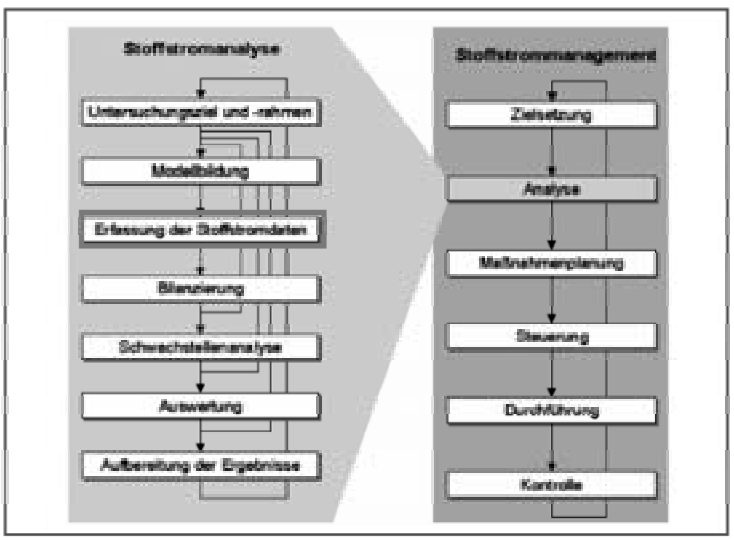


Abbildung 2: Zusammenhang der Prozesse des Stoffstrommanagements und der Stoffstromanalyse, sowie die in diesem Vorhaben fokussierte Phase (rot umrandet) (angelehnt an [Bul00])

Einen wesentlichen Schritt innerhalb des Prozesses des Stoffstrommanagements stellt die Stoffstromanalyse dar (vgl. Abbildung 2). Hier geht es um die Erfassung betrieblicher Stoff- und Energieströme als Informationsgrundlage für Planungsentscheidungen, indem die betrieblichen Stoffströme verursachungsgerecht erfasst, aufbereitet und aggregiert werden.

In diesem Zusammenhang wird jeder Produktionsschritt bzw. –prozess untersucht, da häufig die größten Energie- und Materialverbraucher bzw. Abfallverursacher in Prozessen „versteckt“ sind, die entweder nicht im Fokus der Wertschöpfung stehen bzw. deren Relevanz für die betrieblichen Entscheider nicht transparent ist, weil die mit den Prozessen bzw. Maschinen verbundenen Stoff- und Energieströme nicht bekannt sind bzw. in den betrieblichen Informationssystemen nicht erfasst werden. Dies gilt insbesondere für KMU (vgl. [Woh081] S.213ff).

Zur Erfassung wichtiger Stoff- und Energieströme im Betrieb (Stoffstromdaten gemäß Abbildung 2), d.h. um die relevanten Stoff- und Energieströme entlang der Wertschöpfungskette zu identifizieren und verursachungsgerecht den relevanten Prozessen im Betrieb zuzuordnen, bedarf es einer technischer Unterstützung zur Optimierung der Datenaufnahme, damit die bisherigen Medienbrüche vermieden werden können.

Es ist notwendig die Daten zur Weiterverarbeitung in ein betriebliches Umweltinformationssystem (BUIS) weiterzuleiten, welches diese Daten für die Erfassung, Dokumentation und Bewertung betrieblicher Umweltwirkungen sowie zur Generierung, Planung und Steuerung von Umweltmaßnahmen nutzt. Nur auf dieser Basis ist es möglich, anschließend Erkenntnisse zu erhalten, die für die Weiter- und Neuentwicklung der betrieblichen Prozesse oder Produkte unter sowohl Kostengesichtspunkten als auch stofflichen Aspekten genutzt werden können. Auf der Basis der Stoffstromanalyse könnten z.B. Maschinen mit hohem Energie-, Dampf- oder Wasserverbrauch bzw. mit hohem Ausschuss identifiziert werden, die durch „bessere“ bzw. materialeffiziente Maschinen ersetzt werden könnten oder den Ausschuss durch eine Kreislaufführung reduzieren. Ggf. kommt der Betrieb durch die Stoffstromanalyse auch zur Erkenntnis, den gesamten Prozess unter Material- oder Energieeffizienzgesichtspunkten umzustrukturieren. (vgl. [Woh05], S. 134ff)

Probleme beim Stoffstrommanagement

Der Aufwand der Datenerfassung der Stoff- und Energieströme wird als recht hoch angesehen, da die genannten umweltrelevanten Informationen erst im Betrieb zusammengesucht werden müssen.

Ein weiteres Problem besteht darin, dass KMU nicht wissen, welche Daten in Zusammenhang mit Stoffstromanalysen zur Identifizierung von Effizienzpotentialen überhaupt erfasst werden müssen. Hier gibt es jedoch einige Ansätze, die KMU unterstützen sollen, eine Idee zu entwickeln, welche Daten in welcher Form erfasst werden müssen, z.B. Checklisten, die Methoden Value Stream Mapping, Material Stream Mapping[1]. Allerdings basieren diese Methoden ausschließlich auf der Erfassung der Stoff- und Energieströme durch Fragebögen in Papierform. Diese Erfassungsproblematik von Stoffströmen verschärft sich häufig noch, wenn man spezielle, kaum transparente Stoffströme, wie Energie, Wasserdampf, Druckluft oder Wasser betrachtet.

[1] vgl. http://www.mesor.de/mesor-kompass/material-stream-mapping/, abgerufen am 25.04.2010

Diese Daten können über Messzähler abgelesen werden, allerdings erfolgt die Ablesung häufig noch, insbesondere in KMU, mit Zettel und Papier. Daher müssen diese Daten nochmals in ein IT-System eingegeben werden. In größeren Unternehmen werden zur Speicherung dieser Stoff- und Energieströme zwar schon vielfach IT-Systeme eingesetzt (z.B. ERP-Systeme oder FM-Systeme), jedoch müssen diese Informationen noch aufbereitet werden, um in ein BUIS übertragen werden zu können (vgl. [Wol08], S. 40).

3 Projekt MOEBIUS

Das Projekt vom BMBF geförderte Vorhaben MOEBIUS (moebius.htw-berlin.de) verfolgt als Zielstellung, die Entwicklung mobiler Anwendungen für die Datenaufnahme im Rahmen von Stoffstromanalysen. Diese Daten sollen dann über Schnittstellen BUIS zu Verfügung gestellt werden. Dabei sollen die in Abbildung 3 gezeigten Aspekte explizit unterstützt werden und der Einsatz des Mobile Computing und deren Techniken als Lösungsansatz untersucht werden. Insbesondere auch Fragestellungen der Gebrauchstauglichkeit mobiler Anwendungen in diesem Kontext sollen betrachtet werden. Als Vorteile der mobilen Datenerfassung in diesem Kontext werden insbesondere folgende Punkte betrachtet:

- **Kosteneinsparung**: Die Datenerfassung stellt einen mitunter beträchtlichen Aufwand für ein Unternehmen dar, besonders im Rahmen des Stoffstrommanagements, welcher zum Teil abschreckend auf KMU wirkt.
- **Standardisierung:** Die mobile Datenerfassung umweltrelevanter Daten erleichtert die Datenhaltung und verringert deren spätere Aufbereitung und Verteilung der Daten.
- **Rückmeldefunktion**: Bei einer mobilen Datenerfassung kann ein Vergleich vor Ort mit Vorgängerdaten erfolgen und somit ein schnelle Ursachenforschung für Fehler bei der Zählerablesung in die Wege geleitet werden.
- **Verfügbarkeit der Daten:** Diese steigert die Motivation der Mitarbeiter, weiter Daten zu erheben und somit den betrieblichen Umweltschutz optimal in dem eigenen Unternehmen zu fördern, Daten, die erfasst wurden und nicht weiter verwendet werden, weil sie z.B. in Aktenschränken „schlummern“, demotivieren den Mitarbeiter eine ähnliche Erfassung noch einmal mit der gleichen Sorgfältigkeit vorzunehmen.
- **Kommunikation und Transparenz:** Diese wird zwischen der Umweltabteilung und den Mitarbeitern durch die ständige Integration der Mitarbeiter als auch der beteiligten Fachabteilungen in die innerbetrieblichen Umweltmanagementprozesse erhöht.

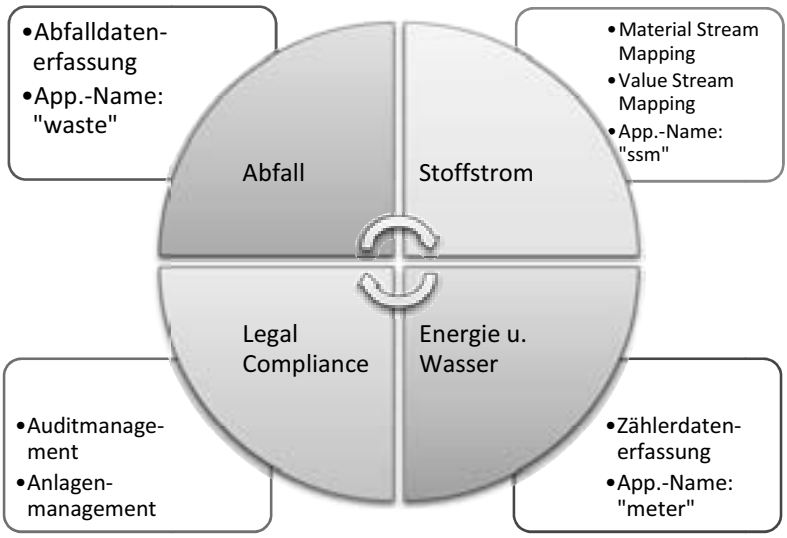


Abbildung 3 Aspekte, die im Rahmen des MOEBIUS-Vorhabens durch mobile Anwendungen unterstützt werden sollen.

4 Material Stream Mapping, als mobile Beispielapplikation

Die Methode des Material Stream Mapping stellt eine strukturierte Aufnahme von Input und Outputdaten entlang einer Prozesskette dar. [2]Es werden die Daten bzgl. Material und die dazugehörigen Kosten erfasst und nach Einsparpotenzialen, anhand der quantitativ und monetär größten Positionen, gesucht. Ziel der mobilen Applikation „ssm“ ist es unter anderem diese, bisher analoge, Methode zu digitalisieren und auf einem PDA ähnlichem Gerät zur Verfügung zu stellen. Dabei wird mit Hilfe des .Net Compact Frameworks 3.5, speziell für MS Windows Mobile-Geräte, entwickelt.

Abbildung 4 veranschaulicht den Prototyp für die Material Stream Mapping Methode, wobei der Nutzer die Möglichkeit besitzt, neben den allg. Unternehmensdaten Daten während einer Begehung aufzunehmen. Die Daten werden lokal auf dem Gerät in einer SQLite-Datenbank gespeichert, wodurch eine ortsunabhängige Persistenz gewährleistet wird. Darüberhinaus können die erfassten Daten in gängigen Exportformaten zur späteren Verwendung verfügbar gemacht werden. Zum Beispiel wird ein spezielles Schnittstellenstandardformat für den Austausch umweltrelevanter Daten, die sog. PAS1025 (vgl. [Wit06]), in Form einer genormten XML-Datei bereitgestellt, welche die direkte Weiterverarbeitung dieser Daten in BUIS ermöglicht.

[2] vgl. http://www.mesor.de/mesor-kompass/material-stream-mapping/, abgerufen am 25.04.2010

Bei der Dateneingabe kann die Kamera als Barcode-Scanner eingesetzt, welche es ermöglicht, Q/R-Codes als Identifikation für die einzelnen Zähler (Zählererfassung) oder Behälter zu nutzen, z.B. in der Abfalldatenerfassung, wodurch die Zuordnung der Verbrauchsdaten zu den Stellen, an denen sie anfallen, optimiert wird und Falscheingaben minimiert werden.



Abbildung 4 Screenshot Datenaufnahme mit der MSM Methode

5 Ausblick

Zum jetzigen Zeitpunkt sind durch das Projekt MOEBIUS die Prototypen „meter“, „ssm“ und „waste“ (vgl. Abbildung 3) zu großen Teilen fertiggestellt und es werden die vorgesehenen Exportformate analog zur folgenden ersten Testphase entwickelt. Es ist zum August 2010 die Fertigstellung aller mobilen Anwendungen als sog. Release Candidates (RC) geplant und neben den Anwendungen wird der C#-Code auf der Internetplattform veröffentlicht, da es sich hier um ein Open-Source Projekt handelt. Als nächster Schritt wird eine Auswertungsanwendung als Desktopanwendung auf Grundlage des Empinia-Frameworks erstellt, welche KMU die Möglichkeit bietet, eigene Schnittstellen zu firmenspezifischen Informationssystemen zu implementieren, darüberhinaus aber auch die Auswertung und graphische Aufbereitung der Daten ermöglicht. Mit der Veröffentlichung der Ergebnisse der zuvor genannten Projektphase ist zum April 2011 zu rechnen. Für weitere Informationen wird an dieser Stelle auf die Internetplattform moebius.htw-berlin.de verwiesen.

Danksagung

Das Forschungsprojekt wird finanziert durch das Bundesministerium für Bildung und Forschung (BMBF). Die Autoren bedanken sich für die finanzielle Unterstützung.

Bibliographie

[Bul00] H.J. Bullinger, S. Beucker: Stoffstrommanagement und Betriebliche Umweltinformationssysteme (BUIS) liefern neue Impulse für das Umweltcontrolling (Stuttgart, 2000).

[LfU04] LfU: Energie und Stoffstrommanagement – Ein positives Fazit für die Unternehmen und die Umwelt. (Karlsruhe, 2004).

[Mor06] R. Morana: Management von Closed-Loop Supply Chains. (Wiesbaden, 2006).
[NOV09] Geschäftsleitung NOVAPAX: Bestandsaufnahme der Istsituation bzgl. des Stoffstrommanagements (Berlin, 2009).
[Wit06] A. Witte, V. Wohlgemuth, I. Mersiowsky: Datenaustausch von Ökobilanzdaten zwischen einem SAP-Business Warehouse und dem BUIS Umberto unter Nutzung der PAS 1025 und ein Vorschlag zur Erweiterung der PAS 1025 um Bilanz- und Kennzahleninformationen. (2006).
[Woh05] V. Wohlgemuth: Komponentenbasierte Unterstützung von Methoden der Modellbildung und Simulation im Einsatzkontext des betrieblichen Umweltschutzes (Aachen, 2005).
[Woh08] V. Wohlgemuth, M. Mäusbacher: Analyse und Implementierung von Import- und Exportschnittstellen zwischen dem stoffstromorientierten BUIS Umberto und SAP (Aachen, 2008).
[Woh081] Volker (Hrsg.) Wohlgemuth, Sevim Demir, Martin Lotter: Durchführung einer Stoffstromanalyse als Ausgangspunkt für eine Potenzialanalyse mit dem Schwerpunkt Material und Energieeffizienz bei der PanTrac GmbH (Berlin, 2008).

Bi-nationale Promotionsprogramme in der Betrieblichen Umweltinformatik am Beispiel des DEEBIS-Net Programms

Liane Haak, Nils Giesen, Jorge Marx Gómez

Universität Oldenburg, Fakultät II, Department für Informatik
Wirtschaftsinformatik I / VLBA
Ammerländer Heerstr. 114-118
D-26129 Oldenburg
{haak, giesen, marx-gomez}@wi-ol.de

Abstract: In diesem Beitrag wird kein einzelnes konkretes Forschungsergebnis, sondern ein ganzes Forschungsnetzwerk im Bereich Betrieblicher Umweltinformationssysteme (BUIS) vorgestellt: das DEEBIS-Net. Es handelt sich hierbei um ein bi-nationales PhD-Programm zwischen Deutschland und Kuba, welches aus der Universität Oldenburg heraus geleitet wird. Inhaltliche Schwerpunkte liegen im Bereich Very Large Business Applicatons (VLBA) und BUIS. DEEBIS-Net ist ein Forschungsnetzwerk mit interessanten Anknüpfungspunkten für andere Wissenschaftler, besonders die Forschungsthemen im BUIS Bereich, aber auch das Netzwerk selber sollen daher hier kurz vorgestellt werden.

1 Einleitung und Hintergrund

Neben der eigentlichen Forschungsförderung im Rahmen von Forschungsprojekten, ist die gezielte Förderung der Promotion ein wichtiges Ziel vieler Förderungseinrichtungen und –initiativen zum Auf- und Ausbau des wissenschaftlichen Nachwuchses. Genannt seien hierbei beispielsweise die Promotionsförderung der Deutschen Forschungsgemeinschaft (DFG)[1] in Form von Stipendien, Graduiertenkollegs oder im Rahmen von Sonderforschungsbereichen und spezielle Förderprogramme der Europäischen Union (EU) wie z.B. das Marie Curie Programm[2]. Des Weiteren stellen deutsche Universitäten für die Ausbildung von wissenschaftlichem Nachwuchs regelmäßig sogenannte Qualifikationsstellen bereit und verschiedene Institute und Stiftungen wie z.B. die Deutsche Bundesstiftung Umwelt (DBU)[3] bieten Möglichkeiten für besonders begabte Doktoranden sich um persönliche Stipendien zu bewerben. Die Ausrichtung der Programme variiert von einer rein nationalen bis hin zu internationalen Maßnahmen.

Besonders die internationale Promotionsförderung und ein regelmäßiger wissenschaftli-

[1] www.dfg.de
[2] http://cordis.europa.eu/fp7/mariecurieactions/home_en.html
[3] http://www.dbu.de/

cher Austausch wird vom Deutsche Akademischen Austauschdienst (DAAD)[4] fokussiert. Die geförderten Aktivitäten reichen vom Export von Studiengängen, Hochschulpartnerschaften mit Entwicklungsländern, Stipendien, Summerschools und Workshops bis hin zu Alumni-Netzwerken und dienen dem Austausch in Forschung und Lehre auf internationaler Ebene. Seit 2001 existiert das Programm „Promotion an Hochschulen in Deutschland (PHD), welches sich mit verschiedenen Förderlinien an der Promotionsförderung in Deutschland beteiligt. Bereits in der ersten Runde (2001-2006) konnte das Programm mit vergleichsweise geringen Mitteln (6 Mio. Euro für ca. 50 Programme: entspricht durchschnittlich 120.000 Euro pro Projekt) rund 3000 Doktoranden (ca. 40% Frauen) unterstützen und eine durchschnittliche Promotionsdauer realisieren [BOD06]. Dieses vom DAAD selbst als Erfolgsgeschichte bezeichnete Programm wurde dann in 2008 erneut zur Förderung ausgeschrieben mit dem Fokus auf bi-nationale Programme (PhD-Net), in dessen Rahmen das hier vorgestellte bi-nationale Promotionsnetzwerk DEEBIS-Net gefördert wird.

Der DAAD fördert hierzu aus Mitteln des Bundesministeriums für Bildung und Forschung die Zusammenarbeit deutscher und ausländischer Hochschulen auf dem Gebiet der Doktorandenausbildung. Ziel des Programms bi-nationales Promotionsnetzwerk (PhD-Net) ist es, ausländischen Doktoranden grenzüberschreitende, bi-nationale Promotionen zu ermöglichen und hochqualifizierte ausländische Nachwuchswissenschaftler für eine wissenschaftliche Karriere in Deutschland und/oder für den Aufbau wissenschaftlicher Kooperationsbeziehungen mit Deutschland zu gewinnen. Grundsätzlich werden dabei drei voneinander unabhängige Förderlinien angeboten: Anbahnungsreisen, Workshops/Summer Schools und Promotionsprogramme.

2 DEEBIS-Net: ein bi-nationales PhD-Programm

DEEBIS-Net (Doctoral Education in Environmental and Business Information Systems) ist eine Promotionsprogramm, gefördert gemäß der Förderlinie 3 als bi-nationales PhD-Programm und dient der Schaffung geeigneter Rahmenbedingungen für Promotionsvorhaben ausländischer Doktoranden, die in Kooperation mit ausländischen Partnerhochschulen durchgeführt werden. Die Heimathochschule des/der Promovierenden soll in die Betreuung einbezogen werden, indem entweder ein Hochschullehrer der Heimathochschule als Co-Betreuer/Mentor eingesetzt und regelmäßig über den Fortschritt der Promotion unterrichtet wird oder aber indem die Promotion in Form einer bi-nationalen oder Sandwich-Promotion durchgeführt wird. Im Falle einer bi-nationalen oder Sandwich-Promotion wird von den Partneruniversitäten ein Konzept für

- eine gemeinsame Ausgestaltung eines Promotionsprogramms,
- eine gemeinsame Betreuung bis hin zur Verleihung gemeinsamer Abschlussgrade (Doppelpromotion),
- den Einsatz ausländischer Gastdozenten/innen,
- die Organisierung von Auslandsaufenthalten der Doktoranden/innen,
- längerfristige Karrierewege der Doktoranden einschließlich ihrer Perspektiven

[4] www.daad.de

nach der Promotion (Karriereentwicklungskonzept) und
- die Schaffung von Praktikumsplätzen/Forschungsmöglichkeiten für Absolventen in Industrie und/oder Universität

erwartet.

DEEBIS-Net vernetzt als bi-nationales Promotionsprogramm im Bereich der Wirtschaftsinformatik und Betrieblichen Umweltinformatik die Carl von Ossietzky Universität Oldenburg (Abteilung Wirtschaftsinformatik I / VLBA) und ihre kubanischen Partneruniversitäten, die Technische Universität Havanna (Instituto Superior Politécnico José Antonio Echeverría, CUJAE, Havanna) und die Universität von Santa Clara (Universidad Central de las Villas (UCLV) Santa Clara). Des Weiteren ist noch die Abteilung Wirtschaftsinformatik der Otto-von-Guericke Universität mit zwei Doktoranden im Projekt vertreten. Ursprünglich ist das Programm mit acht Doktoranden gestartet, wurde aber in 2009 um einen weiteren Doktoranden erweitert, da abzusehen war, dass eine Doktorandin aus persönlichen Gründen nicht mehr an den Auslandsaufenthalten teilnehmen kann. Sie ist aber weiterhin in das Netzwerk integriert, insbesondere da ihr Thema im BUIS Schwerpunkt liegt und viele Anknüpfungspunkte für gemeinsame Fragestellungen zwischen den Doktoranden bietet.

Das Ziel des Projektes ist es, besonders begabten Doktoranden in kürzerer Zeit eine exzellente wissenschaftliche Promotion zu ermöglichen, um darauf aufbauend einen weiteren Karriereweg sowohl in wissenschaftlichen als auch in wirtschaftlichen Bereichen in Führungspositionen einzuschlagen zu können. Die bisherigen Randbedingungen in der Informatik, die sich u.a. durch lange Promotionszeiten und hohe fachliche Diversifizierung (Inselpromotion) auszeichnen sollen durch gezielte Maßnahmen im Rahmen dieses Projektes für die Bereiche Wirtschaftsinformatik und Betriebliche Umweltinformatik deutlich verändert werden. Zudem soll der in der Wirtschaftsinformatik unabdingbare Praxisbezug gestärkt werden. Das mit einer zunächst 3-jährigen Laufzeit am 01.08.2008 gestartete Promotionsprogramm ist international ausgerichtet und richtet sich sowohl an deutschsprachige als auch an englischsprachige Doktoranden.

Wie in den anderen geförderten bi-nationalen PhD-Netzwerken[5] liegt der Schwerpunkt darauf eine anerkanntes und abgesichertes Promotionsverfahren mit schnellen Promotionszeiten und intensivem Forschungsaustausch zwischen Universitäten in zwei Nationen zu implementieren. In DEEBIS-Net wird eine gegenseitige eindeutige Anerkennung der Abschlüsse angestrebt. Erste Maßnahmen wie z.B. notwendige Änderungen in den Promotionsordnungen wurden hierzu durchgeführt. Durch regelmäßige wechselseitige Forschungsaufenthalte von Doktoranden und Wissenschaftlern ist zudem der kontinuierliche Austausch sichergestellt. Erste gemeinsame Publikationen sind schon jetzt ein erfolgreiches Resultat dieser Zusammenarbeit. Begleitend werden Seminare und andere Fortbildungsmaßnahmen für die Doktoranden angeboten, um sich inhaltlich, aber auch persönlich (Networking und Karrieremanagement) weiterzubilden.

Dennoch ist DEEBIS-Net sowohl in der Durchführung kaum mit den anderen vergangenen und aktuellen Projekten in diesem Förderprogramm vergleichbar. Zum einen ist die

[5] http://www.daad.de/hochschulen/internationalisierung/phd-net/08473.de.html

Organisation wie auch die Vereinheitlichung der Promotion bei einem Partner wie Kuba deutlich komplexer als mit einem Europäischen Partner mit ähnlichem Ausbildungssystem und mit vielen Unsicherheiten behaftet. So ist die Jahresplanung der Aktivitäten oft an die Entscheidungshorizonte in Kuba anzupassen, welche wiederum stark von der nationalen Politik geprägt sind. Hinzu kommen technische Probleme in der Kommunikation und kulturelle Unterschiede, insbesondere das unterschiedliche Verständnis von Zeit, sind zu überbrücken. Zum anderen liegen die Differenzen zu ähnlichen Projekten dieser Art im grundsätzlich unterschiedlichen Verfahren er Promotion begründet. Während die Promotion in Deutschland in der Hoheit der Hochschulen und dort der Fakultäten liegt, handelt es sich in Kuba um ein nationales Verfahren. Dies erhöht die Komplexität im Abstimmungsprozess hinsichtlich eines gemeinsamen Abschlusses deutlich gegenüber den anderen PhD-Net Projekten, die fast ausnahmslos in Europa verankert sind. Eine gute Basis stellen daher die bereits nach deutschem Vorbild exportierten Studiengänge in der Wirtschaftsinformatik und der Betrieblichen Umweltinformatik dar. Diese Thematiken bringen aber ebenfalls neue Herausforderungen mit in das Projekt, denn es handelt sich insbesondere beim letzten um eine recht junge Disziplin. Im folgenden Abschnitt sollen insbesondere die Themenstellungen in der Betrieblichen Umweltinformatik daher näher vorgestellt werden.

3 Forschung in DEEBIS-Net

Die Forschung im DEEBIS-Net Forschungsnetzwerk orientiert sich an den Forschungsschwerpunkten der beteiligten Universitäten und zielt daher besonders auf Themenstellungen im Bereich großer betrieblicher Anwendungssysteme (Very Large Business Applications (VLBA)) und Betriebliche Umweltinformationssysteme (BUIS) ab. Im Folgenden werden nach einer kurzen Einführung die aktuellen Forschungsthemen und Fragestellungen im BUIS Kontext der Doktoranden in DEEBIS-Net vorgestellt.

3.1 Betriebliche Umweltinformationssysteme (BUIS)

Betriebliche Umweltinformationssysteme fassen die Klasse der organisatorisch-technischen Systeme zur ganzheitlichen Erhebung, Aufbereitung und Verarbeitung umweltrelevanter Informationen eines Unternehmens zusammen (vgl. [HOJ93]). Dabei werden die Wirkräume dieser Systeme von der messtechnischen Ersterfassung bis hin zur Bereitstellung, zur Weiterverarbeitung oder Publikation betrachtet. Betriebliche Umweltinformationssysteme dienen nicht allein als spezialisierte Ausprägung vorhandener betrieblicher Informationssysteme, sondern haben in ihrer Entwicklung (vgl. [PVO03]) neben Bezügen zu ökonomischen und ökologischen Fragestellungen als Werkzeuge der betrieblichen Umweltinformatik zur Entwicklung dieses Forschungsbereiches als eigenständige wissenschaftliche Disziplin beigetragen. So lassen sich die vorhandenen Fragestellungen und Forschungsthemen der Betrieblichen Umweltinformatik nicht mit den vorhandenen Lösungen der Wirtschaftsinformatik beantworten. Neben der Kenntnis über betriebswirtschaftliche Zusammenhänge und informationstechnische Vorgänge werden für die Lösung der vorhandenen Forschungsfragen in diesem Forschungsbereich zusätzlich ein hohes Verständnis für die ökologischen Auswirkungen,

beispielsweise durch gezielte Simulation (vgl. [WBP01]), des unternehmerischen Handelns sowie ein ganzheitlicher Blick unter dem Aspekt der Nachhaltigkeit benötigt.

3.2 Forschungsschwerpunkte im BUIS Bereich

Der Forschungsschwerpunkt der Betrieblichen Umweltinformatik (in engl. auch Environmental Management Information Systems (EMIS) genannt), insbesondere im Bezug zu großen betrieblichen und überbetrieblichen Informationssystemen, behandelt Softwaresysteme die in vielfältiger Art und Weise bei der Sammlung, Verarbeitung, dem Austausch und der Berichterstattung umweltrelevanter Daten unterstützen: Betriebliche Umweltinformationssysteme (BUIS). Diese Systeme haben den gleichen Applikationsrahmen wie betriebliche Informationssysteme der Wirtschaftsinformatik (vgl. [RAU99]). Dabei liegt der Fokus sowohl innerbetrieblich, z.B. in der notwendigen Integration von BUIS und ERP-Systemen (vgl. z.B. [FNM09]), als auch im Austausch mit Teilhabern in Industrie, Behörden oder gesellschaftlichen Institutionen. Dies hat zu einem starken Zuwachs an Aufmerksamkeit und steigender Popularität dieses jungen Forschungsschwerpunktes geführt. Da Unternehmen sich neuen Herausforderungen stellen müssen - wie dem Klimawandel, dem wachsenden öffentlichen Interesse in Themen wie Umweltschutz und den Umweltauswirkungen unternehmerischen Handelns - können sie BUIS dazu nutzen, diese Auswirkungen zu identifizieren, sowie Maßnahmen zur Vermeidung und Reduzierung des umweltschädlichen Aktivitäten auszubilden und durchzuführen. Dazu ist es Zudem wird der Bedarf für neue und aufwendige Umweltinformationssysteme weiterhin durch externe Faktoren wie die Erfüllungspflicht für gesetzliche Auflagen oder den Druck durch gesellschaftliche Beobachtung gestärkt. Hier spielt beispielsweise die Nachhaltigkeitsberichterstattung eine prägnante Rolle, die jegliche Stakeholder zum Ansprechpartner hat (vgl. [IMG08]). Während viele bestehende und von Unternehmen genutzte BUIS auf einzelne und spezielle Aufgaben fokussiert bzw. durch Gesetzgebung oder Output-Orientierung gekennzeichnet sind, setzen neue Forschungsansätze auf einen Ausbau der bestehenden Arbeiten und einer Entwicklung holistischer Sichtweisen für Betriebliche Umweltinformationssysteme. Dies ergibt viele offene Fragestellungen, wie der weiteren Entwicklung von BUIS, Möglichkeiten eine ganzheitliche und nachhaltige Sicht auf Systemlandschaften zu gewinnen, Strategien zur Integration von Umweltinformationen oder die Verbesserung der Stakeholder-Beziehungen.

Die Umweltwirkungen des betrieblichen Handelns stellen dabei das verbindende Element der Promotionsthemen innerhalb von DEEBIS-Net dar. Unter der Thematik der „Organisations Environmental Performance Indicators (OEPI)" werden Lösungsmöglichkeiten für die Erfassung und Verbreitung von betrieblichen Umweltkennzahlen mit Hilfe eines semantischen standardisierten Konzeptes zur Aufbereitung der Kennzahlen untersucht. Dabei werden unter anderem verschiedene Dienste und Tools entwickelt, die eine Verbreitung der Umweltkennzahlen ermöglichen sollen (bspw. über mobile Endgeräte, Internetzugriff oder die Integration in bestehende Systeme mittels Webservices).

Die Promotion zu „Procedures for measuring social impact of Corporate Social Responsibility (CSR)" untersucht ganzheitliche Verfahren zur Messung der sozialen, aber auch der ökologischen, Auswirkungen von CSR. Insbesondere die Auswirkungen

und Messbarkeit der vorgenommenen Entwicklung soll mit Hilfe verschiedener Methoden erfasst werden und kann so u.a. zur Optimierung genutzt werden. In der betrieblichen Betrachtung wird CSR zudem weniger als rein geschäftliches Mittel, sondern als gesellschaftsbildende Möglichkeit aufgefasst, vgl. [PNU05]. Eine Integration der CSR in die strategischen Ziele des Unternehmens soll dabei die langfristige Absicherung von Forderungen der Gesellschaft ermöglichen, vgl. [SUL01]. Die Anforderungen an Unternehmensleitungen zielen dabei vermehrt auch auf soziale und ökologische Bereiche, was Transparenz in den Entscheidungen und der Kommunikation erfordert, vgl. [ART02]. Dies erfordert strukturierte Vorgehensweisen, die eine Anleitung für die Entwicklung solcher umfassender Strategien, die Wahl der Methoden sowie eine effektive und effiziente Integration der CSR ermöglichen.

Die laufenden Arbeiten zu dem Thema „Environmental Information Management“ sind Teil der wissenschaftlichen Diskussion zur Erarbeitung eines holistischen Umweltinformationsmanagement, welches als Unterbau oder verbindendes Element der eher isolierten verschiedenen Speziallösungen vorhandener betrieblicher Umweltinformationssysteme dienen soll. Dabei sollen weniger die technischen Möglichkeiten für ein übergreifendes Umweltinformationssystem (vgl. [GHM09]) im Kern der Betrachtung stehen als Methoden und Prozesse zur Integration des Umweltmanagement auf der strategischen Entscheidungsebene.

Eine weitere Arbeit widmet sich dem Thema „Improving Stakeholder relations by using Web 2.0 technologies for target oriented and dialogue based sustainability reporting” und untersucht, wie mit Hilfe des Web 2.0 die Konsumenten von Nachhaltigkeitsberichten in die Bewertung, Analyse und Kommentierung von diesen Berichten einbezogen werden können. Die so entstehende bi-direktionale Kommunikation soll einen direkten Dialog zwischen Unternehmern und Stakeholdern fördern. Dabei stehen insbesondere die unterschiedlichen Informationsbedürfe in der Betrachtung, was bei der bisherigen Berichterstattung nicht betrachtet wird (vgl. [SGI08]). Zwar gibt es in größeren Unternehmen eine Zunahme der nachhaltigkeitsbezogenen Berichterstattung, insbesondere im Zusammenhang mit CSR (vgl. [BGH07]), doch im Bereich der kleinen und mittleren Unternehmen sind diese Aktivitäten deutlich geringer ausgeprägt ([KEH08]).

4 Zusammenfassung und Fazit

Das bi-nationale Forschungsnetzwerk DEEBIS-Net ist nun bereits im zweiten Jahr nach Beginn des Programms. Die aktuellen Forschungsaufenthalte sind geplant und haben bereits begonnen. Trotz der beschriebenen Schwierigkeiten überwiegen derzeit die guten Erfahrungen und nach einer ersten Zusammenarbeit wird nun für die Verlängerung, eine Erweiterung und die Integration weiterer Doktoranden geplant. Diese sind bereits in der Abteilung Wirtschaftsinformatik thematisch verankert und sollen gemeinsam mit den DEEBIS-Doktoranden von den Vorzügen des Programms profitieren. Hierbei handelt es sich in erster Linie um eine inhaltliche Integration, weniger eine finanzielle Unterstützung. Ein erster Austausch ist bislang schon während der jeweiligen Forschungsaufenthalte erfolgt, kann nun aber durch die neuen Maßnahmen weiter vertieft werden. Es handelt sich hierbei im Wesentlichen um vier DAAD-Stipendiaten aus Kuba und weitere

Stipendiaten aus Syrien. Besonders wertvoll ist dabei auch die Zusammenarbeit mit Betreuern anderer deutscher Universitäten, wodurch der Netzwerkeffekt von DEEBIS-Net deutlich verstärkt wird und längerfristige gemeinsame Forschungsaktivitäten aufgebaut werden können.

Anzumerken bleibt, dass insbesondere aufgrund politischen, kultureller und organisatorischer Unterschiede einem derartigen Netzwerk durchaus gewisse Herausforderungen gegenüberstehen. Betrachtet man aber die Vielzahl der Themen die diesem Forschungsnetzwerk adressiert werden, ist es aus Sicht der Forschung ein lohnender Aufwand. Der Ausbildungsstand in Kuba ist bekanntlich gut, und der Bedarf sowie das Interesse an diesen Themen in Kuba sind groß. Die Sicht auf die Thematik ist häufig geprägt von einem unterschiedlichen Gesellschaftsbild, dies zeigt sich insbesondere im Promotionsvorhaben zur Messung der sozialen Auswirkungen von CSR. Anderer Themen, wie beispielsweise die erweiterte Nachhaltigkeitsberichterstattung sind in Kuba bislang nur Forschungsgegenstand und von der realen Umsetzung weit entfernt.

Mit DEEBIS-Net ist es gelungen, die weltweit einzigartige und führende Lehre und Forschung im Bereich der Betrieblichen Umweltinformatik in einem Entwicklungsland zu verankern. In einem weiteren Projekt wird zudem parallel ein Masterstudiengang Betriebliche Umweltinformationssysteme in Lateinamerika eingeführt, der ebenfalls für einen Ausbau des wissenschaftlichen Nachwuchses in dieser jungen Disziplin sorgen wird.

Literatur

[ART02]: Arteche, F. (2002) - “Empresa y desarrollo sostenible” SOLUZIONA Calidad y Medio Ambiente.

[BGH07]: Blanke, M.; Godemann, J.; Herzig, C. (2007): Internetgestützte Nachhaltigkeitsberichterstattung. Eine empirische Untersuchung der Unternehmen des DAX30. Lüneburg: Centre for Sustainability Management (CSM) & Institut für Umweltkommunikation (INFU).

[BOD06]: Bode, Ch. (2006): Statement: 5 Jahre PHD-Programm: Zahlen und Fakten einer Erfolgsgeschichte. Im Rahmen der 5. Jahrestagung im Programm „Promotion an Hochschulen in Deutschland (PHD)“. Bonn.

[FNM09] Funk, B.; Niemeyer, P.; Möller, A.: A reference architecture for the integration of EMIS and ERP-systems, Lecture Notes in Informatics, vol. 154, p. 3393-3401, 2009

[GHM09]: Giesen, N; Hashemi Farzad, T.; Marx Gómez J. (2009): A component based approach for overall Environmental Management Information Systems (EMIS) integration and implementation. In: Wohlgemuth, V.; Page, B.; Voigt, K. (Hrsg.): EnviroInfo 2009 - Environmental Informatics and Industrial Environmental Protection: Concepts, Methods and Tools; Shaker, Aachen. Volume 2, pp. 155-160.

[HOJ93]: Hopfenbeck W, Jasch C (1993) Öko-Controlling. Audits, Umweltberichte und Ökobilanzen als betriebliche Führungsinstrumente. Landsberg/Lech, Moderne Industrie.

[IMG08]: Isenmann, R.; Marx Gómez, J.: Internetbasierte Nachhaltigkeitsberichterstattung: Maßgeschneiderte Stakeholder-Kommunikation mit IT. Berlin : ESV, 2008.

[KEH08]: Kuhndt, M.; Eckermann, A.; Herrendorf, M. (2008): Kleine und mittlere Unternehmen als Akteure eines nachhaltigen Wirtschaftens. Aus: Ralf Isenmann (Hrsg.), Jorge Marx Goméz (Hrsg.): Internetbasierte Nachhaltigkeitsberichterstattung: Maßgeschneiderte Stakeholder-Kommunikation mit IT. Schmidt. S. 67

[PNU05]: PNUD (2005) "Hagamos de la competitividad una oportunidad para todos" Sinopsis del Informe sobre Desarrollo Humano Revista de la facultad de ciencias económicas de la UNMSM, año x nº 27. septiembre Perú.

[PVO03]: Page, B., Voigt, K. (2003): Recent History and Development of Environmental Information Systems and Databases in Germany. In: Online Information Review, vol. 27, 1, pages 37-50.

[SUL01]:Sulbarán, Juan Pedro, (2001) "El concepto de Responsabilidad Social de la Empresa" Revista Economía No. 10 p 225-248 Facultad de Ciencias Económicas y Sociales Universidad de Los Andes, Chile.

[SGI08]: Süpke, D.; Marx Goméz, J.; Isenmann, R. (2008): Concept and implementation of a flexible and differentiated shopping cart functionality for creating personalised sustainability reports. Aus: 3rd IEEE International Conference on Information & Communication Technologies: From Theory to Applications (ICTTA' 08). Damascus, Syria. S. 2f

[RAU99]: Rautenstrauch, C. : Betriebliche Umweltinformationssysteme : Grundlagen, Konzepte und Systeme. Berlin u.a. : Springer, 1999.

[WBP01]: Wohlgemuth V, Bruns L, Page B (2001) Simulation als Ansatz zur ökologischen und ökonomischen Planungsunterstützung im Kontext betrieblicher Umweltinformationssysteme (BUIS). In: Hilty LM, Gilgen PW (eds) Sustainability in the Information Society. 15th International Symposium Informatics for Environmental Protection, Zurich 2001. Part 2: Methods/Workshop Papers. Metropolis, Marburg, pp 999-1008

Workshop

Deklarative Modellierung und effiziente Optimierung

Ulrich Geske, Armin Wolf

Deklarative Modellierung und effiziente Optimierung

Ulrich Geske[1], Armin Wolf[2]

[1]Universität Potsdam, [2]Fraunhofer FIRST, Berlin
Ulrich.Geske@uni-potsdam.de, Armin.Wolf@first.fraunhofer.de

Vorwort

Für die Unterstützung bei der Organisation und Durchführung des nunmehr dritten Workshops dieser Serie danken wir dem Programmkomitee, das insbesondere durch die Begutachtung der eingereichten Beiträge und die Unterstützung der Autoren durch Hinweise zu ihren Arbeiten zum Gelingen beigetragen hat: Christoph Beierle, FernUniversität Hagen; Hans-Joachim Goltz, Fraunhofer FIRST, Berlin; Petra Hofstedt, Brandenburgische Technische Universität Cottbus; Walter Hower, Hochschule Albstadt-Sigmaringen; Ulrich John, SIR Plan GmbH, Berlin; Frank Raiser, Universität Ulm; Georg Ringwelski, Hochschule Zittau/Görlitz; Andreas Schutt, University of Melbourne; Dietmar Seipel, Universität Würzburg.

Die Programmierung kombinatorisch komplexer Probleme, basierend auf fundierten theoretischen Grundlagen, weist eine Anwendungsreife auf, dass sie zur Modellierung und Lösung vielfältiger Optimierungsprobleme, bei denen eine Vielzahl meist unterschiedlicher Bedingungen zu berücksichtigen sind, einsetzbar ist. Die aktuelle Entwicklung tendiert immer mehr dahin, auch für große und komplexe Probleme in unterschiedlichsten Lebensbereichen in kürzester Zeit eine Lösung und häufig sogar eine nahezu optimale Lösung zu finden. Anwendungsbereiche von Simulations- und Optimierungs-Verfahren sind vielfältig. Sie können z.B. im Bereich der Planung und Optimierung der Reihenfolge von Operationen, der Personalplanung (Dienstpläne unter Beachtung unterschiedlicher Qualifikationen), aber auch bei Konfiguration und Wartung medizintechnischer Geräte und der Energieeinsparung (z.B. Reduzierung des Energieverbrauchs von Anlagen als ein Optimierungskriterium) liegen. Für Unternehmen ist es häufig wichtig, dass innerhalb kürzester Zeit eine akzeptable Lösung gefunden wird, um wettbewerbsfähig zu sein. Dabei muss auf veränderte Randbedingungen reagiert werden können, um die Lösung anzupassen. Für die Simulationen von "was-wäre-wenn"-Szenarien sind Konsequenzen aus Entscheidungen unmittelbar aufzuzeigen, um bestmögliche Unterstützung der Anwender zu bieten.

Die zur Optimierung eingesetzten Methoden weisen eine große Breite auf. Neben der Weiterentwicklung von mathematischen Verfahren des Operations Research basieren erfolgreiche moderne Verfahren und Programmierwerkzeuge auf Methoden des Soft

Computing (wie Genetische Algorithmen), auf heuristischen (wie Simulated Annealing), wissensbasierten (wie Discrepancy Search) und mathematischen (wie Operations Research) Verfahren. Das besondere Interesse des Workshops ist auf die Untersuchung, Weiterentwicklung und Anwendung der Technik der Constraint-Löser gerichtet. Constraint-Verfahren zeichnet unter den genannten Verfahren der Vorteil der größtmöglichen Transparenz der Lösungsfindung aus. Die Nachvollziehbarkeit der automatisch gefundenen Lösung ist für den Anwender häufig ein wesentliches Auswahlkriterium. Eine weiteres wesentliches Merkmal dieses Verfahrens ist die Möglichkeit einer inkrementellen, anwendergesteuerten Lösungsverbesserung. Der Workshop zielt daher insbesondere auf die Anwendung und Weiterentwicklung von Constraint-Technologien sowie auf die Verbindung bzw. Gegenüberstellung dieses Paradigmas mit anderen Optimierungsverfahren. Konkrete Themen, die für den Workshop eine Rolle spielen sind deshalb einerseits die constraint-basierte Modellierung und Lösung von Anwendungsproblemen, u.a. für die Werkstatt- und Materialbedarfsplanung, das Flottenmanagement und die Fahrplanoptimierung, Personal- und Stundenplanung, Variantenkonfiguration technischer Anlagen aber auch deren Software, die Optimierung verteilter Berechnungs- oder Geschäftsprozesse sowie die Lastverteilung, als auch andererseits die Implementierung von Constraint-Modellen, u.a. durch die Verwendung unterschiedlicher Constraint-Systeme, die Erweiterung von Konsistenztechniken, effiziente Suchverfahren, Nachweis der Optimalität, Visualisierung des Lösungsprozesse, inkrementelle Lösungsanpassung u.a. in Verbindung mit anderen Optimierungstechniken Die für diesen Workshop akzeptierten Arbeiten demonstrieren einige der unterschiedlichen Aspekte des Workshop-Themas.

Petra Hofstedt beschreibt in „The Multiparadigm Programming Language CCFL" die Vorteile eines Sprachansatzes, der verschiedene Programmierparadigmen - hier funktional, constraint-basiert und parallel bzw. concurrent – in einer einheitlichen Art in kombiniert. Die Konstrukte der entwickelten Programmiersprache „Concurrent Constraint Functional Language" (CCFL) werden durch prägnante Beispiele erläutert.

Dagi Trögner behandelt in ihrer Arbeit „Constraint Reasoning in Domain-specific Modeling Languages" die Methode, Unschärfe im Problembereich durch eine Kombination von Fuzzy-Set-Methoden und Constraint-Techniken zu behandeln und beschreibt eine dafür entwickelte domänenspezifische Sprache.

In der Arbeit " Interaktive Komponenten in constraint-basierten Planungssystemen" von Hans-Joachim Goltz werden unterschiedliche Varianten der Kombination von automatischer und interaktiver Lösungssuche und deren Anwendung in constraint-logischen Systemen diskutiert.

In „Constraintbasierte Behandlungsplanung in der Dialyse" von Armin Wolf, Ulrich Geske, Andrej Finsterbusch und Mario Rothe wird die Anwendung des Constraint-Paradigmas zur Sicherung der Patientenzufriedenheit in einem sensiblen medizinischen Behandlungsbereich beschrieben. Durch die Kombination von Constraint-Propagation mit einer geeigneten, problemadäquaten Heuristik wird eine äußerst effiziente Ermittlung eines Behandlungsplanes erreicht.

Interaktive Komponenten in constraint-basierten Planungssystemen

Hans-Joachim Goltz

Fraunhofer FIRST, Berlin
hans-joachim.goltz@first.fraunhofer.de

Abstract: Diskutiert werden Möglichkeiten und Varianten der Integration von interaktiven Komponenten in constraint-basierten Systemen, insbesondere in Planungssystemen.Drei grundsätzlichen Varianten der interaktiven Einzelplanung werden kurz beschrieben.

1 Einleitung

Die constraintlogische Programmierung über endliche Domänenbereiche konnte in vielen Anwendungen zur Lösung komplexer diskreter Probleme, insbesondere für Planungsprobleme, sehr erfolgreich eingesetzt werden (siehe z.B. [St08, Wa96]). Es existieren eine Vielzahl von Arbeiten, in denen praktische Anwendungen der constraintbasierten Programmierung beschrieben werden. In den meisten Fällen werden aber nur Algorithmen beschrieben, die eine automatische Lösungssuche realisieren. In unseren Anwendungssystem verwenden wir erfolgreich eine kombinierte automatische und interaktive Lösungssuche (siehe z.B. [GM99, Go00, Sc01, GP09]). Möglichkeiten und Varianten der Integration von interaktiven Komponenten in praktischen Anwendungen der constraintlogische Programmierung werden im Folgenden diskutiert.

Die Beschreibung der Algorithmen erfolgt teilweise in einem Pseudocode, der sich an die Syntax der constraintlogischen Programmiersprache CHIP orientiert (siehe z.B. [Di88]). Die Syntax von CHIP basiert auf PROLOG. Zusätzlich können Objekte mit Attributen, die auch Variablen sein können, definiert werden. Mit "`Obj@Attr`" kann auf den Wert eines Attributes zugegriffen und durch "`Obj@Attr <- Val`" ein Wert gesetzt werden.

2 Interaktionen bei der Lösungssuche

In erfolgreichen praktischen Anwendungen ist es oft wichtig, das der Nutzer nicht nur die Möglichkeit einer automatischen Lösungserzeugung besitzt, sondern dass er durch Nutzeraktionen einen direkten Einfluss auf die Erzeugung von Lösungen be-

sitzt. Nicht immer lassen sich alle Bedingungen quantitativ spezifizieren. Dann muss aber die Möglichkeit bestehen, dass automatisch erzeugte Lösungen durch den Nutzer modifiziert werden können. Natürlich dürfen nur solche Nutzeraktionen erlaubt werden, die die gegebenen Constraints alle erfüllen. Gerade die constraintbasierte Programmierung ist dafür besonders gut geeignet.

Durch Nutzeraktionen besteht auch die Möglichkeit, in ein Anwendungssystem die Eigenschaft zu integrieren, dass bestimmte Constraints in Ausnahmefällen verletzt werden können. So dürfen bei einer automatischen Schichtplanung die Schichtbedingungen natürlich nicht verletzt werden. Bei einer interaktiven Einzelplanung könnte man dies aber in Ausnahmefällen erlauben, da dann die Verletzungen der Schichtregeln individuell bestätigt werden müssen. Bei der Raumplanung in einer Universität dürfte es beispielsweise nicht sinnvoll sein, das ein Seminar mit wenigen Studenten in einem Vorlesungssaal stattfindet. Bei einer automatischen Planung sollte eine solche Zuordnung nicht erlaubt werden. Bei einer interaktiven Einzelplanung könnte dies aber wieder als Ausnahme ermöglicht werden. Folgende Möglichkeiten der Nutzeraktionen können unterschieden werden, wobei grundsätzlich vorausgesetzt wird, dass der Nutzer Objekte als ausgewählt markieren kann: *alle automatisch bzw. alle Restlichen automatisch einplanen, markierte einplanen, markierte ausplanen, alle ausplanen, Einzelplanung.*

Für die Akzeptanz eines interaktiven Systems ist es wichtig, dass kein automatisches Backtracking von Nutzerkationen erfolgt. Außerdem sollte jeder Planungszustand gespeichert werden können (auch in einer Datei) und die Constraints müssen mit Einbeziehung des aktuellen Planungszustandes erzeugbar sein. In Anwendungssystemen kann es weiterhin sehr wichtig sein, wenn verschiedene Planungszustände der aktuellen Sitzung gespeichert werden können. So könnte beispielsweise immer der vorletzte Planungszustand gespeichert werden, so dass ein einmaliges Zurücksetzen auf einen vorangegangenen Zustand ermöglicht wird. Außerdem könnte eine vorgegebene maximale Anzahl von Planungszustände gespeichert werden können und das Vertauschen von Planungszuständen erlaubt werden. Insbesondere dadurch können dann leicht "Was-Wäre-Wenn"-Simulationen realisiert werden. Im Folgenden werden drei grundsätzlichen Varianten der interaktiven Einzelplanung diskutiert.

3 Varianten der interaktiven Einzelplanung

Für eine interaktive Einzelplanung existieren unterschiedliche Varianten. In Abhängigkeit vom Art des Problems und dem Ziel der Nutzeraktionen ist eine geeignete Variante zu wählen. Folgende Varianten der interaktiven Einzelplanung werden im Folgenden diskutiert:

1. *Umplanung nach Nutzeraktion:* Durch eine Aktion des Nutzers in der grafischen Oberfläche wird eine Umplanung erforderlich.
2. *ein einzelnes Objekt wird interaktiv eingeplant:* Für ein ausgewähltes Objekt werden die möglichen Werte grafisch dargestellt, die der Nutzer auswählen

kann. Nach einer erfolgreichen Wertezuordnung ist eine solche interaktive Einplanung beendet.

3. *interaktive Einplanung als allgemeine Aktion:* Der Nutzer wählt als allgemeine Aktion die interaktive Einplanung und das System muss dann alle Constraints erzeugen. Nachdem der Nutzer ein einzuplanendes Objekt ausgewählt hat, müssen für dieses Objekt die auswählbaren Werte grafisch dargestellt werden. Anschließend kann der Nutzer ein anderes Objekt auswählen oder die interaktive Einzelplanung beenden.

3.1 Umplanung nach Nutzeraktion

Das Erzeugen der Constraints und die Umplanung wird durch eine Aktion des Nutzers in der grafischen Oberfläche gestartet. Solche Nutzeraktionen können beispielsweise sein:

- Ein Objekt wird verschoben, was als eine Änderung der Startzeit interpretiert werden muss.
- Eine Ressource ist nicht mehr verfügbar.
- Eine Ressourcenverfügbarkeit wird zeitlich oder kapazitativ eingeschränkt.

Nach einer solchen Nutzeraktion muss das System eine Umplanung starten, die möglichst die gewünschte Nutzeraktion berücksichtigt. Für die Akzeptanz eines solchen Systems ist es wichtig, dass nicht die Umplanung aller Objekte erlaubt wird. Für eine erlaubte Umplanung können Prioritäten eingeführt werden. Im Folgenden werden drei verschiedene Prioritäten betrachtet:

- *umplanbar:* darf immer umgeplant werden;
- *eingeschränkt umplanbar:* eine Umplanung sollte möglichst vermieden werden;
- *fest geplant:* darf nur umgeplant werden, falls eine Ressourcenzuordnung nicht mehr möglich ist; ein zeitliches Verschieben ist nicht erlaubt.

Auf eine Nutzeraktion muss dass System wie folgt reagieren, wobei angenommen wird, dass das Ziel der Umplanung eine möglichst geringe Änderung des vorangegangenes Planes ist:

1. die relevanten Constraints erzeugen: für die umplanbaren und eingeschränkt umplanbaren Objekte so, als wären sie nicht eingeplant; für ein fest geplantes Objekt mit einer ungültigen Ressourcenzuordnung so, als wäre es nicht eingeplant; für alle anderen fest geplanten Objekten mit allen Zuordnungen der Einplanung.
2. eine neue Lösung suchen: zuerst für ein fest geplantes Objekt mit einer ungültigen Ressourcenzuordnung eine neue Ressourcenzuordnung finden, wobei die

Startzeiten möglichst wenig geändert werden sollten; dann werden die anderen Objekte in folgender Reihenfolge eingeplant: die beschränkt umplanbaren Objekte, ein eventuell vorhandenes verschobene Objekt und zum Schluss die umplanbaren Objekte; bei der Einplanung werden generell die alten Planungswerte bevorzugt und die Abweichung der Zeitzuordnung zum alten Wert wird möglichst minimiert.

3.2 Interaktive Einplanung als Einzelaktion

Der Nutzer wählt ein Objekt für die interaktive Einzelplanung aus. Das Ziel der Aktion kann dabei sowohl eine Einplanung als auch die Modifikation einer Einplanung sein. Bei der Modifikation muss das System das Objekt zuerst intern ausplanen. Das System muss dann alle Constraints erzeugen, die für das ausgewählte Objekt relevant sind und die auswählbaren Werte für die Einplanung grafisch darstellen. Dann muss das System auf einen relevanten Mausklick des Nutzers in der grafischen Nutzeroberfläche warten. Der Nutzer kann so einen möglichen Wert auswählen. Nach einer erfolgreichen Wertzuordnung ist diese interaktive Gesamtaktion beendet. Neben der Wertauswahl müssen mindestens noch zwei weitere Fälle einer Nutzeraktion betrachtet werden: "*Abbruch der interaktiven Einzelplanung*" und "*es wird kein möglicher Wert ausgewählt*".

Im Folgenden wird die Grundstruktur eines Algorithmus für eine solche interaktive Einzelplanung angegeben. Dabei bezeichnet `search@interactive` ein Attribut der Klasse `search` zur Kontrolle der interaktiven Einzelplanung.

```
start_search_interactive(Object) :-
   generate_constraints(Object), % alle relevanten Constraints erzeugen
   additional_propagation(Object), !, % siehe Abschnitt 4
   search_interactive(Object).
start_search_interactive(_) :-
   writeln('There is not any valid value, scheduling is not possible').

search_interactive(Object) :-
   show_domain(Object), % die möglichen Werte grafisch darstellen
   search@interactive <- on,
   waiting_of_user_action,  % warten bis search@interactive \= on
   search1_interactive(search@interactive,Object).

search1_interactive(out, Object) :- !, % keinen gültigen Wert ausgewählt
   search@interactive <- on,
   waiting_of_user_action,  % das Warten wird fortgesetzt
   search1_interactive(search@interactive,Object).
search1_interactive(stop, _) :- !, % Abbruch der Einzelplanung
   redraw_gantt. % grafische Darstellung aktualisieren
search1_interactive(Value, Object) :-
   integer(Value),
```

```
   check_value_choice(Value,Object), !, % Wert zuordnen und überprüfen
   writeln('Interactive search was successful'),
   redraw_gantt. % grafische Darstellung aktualisieren
search1_interactive(Value, Object) :-
   writeln('There is a conflict if this value is selected'),
   notin_domain(Value,Object), !, % Wert aus dem Domänbereich streichen
   search_interactive(Object). % neue Werteauswahl ermöglichen
search1_interactive(Value, _) :-
   writeln('Interactive search was not successful'),
   redraw_gantt. % grafische Darstellung aktualisieren
```

3.3 Allgemeine interaktive Einplanung

Der Nutzer wählt als allgemeine Aktion die interaktive Einplanung und das System muss dann alle Constraints erzeugen. Nachdem der Nutzer ein einzuplanendes Objekt ausgewählt hat, muss das System die auswählbaren Werte darstellen. Nach der Auswahl eines Wertes und einer erfolgreichen Wertzuordnung kann der Nutzer ein weiteres Objekt zur Einplanung auswählen.
Gegenüber der Einzelaktion einer interaktiven Einplanung muss folglich noch eine "Warteschleife" für die Auswahl eines Objektes vorangestellt werden. Folgende Nutzeraktionen könnten noch integriert werden:

- Anstatt eines Wertes der grafisch dargestellten möglichen Werte wählt der Nutzer ein anderes Objekt. Dies bedeutet einen Abbruch der aktuellen Wertezuordnung und ein Start einer neuen Wertezuordnung.
- Der Nutzer wählt ein Objekt, das bereits eingeplant ist. Dies ist als Wunsch einer Modifikation der Einplanung zu interpretieren. Das ausgewählte Objekt ist zuerst auszuplanen und alle Constraints müssen neu erzeugt werden. Algorithmisch kann dies beispielsweise über ein Backtracking der bereits erfolgten Erzeugung der Constraints und einer catch-throw-Konstruktion realisiert werden. Anschließend kann das ausgewählte Objekt wie ein anderes einzuplanendes Objekt behandelt werden.

4 Zusätzliche Propagation

Ein Constraintlöser über endliche Domänenbereiche ist im Allgemeinen nicht vollständig. Im Domänenbereich einer Constraintvariable können folglich Werte enthalten sein, die bei einer direkten Zuordnung sofort einen Widerspruch erzeugen. Für eine interaktive Einzelplanung ist es sehr ungünstig, wenn auswählbare Werte dargestellt werden, die eigentlich nicht zur Auswahl stehen. Für die Akzeptanz einer interaktiven Einzelplanung ist es deshalb wichtig, eine zusätzlich Propagation zu aktivieren, die solche Werte aus den Domänen vorher möglichst streicht. Ein

einfacher Algorithmus für eine zusätzliche Propagation besteht darin, alle aktuell erlaubten Werte der relevanten Constraintvariable zu überprüfen, ob sie bei einer direkten Zuordnung einen Widerspruch erzeugen würden, und die Werte mit Widerspruch aus dem Domänenbereich zu streichen. Ein solcher Algorithmus kann wie folgt implementiert werden:

```
dom_check(DomVar) :-
   dvar(DomVar),
   !,
   domain(DomVar,ValueList),
   dom_check1(LValue,DomVar).
dom_check(_).

dom_check1([],_).
dom_check1([N|List],DomVar) :-
   not N = DomVar,
   !,
   DomVar #\= N,
   dom_check1(List,DomVar).
dom_check1([N|List],DomVar) :-
   dom_check1(List,DomVar).
```

Für eine Constraintvariable `DomVar`, erzeugt `domain(DomVar,ValueList)` die Liste der Werte `ValueList` , die zum Domänenberich von `DomVar` gehören. Die Relation `DomVar #\= N` bedeutet, dass `DomVar` und `N` verschieden sind, und folglich dass der Wert `N` aus dem Domänenbereich von `DomVar` gestrichen wird. Entsprechend dem Problem könnte der Test `N = DomVar` auch durch einen umfangreicheren Test ersetzt werden. Zu beachten ist aber, dass die zusätzliche Propagation nicht zuviel Zeit benötigt, da der Nutzer interaktiv mit dem System arbeiten möchte.

Literatur

[Di88] Dincbas,M.; van Hentenryck,P.; Simonis,H.; Aggoun,A.; Graf,T.; Berthier,F.: The constraint logic programming language CHIP. In *Int. Conf. Fifth Generation Computer Systems (FGCS'88)*, Tokyo 1988, S. 693–702.

[Go00] Goltz, H.-J.: Combined automatic and interactive timetabling using constraint logic programming. In (E. Burke and Erben,W. Hrsg.): *PATAT 2000, Proc. Int. Conf. Practice and Theory of Automated Timetabling 2000*, S. 78–95.

[GM99] Goltz,H.-J.; Matzke,D.: University timetabling using constraint logic programming. In (Gupta,G., Hrsg.): *Practical Aspects of Declarative Languages*, vol. 1551 of *Lecture Notes in Computer Science*, Springer-Verlag, 1999, S. 320–334.

[GP09] Goltz,H.-J.; Pieth,N.: A Tool for generating Partition Schedules of Multiprocessor Systems. In (Geske,U.; Wolf,A., Hrsg.): *Proc. 23^{rd} Workshop on (Constraint) Logic Programming 2009*, Universitätsverlag Potsdam 2010, S. 167–176.

[Sc01] Schlenker,H.; Goltz,H.-J.; Oestmann,J.W.: TAME – Time Resourcing in Academic Medical Environments. In *AIME 2001, Proceedings Int. Conf.*, LNAI 2101, Springer Verlag, 2001, S. 395-404.

[St08] Stuckey,P.J. (Hrsg.): Principles and Practice of Constraint Programming – CP 2008. Vol. 5202 of *Lecture Notes in Computer Science*, Springer-Verlag, 2008.

[Wa96] Wallace,M.: Practical Applications of Contraint Programming. *Constraints, An International Journal*, 1:139–168, 1996.

The Multiparadigm Programming Language CCFL

Petra Hofstedt

hofstedt@informatik.tu-cottbus.de

Brandenburg University of Technology Cottbus

Abstract: Constraints support an efficient modeling and solution of problems in two ways: The most common application is their use for the description of problems with incomplete knowledge. On the other hand constraints may also serve as particular language constructs for the control of the program evaluation process. This paper presents the *C*oncurrent *C*onstraint *F*unctional *L*anguage CCFL which comprises both aspects.

Real-world problems comprise aspects from several realms. They are, thus, often best implemented by a combination of concepts from different paradigms. This combination is comfortably realised by *multiparadigm programming languages*, an area of research and application which has attracted increased interest in the recent years.

The *C*oncurrent *C*onstraint *F*unctional *L*anguage CCFL is a multiparadigm programming language which combines concepts and constructs from the functional and the constraint-based paradigms. The language enables the description of deterministic computations using a functional programming style and of non-deterministic behaviour based on constraints. Moreover, constraints are used to describe systems of concurrent cooperating processes and even typical parallelization patterns. We discuss language concepts and applications by means of examples.

1 Functional Programming

In CCFL, functions are used to express deterministic computations. CCFL's functional sublanguage inherits notions and concepts from the functional languages HASKELL and OPAL [Opa04, PH06]. It is a lazy language with polymorphic type system. A function consists of a type declaration and a definition allowing the typical constructs such as case-expressions, let-expressions, function application, and some predefined infix operator applications, constants, variables, and constructor terms.

Free Variables In CCFL, expressions are allowed to contain free variables; this also applies to function applications. Function applications with free variables are evaluated using the residuation principle [Smo93], that is, function calls are suspended until the variables are bound to expressions such that a deterministic reduction is possible. For example, a function call $(4 + x)$ with free variable x will suspend. In contrast, a com-

putation `length` [1,*x*,3,2,*y*] of the length of a list containing free variables *x* and *y* is possible because a concrete binding of these variables is not necessary to proceed with the evaluation.

2 Constraint abstractions

Besides for the constraint-based problem description, user-defined constraints (or constraint abstractions) serve two further important purposes: dealing with non-determinism and expressing concurrent computations. We sketch on both in this section.

A constraint abstraction consists of a head and a body which may contain the same elements as a function definition. Additionally, the body can be defined by several alternatives the choice of which is decided by guards, i.e. *ask-constraints*. Each body alternative is a conjunction of *tell-constraints*. A constraint (abstraction) always has result type ***C***.

***ask*- and *tell*-constraints** Within a CCFL rule constraints may have two functionalities: *tell*-constraints generate concurrently working processes which propagate knowledge in form of variable bindings (or constraints in general). These processes communicate over common variables. In contrast, *ask*-constraints do not generate knowledge but check for concrete variable bindings or constraints. *ask*-constraints control the choice of (potentially competing) rules and, thus, allow to express the synchronization of concurrently working processes on the one hand and non-deterministic computations on the other hand.

Example 2.1 *Consider the user-defined constraint* `game` *in Prog. 2.1 which describes a game between two players. For example, a constraint application* `game x y 10` *initiates a game where both players throw the dice 10 times each and they reach the overall values* `x` *and* `y`*, resp.*

The expression `dice x1 & dice y1 & ... & game x2 y2` *(*`m`*−1) in lines 5–7 consists of conjunctively connected* tell*-constraints and they express the rules of the game.*

Guards with ask*-constraints can be found e.g. in the* `member`*-constraint (lines 14 and 15) and they realize the non-determinism in this program as discussed below.*

Concurrent Processes CCFL allows the description of systems of communicating and cooperating processes. The main idea is to express concurrent processes by means of conjunctions of *tell*-constraints.

The constraints in the body of the rules are *tell-constraints.* They create processes which may compute bindings for the incorporated variables. Several *tell*-constraints combined by the &-combinator (as in Prog. 2.1, lines 5–7) generate an according number of processes and these communicate over common variables. *tell*-constraints are either applications of user-defined constraints (e.g. `dice x1`) or they are equality constraints `x =:= fexpr` between a variable *x* and a functional expression `fexpr` (e.g. `x =:= x1 + x2`).

Program 2.1 A simple game of dice

```
 1  fun game :: Int → Int → Int → C
 2  def game x y n =
 3    case n of 0 → x =:= 0 & y =:= 0 ;
 4              m → with x1, y1, x2, y2 :: Int
 5                   in dice x1 & dice y1 &
 6                      x =:= x1 + x2 & y =:= y1 + y2 &
 7                      game x2 y2 (m−1)
 8
 9  fun dice :: Int → C
10  def dice x = member [1,2,3,4,5,6] x
11
12  fun member :: List a → a → C
13  def member l x  =
14     l =:= y:ys → x =:= y |
15     l =:= y:ys → case ys of []    → x =:= y ;
16                             z:zs → member ys x
```

Equality constraints are interpreted as strict. That is, the constraint s =:= t is satisfied, if both expressions can be reduced to the same ground data term [HAB+06]. While a satisfiable equality constraint `x =:= fexpr` produces a binding of the variable `x` to the functional expression `fexpr` and terminates with result value `Success`, an unsatisfiable equality is reduced to the value `Fail` representing an unsuccessful computation.

Non-deterministic Computations The atoms of the guard of a user-defined constraint are *ask-constraints.* If a guard of a rule with matching left-hand side is entailed by the current accumulated bindings and constraints, the concerning rule alternative may be chosen for further derivation. In case that the guard fails or cannot be decided yet, this rule alternative is suspended. If all rule alternatives suspend, the computation waits (possibly infinitely) for a sufficient instantiation of the concerning variables.

Example 2.2 *Consider the* `member`*-constraint in Prog. 2.1. The* ask*-constraints (lines 14, 15) of the guards of both alternatives are the same, i.e.* `l =:= y : ys`*, while the bodies differ. Thus, the evaluation of a constraint* `member` $[y_1,y_2,...,y_n]$ `x` *non-deterministically generates a constraint which binds the variable* `x` *to one list element* y_i*.*

For *ask*-constraints, we distinguish between bound-constraints `bound x` checking, whether a variable `x` is bound to a non-variable term (not used in our examples), and match-constraints `x =:= d` $x_1 \ldots x_n$ which test for a matching of the root symbol of a term bound to the variable `x` with a certain constructor `d`. The variables $x_1 \ldots x_n$ are fresh.

Example 2.3 *For Prog. 2.1 the constraint abstraction* `member` *is the only source of non-*

Program 3.1 Functional `map` and constraint-based `farm`

```
 1  fun map :: (a → b) → List a → List b
 2  def map f l =
 3    case l of []      → [];
 4              x : xs → (f x) : (map f xs)
 5
 6  fun farm :: (a → b) → List a → List b → C
 7  def farm f l r =
 8    case l of []      → r =:= [];
 9              x : xs → with rs :: List b
10                       in r =:= (f x) : rs & farm f xs rs
```

determinism. It non-deterministically chooses a value from a list such that a constraint application `dice x` *calling the* `member`*-constraint simulates the dice. The* tell*-constraints* `dice x1` *and* `dice y1` *(line 5) produce values which are consumed by the equality constraints* `x =:= x1 + x2` *and* `y =:= y1 + y2`*, resp. Note that their computation is suspended until the arguments are (sufficiently) instantiated to apply the built-in arithmetic function* +*.*

3 Parallel Programming

The differentiation between functions as computational core and constraints as coordinational core of CCFL allows an explicit control of concurrency and even to express typical parallelization schemes. The following examples are taken from [HL09].

Example 3.1 *Consider Prog. 3.1 defining a function* `map` *and a constraint abstraction* `farm`*. Both have the same general structure, i.e. a function* `f` *is applied to every element of a given list* `l` *and the results are composed into a new list. However, there is one fundamental difference: Since* `map` *is a function, it is evaluated sequentially. In contrast,* `farm` *is a user-defined constraint. Its evaluation yields two concurrently working processes generated from the constraint conjunction in line 10.*

While this uncontrolled form of parallelization as demonstrated in Example 3.1 may yield a huge number of, possibly computationally light-weight, concurrent processes, a selective control of the degree of parallelization of computations is possible in CCFL, too.

Example 3.2 *Prog. 3.2 shows a data parallel farm skeleton* `pfarm` *with granularity control. Here, the number of processing elements* `noPE` *determines the number of generated processes. The abstraction* `pfarm` *calls* `nfarm` *which splits the list to be processed into* `noPE` *sub-lists and generates an according number of processes for list processing. These are distributed across the parallel computing nodes by the run-time system of* CCFL*.*

Program 3.2 Farm parallelization patterns

```
1  fun nfarm :: Int → (a → b) → List a → List b → C
2  def nfarm n f l r =
3    with rs :: List (List b)
4    in let parts = partition n l;
5           pf    = map f
6       in farm pf parts rs & r =:= concat rs
7
8  fun pfarm :: (a → b) → List a → List b → C
9  def pfarm f l r = nfarm noPE f l r
```

CCFL does not feature specialized data structures to support data parallelism in contrast to other approaches [CLJ+07, Nit05]. Instead, the user provides a regular splitting of the data structure controlling the granularity of parallelism in this way[1], while the runtime system is responsible for an equitable distribution of the data (and tasks) onto the processing nodes. Thus, the step from data to task parallel skeletons is smooth in CCFL.

More examples, in particular on parallel processing, and a brief sketch of semantics and implementation details (including further references) of our language can be found in the extended version of this paper (attached to the proceedings and in electronic form).

4 Conclusion

Constraints support an efficient modelling and solution of problems in two forms: First, they are used to model and solve problems with incomplete knowledge. But secondly, they also allow to guide the program evaluation process. The *multiparadigm programming language* CCFL presented in this paper comprises both aspects. It is a successful approach on the integration of the functional and constraint-based paradigms allowing a comfortable modelling of systems of concurrent processes and typical parallelization patterns on the one hand, and of deterministic and non-deterministic behavior on the other hand.

Related Work GOFFIN [CGKL98] is a constraint functional language with a similar structure like CCFL. However, there are fundamental differences, e.g. in the nature of constraint abstractions, the *ask*-constraint's functionalities, and in expressing parallelism. Moreover, in [Hof09] we discuss the extension of CCFL by e.g. arithmetic constraints.

Concentrating on functional programming approaches, there are, e.g. the functional languages EDEN [LOMP05] and ERLANG [AVWW07] which, similar to CCFL, allow concurrent computation of processes. However, both use explicit notions for the generation of

[1] Thus, in our approach the number of processing elements *noPE* plays a role not only in the machine space but also on the level of the problem description.

processes and their communication. CONCURRENT HASKELL [PGF96] supports threads via the IO monad. DATA PARALLEL HASKELL [CLJ+07] targets multicore architectures and allows nested data-parallel programming based on a built-in type of parallel arrays.

Acknowledgment This work has been partially supported by a postdoctoral fellowship of the author, No. PE 07542, from the Japan Society for the Promotion of Science (JSPS).

References

[AVWW07] J. Armstrong, R. Virding, C. Wikstrom, and M. Williams. *Concurrent Programming in Erlang*. Prentice Hall, 2nd edition, 2007.

[CGKL98] M.M.T. Chakravarty, Y. Guo, M. Köhler, and H.C.R. Lock. GOFFIN: Higher-Order Functions Meet Concurrent Constraints. *Science of Computer Programming*, 30(1-2):157–199, 1998.

[CLJ+07] M.M.T. Chakravarty, R. Leshchinskiy, S.L. Peyton Jones, G. Keller, and S. Marlow. Data Parallel Haskell: A status report. In *Workshop on Declarative Aspects of Multicore Programming – DAMP*, pages 10–18. ACM, 2007.

[HAB+06] M. Hanus, S. Antoy, B. Braßel, H. Kuchen, F.J. Lopez-Fraguas, W. Lux, J.J. Moreno Navarro, and F. Steiner. Curry: An Integrated Functional Logic Language. Technical report, 2006. Version 0.8.2 of March 28, 2006.

[HL09] P. Hofstedt and F. Lorenzen. Constraint Functional Multicore Programming. In S. Fischer, E. Maehle, and R. Reischuk, editors, *Informatik 2009. GI Jahrestagung*, volume 154 of *Lecture Notes in Informatics (LNI)*, pages 367, 2901–2915. GI, 2009.

[Hof09] P. Hofstedt. Multiparadigm Constraint Programming Languages, 2009. Habilitation thesis. Technische Universität Berlin.

[LOMP05] R. Loogen, Y. Ortega-Mallén, and R. Peña. Parallel Functional Programming in Eden. *Journal of Functional Programming*, 15(3):431–475, 2005.

[Nit05] T. Nitsche. *Data Distribution and Communication Management for Parallel Systems*. PhD thesis, Technische Universität Berlin, 2005.

[Opa04] The OPAL Project. `http://uebb.cs.tu-berlin.de/~opal/`, 2004. last visited 2010-04-24.

[PGF96] S. Peyton Jones, A. Gordon, and S. Finne. Concurrent Haskell. In *Principles of Programming Languages – POPL*, pages 295–308, 1996.

[PH06] P. Pepper and P. Hofstedt. *Funktionale Programmierung: Sprachdesign und Programmiertechnik*. Springer, 2006.

[Smo93] G. Smolka. Residuation and Guarded Rules for Constraint Logic Programming. In Frédéric Benhamou and Alain Colmerauer, editors, *Constraint Logic Programming. Selected Research*, pages 405–419. The MIT Press, 1993.

Combination of fuzzy sets with the Object Constraint Language (OCL)

Dagi Troegner
Institute of Systems Engineering, Department of Simulation,
Leibniz Universität, Welfengarten 1, 30167 Hannover
Dagi.Troegner@dlr.de

Abstract: In domain-specific modeling languages a variety of constraints originating from different resources have to be considered and integrated into a verification process to ensure a model's correctness. The basis of common model checking methods are constraints derived from meta-model specifications. Besides the meta-model, these constraints often derive from a domain's knowledge. In principle, the heuristic rules and data formulated by domain experts are not precise. To cope with vagueness in the domain knowledge, the concept of fuzzy sets is used. This work therefore describes an approach to integrate fuzzy sets and constraints expressed in the Object Constraint Language (OCL) in a combined constraint reasoning process. The approach is exemplified by introducing a domain-specific modeling language (DSL) for the scope of arrival management.

1 Introduction

A verification process for domain-specific modeling languages has to consider a variety of constraints that may originate from different resources. The basis of common model verification is constraints derived from the meta-model specifications. Correctness of data types of attributes, inheritance or association cardinalities in models are examples of these constraints, which can be added to a meta-model specification in terms of declarative languages (e.g. OCL, Prolog). Furthermore, modeling languages reflect the knowledge of their domain. A domain's knowledge is, besides common basic information (e.g. specifications, legal requirements), based on the experiential knowledge of the domain's experts (e.g. heuristic data and procedures, preferences of application users). The constraints derived from this knowledge can be referred to as domain constraints. In complex domains they can improve the correctness and reasonability of model states in large solution spaces as they enhance the available knowledge in the constraint base. A combination of OCL with fuzzy sets therefore contributes to an enrichment of model checking processes by adding further requirements representing a domain's knowledge and indicating a correct model. In [KPP06] an alignment of OCL with domain-specific languages is presented. This approach is adopted in this paper and enriched by a concept to transform the elements of a DSL to according class concepts of an object-oriented programming language. The set of constraints that is applicable in the verification process is considered as a constraint hierarchy, which labels each constraint as either required or preferred at different strengths. A constraint satisfaction algorithm for the hierarchical constraints is formulated according to an approach presented in [BFW92].

2 Motivating Example

The domain-specific modeling language AMAN-ML (Arrival Management Modeling Language) has been developed with the purpose of designing and implementing an arrival management system that is adaptable to airport-specific requirements and the heuristic knowledge of air traffic controllers. One of the components of AMAN-ML is the *RunwayConfiguration*-model that is based on the layout of an airport's runway system. In addition to the runway layout the model enables a visualization of the dependence-relationships between runways (parallel-, crossing- or V-layout) and resulting separation requirements for aircraft operations. Figure 1 shows an exemplary model for the airport of Hamburg-Fuhlsbüttel in Germany.

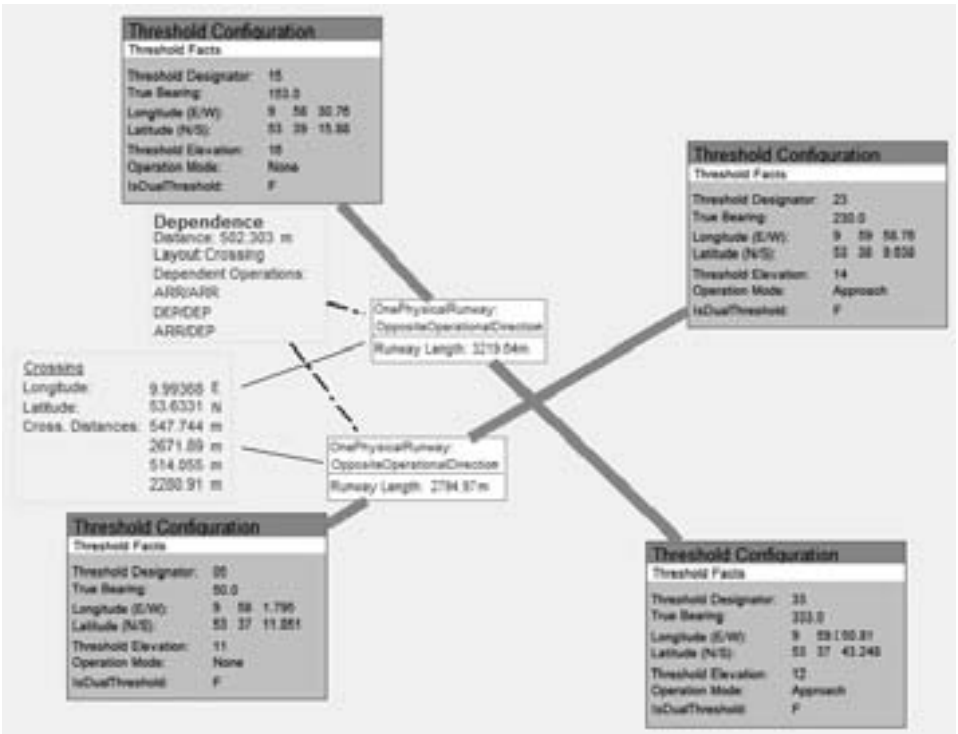


Figure 1: *RunwayConfiguration*-model for Hamburg-Fuhlsbüttel, Germany

The basic elements in the *RunwayConfiguration*-model for Hamburg-Fuhlsbüttel are the runway thresholds containing runway designator, bearing, geographic coordinates and operation mode (arrival, departure or multi mode) as attributes. The created runway thresholds can be related to each other using different relationship types to enhance the level of information for the modeler and consequently for the subsequent code generation. Two opposing thresholds of one runway can be related by a relationship of type 'OnePhysicalRunway' in the category 'OppositeOperationalDirections'. This relationship has an attribute specifying the length of a runway, calculated as the geographical distance between the paired thresholds.

The two runways in Figure 1 (15/33 and 05/23) have intersecting centerlines, which can be specified by a relationship of type 'Crossing'. The 'Crossing'-relationship provides further information regarding the crossing coordinates and the distances from the thresholds to the intersection point. A crossing implies dependencies for all operation modes for the runways involved. These dependencies are defined in a referring 'Dependence'-relationship, containing all operation mode procedures leading to separation requirements for arriving or departing aircraft.

A variety of domain-inherent constraints prescribe the correctness of a model instance. Referring to the *RunwayConfiguration*-example, e.g. legal requirements have to be considered that define the degree of dependence between runways based on their distance, potential intersection and intended operation modes (departure, approach or mixed operations). The dependence in turn has an impact on allowed operation modes and separation requirements between aircraft. Additional domain constraints are considered in form of vague concepts or heuristic procedures to enhance the correctness of the model. Particularly an *OnePhysicalRunway*-relationship can be constrained by verifying the calculated distance between the included runway thresholds against the heuristic concept of runway length. Reasonable values for the length of runways can be derived from heuristic considerations and statistics in airport development.

3 Combination of OCL and fuzzy sets

Model checking is a method for verifying the correctness of a given model state on the basis of a given specification. A variety of constraint languages can be used for this purpose as constraint language specifications differ particularly in their coverage. Therefore, different types of constraints may be necessary to represent their requirements appropriately. In this paper an approach for the integration of fuzzy constraints with meta-model based constraints is presented. Thus, model verification methods are enhanced with requirements representing descriptions heuristically and the rules of domain experts can be formulated in combination with requirements based on a model's specification.

3.1 OCL (Object Constraint Language)

The Object Constraint Language (OCL) is a declarative language for capturing constraints that can be applied to model specifications that are conform to Meta Object Facility (MOF) meta-modeling architectures. OCL provides a set of powerful facilities for navigating and querying models meeting these requirements. An alignment specification is needed for mapping the OCL concept to domain-specific modeling languages (DSL-OCL), as presented e.g. in [KPP06]. In Figure 2 an exemplary constraint is illustrated. The context specifies the correlation of a constraint to a model element. In this case, the constraint restricts relationships of type 'OnePhysicalRunway' with the category 'OppositeOperationalDirections' in a model state. The keyword *inv* implies that an invariant is applied to the given model element determining that the following condition has to be satisfied for each instantiated model element at any time.

```
[context OnePhysicalRunway.OppositeOperationalDirections inv:
 self.length >= 0 AND
 self.length =  fuzzy (RunwayLength)];
```

Figure 2: Example of fuzzy variables in OCL expressions

The body definition of the adapted OCL-constraint contains the set of conditions that have to be satisfied for the given model element. The conditions can be connected by standard Boolean operators. The resulting DSL-OCL is further adapted in this approach to the requirements of integrating fuzzy logic variables into an OCL expression. In the example DSL, AMAN-ML, attribute values can be constrained by approximate degrees of truth. This can be helpful when a values' margin can't be limited exactly, for example when determining the correspondent pair of thresholds to define a runway. The length of a runway may lie between 700m and 4000m with an average of 2918m for Germany. This can be expressed using a fuzzy variable named *RunwayLength* with a Gaussian membership function centering the average value (as can be seen in Figure 3). An integration of fuzzy logic variables into OCL expressions is enabled in the constraint model by invoking an implicit evaluation over the keyword *fuzzy* (Figure 2). A fuzzy inference system is implemented as a separated process calculating the degree of truth and returning the result to the invoking process.

3.2 Fuzzy Constraints

Fuzzy logic can be defined as a form of multi-valued logic based on the paradigm of inference under vagueness [SE10]. This vagueness may result from applying linguistic variables to express natural language concepts or from impreciseness implicated with computational methods. In this approach fuzzy constraint reasoning is used to address constraints expressing a domain's knowledge. Heuristic procedures and values for a domain are mapped to corresponding rules and facts using fuzzy sets. The model verification processes invoke constraint reasoning procedures over selected sets of constraints that may include fuzzy sets. These can be simple fuzzy constraints or OCL constraints using fuzzy variables. In case of rules based on simple fuzzy constraints, a fuzzy reasoning is initiated that follows a standard Mamdani's fuzzy inference method [Ma76]. In Figure 3, a fuzzy reasoning for the correctness of a new relationship of type *OnePhysicalRunway* is shown.

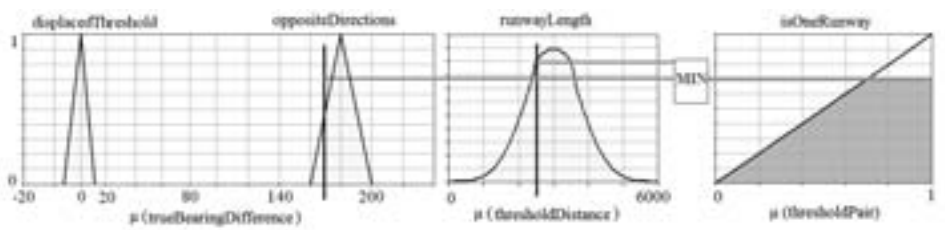


Figure 3: Fuzzy Reasoning example

The correctness of a relationship of type *OnePhysicalRunway* between two objects *Threshold* is dependent to variables that can be expressed most appropriate using fuzzy sets. A pair of thresholds is most likely to be lying on opposite directions of one runway, when their true bearings differ about 180 degrees and the calculated distance between them wouldn't exceed the interval mentioned in the section before. Therefore the *thresholdDistance* and the *trueBearingDifference* are defined as fuzzy variables. According membership functions denote their degrees of membership to the vague sets of *runwayLength* and *oppositeDirections*. The fuzzy operation resolving the AND-operator of the rule defined, determines the resulting output fuzzy set. Further aggregation methods and a defuzzification of the output set (in this case centroid) are used to identify the satisfaction of a fuzzy constraint. In the example shown, a threshold pair with 2600m distance and a bearing distance of 170° is likely to be part of one runway with a vague value of 0.75. Together with an assumption of a marginal value of 0.7 for the result variable, the object pair would be evaluated as being one physical runway. A second case is given when a constraint includes a combination of fuzzy constraints and crisp constraints (e.g. OCL). The constraint reasoning initiated in this situation applies an intermediate step mapping the included crisp variables to corresponding fuzzy variables before passing them to the fuzzy inference process. Crisp constraints can be mapped to fuzzy constraints using balk membership functions (for single values) or sharp-edged trapezoidal membership functions (for value intervals). In the example given, the condition 'self.length>=0' would be mapped to an according fuzzy variable with membership values defined as 1 for all values greater than 0.

4 Constraint Reasoning

The difference regarding constraint verification based on crisp constraints and constraints including fuzzy sets lies basically in the result of the constraint satisfaction problem. Constraint verification over a set of crisp constraints always has a return value of Boolean satisfaction. In this approach, a domain-specific constraint solver is used to cope with satisfiability in combination with vague values. The constraint solver is based on a knowledge base for the specified modeling language. In the context of AMAN-ML, an actual model state is transformed into corresponding constructs of a programming language with instantiations of their classes representing the actual elements of a model state. The relationships of a model instance are represented using self-contained classes with pointers to any objects they are bound to in the actual model instance. The constraints defined in the constraint base are directly related to the class definitions of a model and can be assigned to them through a defined context. A model verification process is invoked either by a modeler's initiation or automatically by modifications to model elements. In the first mentioned scenario all elements of the regarded model will be verified, whereas in model alterations only the elements concerned are going to be checked for maintaining the correctness. When a model verification is initiated, like the creation of a relationship, the corresponding model checking process invokes a constraint verification for the elements included. The set of objects and the relationship created are given to a verification method (e.g. *addOppositeoperationalDirection (*objectSource, *objectTarget, *relationship)*). The model checking process determines the set of applicable constraints through their defined contexts. Boolean satisfaction

results are expected to identify constraints and corresponding model elements not satisfying a given requirement. Invalid model elements are highlighted in the model and stated in an error report with a reference to the constraint violated. The constraint verification uses three sets of constraints: simple OCL, simple fuzzy and combined OCL-Fuzzy. In each set, the required constraints are labelled with degrees of strength to enable a hierarchical constraint solving. The constraints referring to the objects and their attributes are labelled with the highest degrees of strength, followed by the roles and relationships of the model with lower values. This assures the correctness of the bound elements first before verifying a relationship's correctness. Due to the interrelation of the correctness of *Dependence*-relationships to other relationship types, these element types are labelled with the lowest strength value in the constraint hierarchy. In the *RunwayConfiguration*-example, all constraints referring to the context of *Thresholds* will be verified, followed by *OnePhysicalRunway*, *Crossing* and the *Dependence* at last. For each of the three constraint sets this hierarchy is applied. The constraint sets with combined OCL-fuzzy include an implicit fuzzy reasoning process after transforming the included crisp variables to according fuzzy sets. The fuzzy result values are compared to marginal values that are predefined for each fuzzy variable and retransformed to according Boolean values, like in simple fuzzy sets. Finally, after evaluating each constraint set separately, the results are combined by standard Boolean AND-operators.

5 Conclusion

In domain-specific modeling a variety of constraints have to be taken into account representing not only meta-model specifications but also the requirements derived from a domain's knowledge. This knowledge can often be expressed by domain experts most appropriately using vague concepts. Thus, the presented constraint model in this paper enables the usage of fuzzy variables within OCL constraints to combine requirements on meta-model level with domain heuristics. A combined constraint verification process is used for model checking purposes to assist a modeler in creating correct model instances. The verification process described in this paper defines different groups of constraints (simple OCL, simple fuzzy and OCL-Fuzzy) and solves them separately using hierarchical constraint solving. For model checking purpose, basic Boolean operations are used to determine the final result over the subsets. Fuzzy result values are mapped to according Boolean satisfaction values by assigning a marginal value to each defined fuzzy variable.

References

[SE10] Stanford Encyclopedia of Philosophy, http://plato.stanford.edu/entries/logic-fuzzy

[KPP06] Kolovos, D.S.; Paig, R.F.; Polack, F.A.: Aligning OCL with DSLs to Support Instance-Level Queries. In: Journal of the Electronic Communications EASST, 2006.

[BFW92] Borning , A.; Feldman-Benson, B.; Wilson, M.: Constraint Hierarchies. In: LISP and Symbolic Computation. Kluwer Academic Publishers, p. 223-270, 1992.

[Ma76] Mamdani, E.H.: Application of fuzzy logic to approximate reasoning using linguistic synthesis. In: Proceedings of the sixth international symposium on Multiple-valued logic. IEEE Computer Society Press, p. 196 – 202, 1976.

Constraintbasierte Behandlungsplanung in der Dialyse

Armin Wolf*, Ulrich Geske†, Andrej Finsterbusch†, Mario Rothe†

*Fraunhofer FIRST, Kekuléstraße 7, 12489 Berlin, armin.wolf@first.fraunhofer.de
†Universität Potsdam, Institut für Informatik, August-Bebel-Straße 89, 14482 Potsdam
ugeske@uni-potsdam.de, diaatch@gmail.com, mrotheup@gmail.com

1 Einleitung

Steigender Kostendruck im Gesundheitswesen zwingt medizinische Leistungserbringer zu mehr Effizienz. Neben der Erfüllung hoher Qualitätsanforderungen muss die Behandlung den Patienten gerecht werden. Obwohl diese Zielgrößen scheinbar widersprüchlich erscheinen, kann durch eine geeignete Optimierung von Behandlungsabläufen neben einer hohen Termintreue und der Gewährleistung des Behandlungsverlaufs auch eine gute Auslastung der Geräte und des Personals erreicht werden. Aufgrund der damit erzielten Erfolge in vielen Anwendungsbereichen und seiner Flexibilität hinsichtlich problemspezifischen Anpassungen wird hier als Planungsansatz für den Dialysebetrieb *Constraint-Based Scheduling* [BlPN01] gewählt und im Folgenden beschrieben.

2 Behandlungsabläufe in der Dialyse

Zur reibungslosen Durchführbarkeit von Dialysebehandlungen ist sicherzustellen, dass die dazu erforderlichen Ressourcen zur richtigen Zeit verfügbar sind. Die kann durch eine entsprechende Zeitplanung und Ressourcendisposition erreicht werden, wenn dabei die Charakteristika der Behandlungsabläufe adäquat berücksichtigt werden; diese sind:

- selbstständige Vorbereitung des Patienten auf die Dialyse
- Übernahme des Patienten durch eine Pflegekraft und Anschluss an ein geeignetes Dialysegerät (connect)
- Durchführung der Dialyse des Patienten am angeschlossenen Gerät (dialyze)
- Trennung des Patienten vom Dialysegerät durch eine Plegekraft und Entlassung des Patienten (disconnect)
- selbstständige Nachbereitung des Patienten
- Reinigung des genutzten Dialysegeräts durch eine Pflegekraft (clean).

Diese Analyse zeigt, dass für die Behandlungsschritte jeweils unterschiedliche Ressourcen zu deren Durchführung verfügbar sein müssen: Während im ersten und vorletzten Schritt nur die Ressource „Patient“ verfügbar sein muss, sind es im zweiten und vierten Schritt neben dem Patienten eine Pflegekraft und ein geeignetes Dialysegerät, das auch für die Dauer der eigentlichen Dialyse belegt ist und währenddessen für keine andere Behandlung zur Verfügung steht. Gleiches gilt auch für die anderen beteiligten Ressourcen. Die Untersuchung zeigt auch, dass der Nachbereitung und Reinigung voneinander unabhängig sind, d. h. nach Trennung des Patienten vom Dialysegerät sind diese unabhängig ausführbar. Neben diesen Freiheiten gibt es aber auch Einschränkungen: So muss die Reinigung eines Dialysegeräts innerhalb einer vorgegebenen Frist nach seiner Benutzung erfolgen. Des weiteren sind die Pflegekräfte nur zu bestimmten Zeiten (Schichten) verfügbar.

Die Dauern der einzelnen Behandlungsschritte unterscheiden sich individuell und behandlungsbedingt. Zwischen der Vorbereitung des Patienten und seiner Übernahme durch eine Pfegekraft zum Anschluss an das Gerät ist keine Wartezeit vorgesehen.[1] Wunsch- aber auch krankheitsbedingt steht für jede Patientenbehandlung eine vorgegebene Auswahl an Dialysegeräten zur Verfügung. Zudem hat jeder Patient eine Wunschzeit für den Beginn seiner Behandlung, die bei der Planung bestmöglich zu berücksichtigen ist.

Ziel der Planung von Behandlungsabläufen ist neben der Terminierung der Behandlungsschritte die Gewährleistung, dass alle benötigten Ressourcen zu den geplanten Zeiten für die jeweiligen Behandlungsschritte verfügbar sind. Ferner ist eine bestmögliche Erfüllung der Wunschzeiten anzustreben.

3 Constraintbasierte Behandlungsplanung in der Dialyse

Um Behandlungen in der Dialyse constraint-basiert zu planen, ist das Planungsproblem als *Constraint Satisfaction Problem* geeignet zu modellieren. Dazu sind die beteiligten und zu bestimmenden Größen zu identifizieren und durch Variablen und deren endliche ganzzahlige Wertebereiche geeignet zu repräsentieren. Ergänzend sind die einzuhaltenden Randbedingungen in Relationen (*Constraints*) zu übertragen.

3.1 Problemmodellierung

Das Planungsproblem sei durch eine endliche Menge von Patienten P, eine endliche Menge von Geräten D und eine endliche Menge von Pflegekräften N charakterisiert sowie durch ein Planungsintervall $[T_{\min}, T_{\max}]$. Sowohl Geräte als auch Pflegekräfte seien durch Nummern repräsentiert, d. h. es gelte $D = \{1, \ldots, i\}$ und $N = \{i + 1, \ldots, j\}$.

Da die Startzeiten der Behandlungsabläufe sowie die Zuordnungen zu Pflegepersonal und Dialysegeräte zu planen sind, ist es naheliegend, diese durch Variablen über endlichen ganzzahligen Wertebereichen zu repräsentieren. Während die Wertebereiche der Startzeiten Teilmengen des diskretisierten Planungshorizonts sind, sind die Wertebereiche der

[1] Ansonsten kommt der Patient entsprechend später in das Dialysezentrum.

Gerätebezeichner einfache Nummern. Somit wird jeder Behandlungsschritt durch eine *Aktivität* a repräsentiert, die eine *variable Startzeit* $a.\mathsf{start}$, eine *variable Dauer* $a.\mathsf{duration}$ und eine *variable Ressource* $a.\mathsf{resource}$ hat. Es gelte $a.\mathsf{start} \in [T_{\min}, T_{\max} - a.\mathsf{duration}]$, d. h. jede Aktivität liegt vollständig im Planungsintervall.

Für jede Pfegekraft $n \in N$ sei deren Schicht $[b_n, e_n] \subseteq [T_{\min}, T_{\max}]$, d. h. deren Arbeitszeit im Planungsintervall gegeben. Die Planung eines Behandlungsschritts außerhalb der Schicht wird durch eine Vorbelegung dieser Zeiten mit zwei Pseudoaktivitäten before_n und after_n erreicht: $\mathsf{before}_n.\mathsf{start} = T_{\min}$, $\mathsf{after}_n.\mathsf{start} = e_n$, $\mathsf{before}_n.\mathsf{duration} = b_n - T_{\min}$, $\mathsf{after}_n.\mathsf{duration} = T_{\max} - e_n$, $\mathsf{before}_n.\mathsf{resource} = n$ und $\mathsf{after}_n.\mathsf{resource} = n$.

Da bei der Zuordnung der Behandlungsschritte zu den Patienten keine Konflikte auftreten können, brauchen diese nicht explizit für diese „Ressource“ geplant werden. Somit reduziert sich der zu planende Behandlungsablauf für jeden Patienten $p \in P$ auf folgende Aktivitäten: $\mathsf{connect}_p \rightarrow \mathsf{dialyze}_p \rightarrow \mathsf{disconnect}_p \rightarrow \mathsf{clean}_p$, wobei wir die Dauern für den Anschluss des Patienten und seiner Trennung vom Dialysegerät als Teil der eigentlichen Dialysezeit betrachten, d. h. die Dauer der Dialysezeit $\mathsf{dialyze}_p.\mathsf{duration}$ wird um die Dauern von Anschluss und Trennung verlängert. Somit soll $\mathsf{connect}_p.\mathsf{start} = \mathsf{dialyze}_p.\mathsf{start}$ und $\mathsf{disconnect}_p.\mathsf{start} + \mathsf{disconnect}_p.\mathsf{duration} = \mathsf{dialyze}_p.\mathsf{start} + \mathsf{dialize}_p.\mathsf{duration}$ gelten. Durch dieses Vorgehen vereinfacht sich das Problemmodell, da Anschluss und Trennung eines Patienten nur noch einer verfügbaren Pflegekraft zuzuordnen ist.

Die Dialysezeit jedes Patienten $p \in P$ sei fest behandlungsbedingt festgelegt, d. h. es gelte $\mathsf{dialyze}_p.\mathsf{duration} = c_p$ mit $c_p \in \mathbb{N}$; anders die anschließende Reinigungszeit, deren Wert durch die Wahl des Dialysegerät bestimmt ist, d. h. es gelte $\mathsf{clean}_p.\mathsf{duration} = e_{\mathsf{clean}_p.\mathsf{resource}}$, wobei $e_1 \in \mathbb{N}, \ldots, e_i \in \mathbb{N}$ die gerätespezifischen Endreinigungsdauern seien.

Für jeden Patienten $p \in P$ sind die möglichen Dialysegeräte $D_p \subseteq D$ gegeben, diese schränken die Ressourcen für $\mathsf{dialyze}_p$ und clean_p – Aktionen, die beide das gleiche Gerät beanspruchen – ein, d. h. es gilt: $\mathsf{dialyze}_p.\mathsf{resource} \in D_p, \mathsf{clean}.\mathsf{resource} \in D_p, \mathsf{dialyze}_p.\mathsf{resource} = \mathsf{clean}.\mathsf{resource}$. Des weiteren gelte $\mathsf{connect}_p.\mathsf{resource} \in N$ und $\mathsf{disconnect}_p.\mathsf{resource} \in N$, d. h. jeder Patient $p \in P$ kann von jeder Pflegekraft betreut werden; die Pflegekräfte für Anschluss und Trennung können verschiedene Personen sein. Für jeden Patienten $p \in P$ ist das zu disponierende Dialysegerät über den gesamten Behandlungsverlauf gleichbleibend, was sich durch den gewählten Modellierungsansatz auf die Erfüllung der Gleichung (siehe oben) $\mathsf{dialyze}_p.\mathsf{resource} = \mathsf{clean}.\mathsf{resource}$ widerspiegelt. Da die Reinigung eines Dialysegeräts auch eine Pflegekraft bindet, wird sie durch zwei Aktivitäten mit identischen Startzeiten und Dauern aber unterschiedlichen Ressourcen repräsentiert: clean_p^d (d für *device*) und clean_p^n (n für *nurse*) mit $\mathsf{clean}_p.\mathsf{start} := \mathsf{clean}_p^d.\mathsf{start} \equiv \mathsf{clean}_p^n.\mathsf{start}$ und $\mathsf{clean}_p.\mathsf{duration} := \mathsf{clean}_p^d.\mathsf{duration} \equiv \mathsf{clean}_p^n.\mathsf{duration}$. Ferner legen die Bedingungen $\mathsf{dialyze}_p.\mathsf{resource} = \mathsf{clean}_p^d.\mathsf{resource}$ und $0 \leq \mathsf{clean}_p.\mathsf{start} - \mathsf{dialyze}_p.\mathsf{start} - \mathsf{dialyze}_p.\mathsf{duration} \leq B_{\max}$ fest, dass das Dialysegerät nach seinem Gebrauch nach spätestens $B_{\max}$ Zeiteinheiten zu reinigen ist. Da diese maximale Verzögerung (in aller Regel unter einer Stunde) kleiner als jede Dialysezeit ist (in aller Regel mehrere Stunden), kann zwischen dem Ende einer Dialyse auf einem Gerät und dessen Reinigung keine weitere Dialyse geplant werden, d. h. diese unzulässige Situation ist per-se ausgeschlossen.

Für zwei beliebige aber verschiedene Aktivitäten r und s ($r \neq s$) wird durch die Bedingung $r.\mathsf{resource} = s.\mathsf{resource} \Rightarrow r.\mathsf{start} + r.\mathsf{duration} \leq s.\mathsf{start} \vee s.\mathsf{start} + s.\mathsf{duration} \leq r.\mathsf{start}$ der Umstand formuliert, dass zu keiner Zeit weder ein Dialysegerät noch eine Pflegekraft mehrfach belegbar sind.

Jede *Lösung* des formulierten CSP, d. h. jede Belegung der Variablen, die die aufgezählten Randbedingungen erfüllen, repräsentiert einen umsetzbaren Behandlungsplan, da alle disponierten Ressourcen zu den festgelegten Zeiten zur Verfügung stehen. Jedoch werden die Wunschzeiten der Patienten, je nach gewählter Lösung, unterschiedlich gut berücksichtigt. Um diese bestmöglich zu berücksichtigen, wurde als Zielfunktion für eine Optimierung die Summe $\sum_{p \in P}(\mathsf{dialyze}_p.\mathsf{start} - w_p)$ gewählt, wobei $w_p \in \mathbb{N}$ die gewünschte Startzeit der Dialyse (nach der eigenen Vorbereitung) ist. Da es sich bei der Wunschzeit im aller Regel auch um die frühestmögliche Zeit handelt, ist zusätzlich $\mathsf{dialyze}_p.\mathsf{start} \geq w_p$ gefordert.

3.2 Implementierung

Das in Abschnitt 3.1 vorgestellte Problemmodell wurde in Java unter Verwendung der Constraintprogrammierbibliothek `firstCS` [HMSW03] implementiert.

Zur Optimierung wird ein *monotones Branch&Bound*-Verfahren eingesetzt, das ausgehend von einer Lösung mit einem Zielfunktionswert $\mathsf{objective} = o$ mit $o \in \mathbb{N}$ nach einer besseren Lösung sucht, indem das Problemmodell um die Bedingung $\mathsf{objective} < o$ erweitert wird, bis letztlich eine beste Lösung gefunden ist.

In einer ersten Variante wurden zur Optimierung verschiedene, in `firstCS` verfügbare chronologische Tiefensuchverfahren eingesetzt; diese sind:

- `BtLabel`: wählt die Variablen in einer vorgegebenen Reihenfolge aus, um diese dann beginnend mit dem kleinsten Wert ihres Wertebereichs zu belegen und bei Auftreten einer Inkonsistenz mit dem nächstgrößeren Wert fortzufahren, bis entweder eine Lösung gefunden wurde oder die Unlösbarkeit des Problems festgestellt wurde.

- `DomainReductionLabel`: wählt die Variable mit dem größten Wertebereich als nächstes aus, um dann zuerst deren Wertebereich auf die untere Hälfte einzuschränken und bei Auftreten einer Inkonsistenz mit der oberen Hälfte fortzufahren, bis letztlich die Wertebereiche aller Variablen auf einen Wert eingeschränkt sind und damit eine Lösung gefunden ist oder die Unlösbarkeit des Problems festgestellt wurde.

- `AlternativeResourceLabel`: legt zuerst die Ressourcen der Aktivitäten fest, um dann die Aktivitäten, die einer gemeinsamen Ressource zugeordnet sind linear zu ordnen.

Es hat sich gezeigt, dass keines dieser Verfahren ohne Einsatz einer zusätzlichen Heuristik für Probleme realistischer Größe (zirka ein Dutzend Patienten, ein halbes Dutzend Dialysegeräte und 3 Schwestern im 3-Schichtbetrieb) geeignet ist, da die Optimierung nicht

in akzeptabler Zeit (wenige Minuten) terminiert. Im Folgenden wird daher eine Heuristik vorgestellt, die speziell für die Behandlungsplanung in der Dialyse entwickelt wurde und insbesondere in Kombination mit den beiden ersten Suchverfahren innerhalb kürzester Zeit zu optimalen Lösungen führt.

3.3 Problemspezifische Heuristik

Die in diesem Abschnitt vorgestellte Heuristik beruht auf der Erkenntnis, dass die geeignete Zuteilung der Dialysegeräte maßgeblich für eine optimale Planung ist. Eine konfliktarme Zuteilung der Dialysegeräte aufgrund der Wunschzeiten kann somit vermutlich schnell zu optimalen Plänen führen. Dazu wird der *Planungshorziont* $T_{\max} - T_{\min}$ in *Sektoren* unterteilt. Die *Dauer der Sektoren* entspricht der durchschnittlichen Behandlungsdauer. Dazu sei $\mathsf{treatmentDuration}_p = \mathsf{connect}_p.\mathsf{duration} + \mathsf{dialyze}_p.\mathsf{duration} + \mathsf{disconnect}_p.\mathsf{duration}$ die *Gesamtbehandlungsdauer* eines jeden Patienten $p \in P$ und folglich $\mathsf{meanTime} = \sum_{p \in P} \mathsf{treatmenDuration}_p / |P|$ die *durchschnittliche Behandlungsdauer*. Damit besteht der Planungshorizont aus $z = \lfloor (T_{\max} - T_{\min}) / \mathsf{meanTime} \rfloor$ Sektoren. Basierend auf diesen Parameternwerden den Patienten nicht bereits zugeordnete Sektoren auf den Dialysegeräten zugeordnet, sofern dies möglich ist. Dazu werden die Patienten so sortiert, dass Patienten mit weniger Geräteauswahlmöglichkeiten zuerst betrachtet werden und bei gleicher Gerätezahl die Patienten mit früheren Wunschzeiten bevorzugt werden. Dazu wird eine Matrix L verwaltet, die für jedes Gerät die Sektorenauslastung wiedergibt; initial ist diese Null, d. h. $L_{d,s} = 0$ für $d \in D$ und $s = 1, \ldots, z$.

Für jeden Patienten $p \in P$ bezeichnen $d_{1_p} < \cdots < d_{l_p}$ die möglichen Dialysegeräte im Wertebereich von $\mathsf{dialize}_p.\mathsf{resource}$. Für jeden Patienten $p \in P$ wird nun gemäß der Sortierung wie folgt vorgegangen:

1. Bestimme die kleinste Sektorzahl $k \in \{1, \ldots, z\}$, so dass mindestens die halbe Gesamtbehandlungsdauer bei Beginn zur Wunschzeit in diesem Sektor liegt: $w_p - T_{\min} + \mathsf{treatmentDuration}_p / 2 < k \cdot \mathsf{meanTime}$.
2. Bestimme die kleinste Sektorenzahl $k' \in \{k, \ldots, z\}$ und dann das kleinste $d'_p \in \{d_{1_p}, \ldots, d_{l_p}\}$, so dass $L_{d'_p, k'} = 0$ gilt.
3. Falls es ein solches d'_p gibt, setze $\mathsf{dialize}_p.\mathsf{resource} = d'_p$ und lege damit das Dialysegerät fest; markiere den Sektor auf diesem Gerät mittels $L_{d'_p, k'} = 1$ als ausgelastet.

Anschließend werden alle Patienten entsprechend dem gewünschten Behandlungsende $w_p + \mathsf{treatmentDuration}_p$ aufsteigend sortiert. Die Startzeiten der Dialysen und Reinigungen sowie die Geräte und Pflegekräfte werden dann aufgrund dieser Sortierung bestimmt.

3.4 Laufzeituntersuchungen

Laufzeituntersuchungen wurden anhand vorgegebener Planungsprobleme unterschiedlicher Größe – teilweise aus der Praxis – durchgeführt. Die Messwerte wurden auf einem

Beispiel	DR	Bt	AR
3on1	**0.02s** (6)	**0.14s** (737)	**0.09s** (7)
3on3	**0.01s** (0)	**0.01s** (15)	**0.15s** (26)
7on4	**0.06s** (0)	**0.05s** (32)	**2.7s** (233)
12on6	**0.12s** (1)	**0.2s** (260)	-
14on6	**0.19s** (3)	**0.24s** (292)	-
14on6_2	**0.18s** (3)	**0.23s** (292)	-

Tabelle 1: Laufzeiten (Zahl der Rücksetzschritte) der eingesetzten Suchverfahren: DR - `DomainReductionLabel`, Bt - `BtLabel`, AR - `AlternativeResourceLabeler`.

PC mit der Java-Version 1.6.0_15 ermittelt und sind in Tabelle 1 zusammengefasst. Die erste Spalte enthält die verwendeten Beispiele, wobei die gewählten Namen dem Schema <*Anzahl der Patienten*>on<*Anzahl der Geräte*> genügen und damit Auskunft über die Problemgröße geben. Allen Beispielen ist gemein, dass jeweils eine Pflegekraft zur Früh-, Mittel-, und Spätschicht arbeitet und zu jedem Zeitpunkt mindestens eine Pflegekraft verfügbar ist. Bei dem Beispiel mit Index 2 wurden gegenüber dem Beispiel gleicher Größe andere Alternativen von Dialysegeräten gewählt. Die folgenden drei Spalten zeigen die Messergebnisse: Jede Zelle enthält zwei Werte: Die Laufzeit in Sekunden (fett) und die Anzahl der erforderlichen Rücksetzschritte (in Klammern) zum Finden einer optimalen Lösung.

4 Zusammenfassung und Ausblick

Es wurde ein constraint-basierter Lösungsansatz zur Behandlungsplanung in der Dialyse vorgestellt. Wie bei vielen anderen praktischen Anwendungen des *Constraint Programming* hat sich gezeigt, dass neben einen adäquaten Problemmodell Heuristiken zur schnellen Lösungsfindung unerlässlich sind. Mit der erarbeiteten problemspezifischen Heuristik ist ein schnelles Finden optimaler Lösungen möglich, jedoch ist offen, ob durch die dabei getroffene Vorauswahl von Dialysegeräten Lösungen nicht mehr bestimmbar sind; dies bedarf einer eingehenden Untersuchung in der Zukunft.

Literatur

[BlPN01] Philippe Baptiste, Claude le Pape und Wim Nuijten. *Constraint-Based Scheduling*. Number 39 in International Series in Operations Research & Management Science. Kluwer Academic Publishers, 2001.

[HMSW03] Matthias Hoche, Henry Müller, Hans Schlenker und Armin Wolf. `firstCS` - A Pure Java Constraint Programming Engine. In Michael Hanus, Petra Hofstedt und Armin Wolf, Hrsg., *2nd International Workshop on Multi-paradigm Constraint Programming Languages – MultiCPL'03*. Online available at `uebb.cs.tu-berlin.de/MultiCPL03/Proceedings.MultiCPL03.RCoRP03.pdf`, 29th September 2003. (Last visited: 2009/05/13).

Workshop

2. Workshop über Daten in den Lebenswissenschaften Informatik-Technologien und Support

Wolfgang Müller, Ralf Hofestädt, Klaus Kuhn, Can Türker

Daten in den Lebenswissenschaften: Informatik-Technologien und Support

Wolfgang Müller, Ralf Hofestädt, Klaus Kuhn, Can Türker

Dieser Workshop befasst sich mit Problemen, die bei der Datenhaltung biologischer un medizinischer Daten auftreten. Er hat zum Ziel, Vertreter der Datenbanken, des Information Retrieval und Vertreter der Bio- und Medizinischen Informatik zusammenzubringen.

Innerhalb der Informatik in den Lebenswissenschaften gibt es Communities, aus verschiedenen Motivationen an ähnlichen Fragestellungen arbeiten. Auf der einen Seite stehen die Informatiker, die sich von der Informatik ausgehend mit den spezifischen Problemen von Anwendungen in der Biologie und Medizin befassen. Auf der anderen Seite stehen Biologen und Mediziner, die aus ihrem Fachgebiet heraus sich auch mit informatischen Fragestellungen befassen. Diese Gruppen miteinander besser zu vernetzen ist ein Ziel, das von den beteiligten Organisationen und auch mit dem Workshop verfolgt wird.

In diesem Jahr lassen sich die Einreichungen in drei große Gruppen einteilen: Papers zur Datenintegration in der Systembiologie, Papers zum Datenmanagement im Krankenhaus mit seinen spezifischen Sicherheits- und Anonymitäts-Fragestellungen und schließlich Textverarbeitung zur Suche von Dokumenten bzw. zur Informationsextraktion.

Wie im letzten Jahr ist allen Papers gemeinsam, dass sie die Datenintegration bezwecken, denn beispielsweise ist die Informationsextraktion aus der wissenschaftlichen Literatur nichts anderes als die Integration von Daten aus nicht-strukturierten Datenquellen. Jedoch sind die Ansätze und die Schwerpunktsetzungen unterschiedlich. Wir freuen uns auf spannende Diskussionen.

Die Organisatoren dieses Workshops

Wolfgang Müller (Kontakt)
Ralf Hofestädt
Klaus Kuhn
Can Türker

DAWIS-M.D. - a data warehouse system for metabolic data

Klaus Hippe, Benjamin Kormeier, Thoralf Töpel, Sebastian Janowski and Ralf Hofestädt

Bioinformatics Department
Bielefeld University
Universitätsstraße 25
33501 Bielefeld

khippe@techfak.uni-bielefeld.de

Abstract: Today, experiments like microarrays and other high-throughput methods generate a large number of life sciences data. Biologists need to be supported in their research by tools and applications that can analyze and interpret experimental data while considering external data sources or data from experiments of other researchers. Moreover, relationships and interactions between different data sets and biomedical domains have to be detected and represented in a clear and understandable manner. These questions are some of the major challenges in bioinfomatics. Therefore, we present DAWIS-M.D., a platform-independent data warehouse system for metabolic data. This data warehouse information system provides an integrated and consistent view of large scale biomedical data. DAWIS-M.D. is a publicly available web-based system that integrates data from 11 different biomedical databases.

1 Introduction

Biomedical research generates a large amount of heterogeneous data. This data is usually stored in several heterogeneous biomedical databases. Typically, these databases are distributed over the entire world. The information of these databases comes from various experiments and from literature as well. Today, approximately 1230 publicly available databases and information systems for life science data are listed in the NAR catalogue [CG10]. Therefore, data integration is necessary to provide a consistent view of the distributed, heterogeneous and duplicated data.

Scientists and research groups usually use several data sources and information systems simultaneously, e.g. in parallel windows of their web browser. It is then difficult for the scientist to understand complex biological questions and find missing links, because of lacking information. Another problem is the incompatibility between the different information systems and software solutions.

In this paper we present DAWIS-M.D., a platform-independent data warehouse approach for metabolic data. DAWIS-M.D (*Data Warehouse Information System for Metabolic Data*). The information system contains information from 11 different databases. The following data sources are integrated into the data warehouse: BRENDA, EMBL, HPRD,

KEGG, OMIM, SCOP, Transfac, Transpath, ENZYME, GO and UniProt. The data in DAWIS-M.D. is divided into 12 various biological domains, which can be accessed via the web-based graphical user interface (GUI). The application provides search forms for the biological domains *Compound*, *Disease*, *Drug*, *Transcription Factor*, *Enzyme*, *Gene*, *Glycan*, *Gene Ontology*, *Pathway*, *Protein*, *Reaction* and *Reaction Pair* domain.

A wide range of biomedical information in DAWIS-M.D. supports scientists in understanding complex biological systems and their properties. In addition, DAWIS-M.D. identifies relationships and interactions spanning multiple biological domains and is able to display this information.

Networks can be displayed, edited and extended by using the VANESA (*Visualization and Analysis of Networks in System Biology Applications*) [JKT+10] software application. All information and biological knowledge in VANESA is provided by the DAWIS-M.D. data warehouse that is connected via web service to the VANESA network editor.

2 Related Works

One of the major topics in bioinformatics is data integration. The integration of heterogeneous, autonomous and distributed data can be realized using different approaches. Generally, in bioinformatics, integration approaches are divided into following classes: text indexing systems (e.g. SRS [EUA96]), multi database and federated database systems (e.g. DiscoveryLink [HSK+01]) and data warehouses systems (e.g. BioWarehouse [LPW+06], ONDEX [KBT+06] and SYSTOMONAS [CMB+07]). Data warehouse systems are widely used in bioinformatics for data integration. Advantages of the data warehouse techniques are high performance, data cleansing and full and easy access to the integrated data that is pivotal for bioinformatics.

In bioinformatics data warehouse approaches can be divided into two groups: General software infrastructures that can be locally installed and configured like BioWarehouse. And project-oriented data warehouse systems that are implemented for specific biological questions like SYSTOMONAS. Some software solutions cannot be categorized, because they have characteristics of both groups. These solutions are hybrid data warehouse approaches.

All these existing systems have several disadvantages: they are not platform-independent, the local installation is quite time-consuming or they are not available via the web. Furthermore, they are restricted to an operating system concerning their specific programming language. Therefore, the applications are not extensible or flexible enough for other developers. Another problem is the need for up-to-date content in the data warehouse. Most software solutions have to be updated manually, and this is a time-consuming process. Additionally, all systems are restricted to a specific relational database management system (RDBMS) and therefore are not independent from manufacturers such as MySQL, PostreSQL and Oracle. These systems cannot be widely used, because of their limitation to one RDBMS. Further problems are the usability, the project-specific data and missing interfaces for more flexibility.

Table 1 illustrates the main data warehouse systems used in bioinformatics. This table also displays the advantages and disadvantages of each software solution. Based on the advantages and disadvantages of existing data warehouse approaches we developed the concept of DAWIS-M.D. The main goal was to develop a platform-independent and flexible data warehouse system for metabolic data that integrates multiple heterogeneous data sources into a local database. The system should enable an intuitive search of integrated life science data, simple navigation to related information as well as visualization of biological domains and their relationship.

Feature	**BioWarehouse**	**ONDEX**	**SYSTOMONAS**
Integration	Close integration, ready-made relational schemas	Loose integration	Unknown
DBMS	MySQL, Oracle	PostgreSQL	PostgreSQL
Programming language	Java, C	Java	PHP
Architecture	Software infrastructure	Software infrastructure, Web application	Web application
Platform independence	Unix-based	Yes	Yes
Up-to-dateness	Manually	High	Unknown
Maintenance and further development	Regular maintenance and further development	Regular maintenance and further development	Regular maintenance and further development
License	MPL	GNU General Public License	Database dump available, web application free
Open source	Yes	Yes	Unknown

Table 1: Comparison of three exemplary bioinformatics data warehouses systems. For each software solution the advantages are highlighted green and the disadvantages are red.

3 Design and Implementation

The system architecture of DAWIS-M.D. consists of five different layers. The source layer contains the multiple data sources such as BRENDA, EMBL, GO, Enzyme, KEGG, HPRD, OMIM, SCOP, Transfac, Transpath and UniProt.

The different biomedical databases, which are integrated in DAWIS-M.D., are available in the data sources. Most of these databases provide parseable flat files, XML files or Structured Query Language (SQL) dumps that can be processed by the BioDWH data

warehouse infrastructure [TKKH08] that is used for the integration process. A monitor component that is part of the integration layer controls the different data sources. It recognizes changes in the original sources and starts the download if necessary. In a defined cycle the parser will be activated to start the Extraction-Transform-Load (ETL) process. Furthermore, BioDWH supports the technique of object-relational mapping (ORM).

The database layer is essential for the DAWIS-M.D. information system. It contains all information of the integrated data. Furthermore, modules for administration and meta data for the 12 biological domains are located in this layer. As a relational backend for DAWIS-M.D. data warehouse the open source RDBMS MySQL is used. The databases layer communicates via JDBC with the persistence layer. The persistence layer provides all components for the ORM, whereby the application layer is independent. Thus, it is possible to support different RDBMS such as MySQL, Oracle or PostgreSQL without changing the application logic. Accordingly, the ORM technique enables a high level of flexibility and independence between database layer and application logic. The persistence layer was realized by using the open source framework Hibernate.

The application layer communicates with the persistence layer. A user can interact via the application layer with the system. Therefore, the web application provides a homogenous and integrated view of the data. A web service controls the communication between the server and external tools and the network editor VANESA. The web service communicates with external applications by using SOAP. For the implementation of the web service the open source framework Apache Axis2 is used. A user is able to send a single element or all elements of a DAWIS-M.D. entry to the VANESA network editor. Then, the VANESA software is able to generate a network of the entry, i.e. an entry of the web application. Furthermore, the resulting network can be extended and edited in VANESA by the scientist.

The web-based graphical user interface of DAWIS-M.D. is implemented with JSP that is based on the programming language Java and runs on an Apache Tomcat web server. For dynamic and interactive websites a web server with JSP-/Servlet engine and additional Web 2.0 technologies are required. Each domain provides a special search form to formulate simple queries to the data warehouse. The results of a query to the database are presented on a new web site. The information is presented as a table. Information that is larger than a certain length is hidden to keep a clear and well-structured overview. This information can be displayed by clicking a specific keyword. Using this ”toggle” function it is possible to show important information and also to hide information which is not of interest.

Moreover, the web application provides a local navigation bar that is shown in figure 1. It is possible to navigate easily and fast within the result page. The navigation bar is only displayed for large result pages. Furthermore, the relationships and interactions between the different biological elements such as genes and proteins are presented and highlighted for each entry, if available. For proteins and enzymes, a graphical representation of a protein classification, using SCOP hierarchy, is available in DAWIS-M.D.

One of the most important features of DAWIS-M.D. is the interaction with the network editor VANESA. The web application provides two different views for the interaction.



(a) Navigation bar (b) Standard view (c) Expert view

Figure 1: Graphical representation of the navigation and the communication bridge between DAWIS-M.D. and VANESA.

Figure 1 shows the Standard and Expert view of the communication bridge. Using the standard view a user is able to display the root element of the entry in the web application as a node in the network editor. The functionality of the expert view is nearly the same, but in this case the user is able to select elements more precisely. Therefore, it is possible to create user-defined networks in VANESA to the current entry.

Finally, the data warehouse information system provides additional information about the integrated databases such as release or update information. The statistics are presented as table or diagram. Based on the statistics it is clear how the database of DAWIS-M.D. is composed.

4 Summary

One of the major challenges in bioinfomatics is the integration and management of data from different sources and their presentation in a user-friendly format. Therefore, in this paper we presented DAWIS-M.D., a platform-independent data warehouse information system for metabolic data. The information system integrates data from 11 widely-used life science databases. The information of integrated databases is divided into 12 various biological domains, which are available via the graphical user interface of the web application.

The data warehouse architecture provides a platform-independent web-interface that can

be used with any common web browser. The system enables intuitive search of integrated life science data, simple navigation to related information as well as visualization of biological domains and their relationships. To ensure maximum up-to-dateness of the integrated data the BioDWH data warehouse infrastructure including a monitor component is used. The persistence layer of DAWIS-M.D. uses the ORM technique whereby the application layer is independent from database layer. Thereby, it is possible to support different database management systems.

The DAWIS-M.D. data warehouse incorporates the advantages of a navigation and informational system and builds a bridge to the network editor approach VANESA. Hence, it is possible to browse through the integrated life science data and bring the information into a modeling and visualization environment. Therefore, it is easy for the scientists to search information of interest, find relationships and interactions between different biomedical domains and bring them for editing, manipulation and analyzing directly into the VANESA network editor. Finally, the scientists gain a better understanding of complex biological problems and are able to develop new theoretical models for further experiments. DAWIS-M.D. is available at http://agbi.techfak.uni-bielefeld.de/DAWISMD/.

References

[CG10] G. R. Cochrane and M. Y. Galperin. The 2010 Nucleic Acids Research Database Issue and online Database Collection: a community of data resources. *Nucleic Acids Research*, 38:D1–D4, 2010.

[CMB+07] C. Choi, R. Münch, B. Bunk, J. Barthelmes, C. Ebeling, D. Schomburg, M. Schobert, and D. Jahn. Combination of a data warehouse concept with web services for the establishment of the Pseudomonas systems biology database SYSTOMONAS. *Journal of Integrative Bioinformatics*, 4(1), 2007.

[EUA96] T. Etzold, A. Ulyanov, and P. Argos. SRS: Information retrieval system for molecular biology data banks. *Methods in Enzymology*, 266:114–128, 1996.

[HSK+01] L. M. Haas, P. M. Schwarz, P. Kodali, E. Kotlar, J. E. Rice, and W. C. Swope. DiscoveryLink: a system for integrated access to life sciences data sources. *IBM Systems Journal*, 40:489–511, 2001.

[JKT+10] S. Janowski, B. Kormeier, T. Töpel, K. Hippe, R. Hofestädt, N. Willassen, R. Friesen, S. Rubert, D. Borck, P. Haugen, and M. Chen. Modeling of cell-to-cell communication processes with Petri nets using the example of quorum sensing. *In Silico Biology*, 10:0003, 2010.

[KBT+06] J. Köhler, J. Baumbach, J. Taubert, M. Specht, A. Skusa, A. Rüegg, C. Rawlings, P. Verrier, and S. Philippi. Graph-based analysis and visualization of experimental results with ONDEX. *Bioinformatics*, 22(11):1383–1390, 2006.

[LPW+06] T. J. Lee, Y. Pouliot, V. Wagner, P. Gupta, D. W. J. Stringer-Calvert, J. D. Tenenbaum, and P. D. Karp. BioWarehouse: a bioinformatics database warehouse toolkit. *BMC Bioinformatics*, 7, 2006.

[TKKH08] T. Töpel, B. Kormeier, A. Klassen, and R. Hofestädt. BioDWH: A Data Warehouse Kit for Life Science Data Integration. *Journal of Integrative Bioinformatics*, 5(2), 2008.

Integrierte Informationsverwaltung für die Lebenswissenschaften mit der Information Workbench

Tobias Mathäß
fluid Operations
Altrottstr. 31
69190 Walldorf, Germany
tobias.mathaess@fluidops.com

Peter Haase
fluid Operations
Altrottstr. 31
69190 Walldorf, Germany
peter.haase@fluidops.com

Abstract: Forscher in den Lebenswissenschaften sind täglich mit den Problemen im Umgang mit riesigen Mengen heterogener und verteilter Daten konfrontiert. Oftmals ist es nicht möglich, komplexe Informationsbedürfnisse mit Hilfe von nur einer einzigen Datenquelle zu beantworten, sondern nur durch Kombination von Wissen aus verschiedenen Datenquellen. In diesem Beitrag stellen wir die Information Workbench vor: eine kollaborative Informations-Management-Plattform, welche die Integration heterogener strukturierter und unstrukturierter Daten unterstützt, eine einheitliche Sicht auf diese Daten bietet und somit quellübergreifende Suche über den gesamten Datenbestand ermöglicht. Die flexible und erweiterbare Benutzeroberfläche der Information Workbench erlaubt die datenabhängige Anzeige zusätzlicher Informationen zu gespeicherten Datenelementen. In diesem Beitrag demonstrieren wir den Einsatz der realisierten Ansätze, insbesondere hinsichtlich Datenintegration, Suchfunktionalität und Datenpräsentation in der Domäne der Lebenswissenschaften.

1 Motivation

Die Verwaltung, Speicherung und Verarbeitung digitaler Daten in den Lebenswissenschaften stellt Entwickler von Informations-Management-Systemen vor große Probleme. Die Daten, welche von Wissenschaftlern im Rahmen von Experimenten erzeugen, sind in hohem Maße heterogen. Arbeiten mehrere Wissenschaftler gemeinsam mit einem solchen System und erzeugen diese die anfallenden Daten nicht nach gemeinsamem Schema trägt dies weiter zur Heterogenität des Datenbestandes bei.

Hinzu kommt die Tatsache, dass auch öffentlich zugängliche Datensätze selten nach einem gemeinsamen Schema entworfen und erzeugt wurden. Es wird ein Werkzeug benötigt, welches eine gemeinsame Sicht auf die Daten ermöglicht. Dies gibt Forschern die Möglichkeit, Daten aus verschiedenen Datenquellen gleichzeitig und integriert zu nutzen. Diese Integration sollte transparent für die Benutzer sein, d.h. die Interaktion mit den integrierten Daten soll genau so von statten gehen wie mit einem einzigen Datensatz.

Die Information Workbench ist in der Lage, große Datenbestände zu verwalten, unabhängig davon ob diese Daten in strukturierter oder unstrukturierter Form vorliegen. Sie erlaubt den Benutzern, Daten zu verändern oder neu zu erzeugen, bereitet relevante In-

formationen sinnvoll auf und gibt dem Benutzer die Möglichkeit, den Datenbestand zu explorieren, also sich von einem Datenelement aus zu verwandten Elementen durch den Datengraphen zu bewegen. Eines der wichtigsten Merkmale der Information Workbench ist die Unterstützung für verschiedene Arten der Suche. Je nach Art des Informationsbedürfnisses und nach Fähigkeiten des Benutzers bieten sich bestimmte Suchparadigmen an. Das Spektrum der umgesetzten Arten der Suche reicht von einer Volltext-Suche über die strukturierten und unstrukturierten Daten über formular-gestützte Suche, die den Benutzer bei der Formulierung strukturierter Anfragen unterstützt, bis hin zu Anfragen in einer strukturierten Anfragesprache.

Für die Datenintegration und Wissensrepräsentation baut die Information Workbench auf den Standards des Semantic Web und Linked Data auf, welche wir im Folgenden kurz einführen.

2 Grundlagen

In diesem Abschnitt stellen wir grundlegende Konzepte wie das Resource Description Framework (RDF) – das Standard-Datenmodell des Semantic Web – und die Linking Open Data Initiative (LOD) vor.

Das Resource Description Framework Das im Kontext des Semantic Web wichtigste Datenmodell ist das Resource Description Framework (RDF), welches vom World Wide Web Consortium (W3C) standardisiert wurde[1]. Ein RDF-Datenbestand besteht aus so genannten Statements. Ein Statement besteht aus einem Subjekt, einem Prädikat und einem Objekt. Das Subjekt ist hierbei ein Datenelement, eine so genannte Entität, das Objekt ist entweder eine weitere Entität, welche mit dem Subjekt in Beziehung steht, oder ein das Subjekt beschreibender Datenwert. Das Prädikat beschreibt die Art der Beziehung, in der Subjekt und Objekt stehen. Subjekt und Objekt eines solchen Statements können als Knoten im Datengraphen angesehen werden, das Prädikat als die verbindende Kante. Da die gleiche URI in verschiedenen Statements als Subjekt oder als Objekt benutzt werden kann definiert eine Menge von Statements einen gerichteten Datengraphen. Ein solcher Datengraph kann mit Hilfe von formalen Anfragesprachen angefragt werden.

Linked Open Data Die Idee hinter Linked Open Data (LOD) ist die Möglichkeit der Vernetzung verschiedener, eigenständiger RDF-Datensätze, indem URIs aus anderen Datensätzen benutzt werden. Betrachtet man zwei solcher Datensätze aus einer integrierten Sicht, so ergänzen sich diese zu einem gesamtheitlichen Bild. Auf Basis dieser Idee ist es möglich, bestehende Daten um bestimmte Aspekte zu erweitern. Enthält beispielsweise ein Datensatz Informationen über chemische Reaktionen sowie die an den Reaktionen beteiligten chemischen Verbindungen und ein weiterer Datensatz Detailinformationen über diese chemischen Verbindungen, so gibt eine integrierte Sicht auf beide Datensätze dem Benutzer die Möglichkeit. für eine chemische Verbindung sowohl alle Reaktionen, an

[1] http://www.w3.org/TR/WD-rdf-syntax

denen diese Verbindung beteiligt ist, als auch weitere Informationen wie Schmelzpunkt, Molekülmasse oder Toxizität gleichzeitig einzusehen und zu analysieren.

Bio2RDF Das Bio2RDF-Projekt[BNT+08] hat zum Ziel, Daten die in der Domäne der Bioinformatik erzeugt wurden in das RDF-Format zu überführen, um sie dann als Teil des LOD-Projektes zu veröffentlichen. Hierzu bietet Bio2RDF Werkzeuge an, die es Wissenschaftlern ermöglichen, ihre Daten mit bereits existierenden Bio2RDF-Daten abzugleichen, und in die bestehende Linked Open Data Umwelt einzupflegen.

3 Szenario

In diesem Abschnitt sollen anhand eines Beispielszenarios die Anforderungen an ein Informationsmanagement-System, das Einsatz in der Domäne der Lebenswissenschaften findet, erläutert werden.

Ein Wissenschaftler in einem Pharmaunternehmen, der auf der Suche nach einem Heilmittel oder einem Impfstoff gegen das HIV-Virus ist, versucht die HIV-Infektion zu unterbinden, indem er den Prozess der Ansteckung unterbricht. Der Forscher versucht dies zu erreichen, indem er dem Organismus für eine für die Ansteckung notwendige Reaktion notwendige Katalysatoren entzieht.

Um eine Übersichtsseite über den HIV-Infektionsprozess sowie alle in dessen Verlauf statttfindenden chemischen Reaktionen zu erhalten, muss dem Wissenschaftler die Möglichkeit gegeben werden, eine strukturierte Anfrage auszuwerten. Da dem Benutzer die genaue Struktur der geladenen Daten unter Umständen nicht bekannt ist, muss er bei der Formulierung dieser Anfrage unterstützt werden. Der Wissenschaftler findet so heraus, dass sehr früh im Prozess eine Umformung der Ribonukleinsäure (RNA) stattfindet. Nachforschungen über diese Reaktion ergeben, dass das Protein Xeroderma Pigmentosum B (XPB) ein für die Reaktion notwendiger Katalysator ist. Durch Experimente verifiziert der Wissenschaftler, dass der Gesamtprozess durch Entzug von XPB unterbrochen werden kann. Er notiert eine Beschreibung der Durchführung seiner Experimente sowie deren Ergebnisse und speichert sie in der Datenbank, sodass andere Mitarbeiter diese einsehen können, um das gleiche Experiment nicht wiederholen zu müssen.

Nun versucht der Wissenschaftler weitere Informationen über das Protein XPB zu finden. Hierzu müssen Volltextsuch-Anfragen unterstützt werden, die ihm bei einer Suche nach dem Schlüsselwort “XPB“ Ergebnisse sowohl aus den strukturierten als auch den unstrukturierten Teilen der Daten liefern. Zusätzlich sollen weitere Informationen über Medikamente, welche dieses Protein enthalten, und eine Liste der Reaktionen, an denen XPB beteiligt ist, geliefert werden. Der Forscher findet so heraus, dass XPB für viele im menschlichen Körper ablaufende DNA- und RNA-Reparatur-Vorgänge notwendig ist. Er sucht mit Hilfe eines Suchformulars nach chemische Verbindungen mit ähnlichen chemischen Eigenschaften wie XPB, die als Ersatzstoff in diesen Reparatur-Prozessen dienen könnten. Das Ergebnis dieser Anfrage wird mit Hilfe von Diagrammen visualisiert, was einen direkten analytischen Vergleich der gefundenen Verbindungen erlaubt. Da die Men-

ge der so gefundenen Stoffe sehr groß ist, muss dem Benutzer die Möglichkeit gegeben werden, diese Menge auf eine Teilmenge einzuschränken, die gemeinsame strukturelle Eigenschaften haben.

Der Wissenschaftler findet auch auf diesem Wege keine geeigneten Ersatzstoffe. Da die voraussichtlichen Nebenwirkungen eines Entzuges von XPB aus dem System zu gravierend wären muss der Wissenschaftler ausgehend von der Übersicht über alle Reaktionen im Rahmen des Infektionsprozesses von neuem mit der Suche nach aussichtsreichen Reaktionen, welche unterbunden werden könnten, beginnen.

Zusammenfassend stellt die Suche nach Informationen zur Lösung von Problemen in den Lebenswissenschaften hohe Anforderungen an ein Informations-Verwaltungs-System in den Bereichen der transparenten Datenintegration, der Datenquellen übergreifenden Suche und Analyse sowie der Datenmanipulation und -annotation.

4 Information Workbench

Die Information Workbench ist eine Plattform für die kollabarative Verarbeitung von Informationen. Dabei werden insbesondere die folgenden Prozesse in der Interaktion mit den Informationen unterstützt:

- Integration von heterogenen und verteilten Informationsquellen,
- Erzeugung von Informationen durch den Endnutzer, z.B. in Form von wiki-basierter Dokumentation and Annotation,
- Browsing und Navigation über die aggregierten Informationen,
- Visualisierung von und Interaktion mit den Informationen über eine Vielzahl von Widgets,
- Integrierte Suche und Exploration,
- Verwaltung von Provenance, d.h. Daten über die Herkunft der Informationen.

In diesem Beitrag sollen die Fähigkeiten der Information Workbench hinsichtlich der Datenintegration und der verschiedenen Arten der Suche und Exploration näher erläutert werden.

4.1 Datenintegration

Die Information Workbench ist in der Lage, sehr große Mengen von Daten zu integrieren. Neben der Möglichkeit der zentralisierten Integration, bei der die Datensätze in eine (lokale) Datenbank geladen werden, ist es alternativ möglich, auf verteilte Datenquellen föderiert zuzugreifen. Dies hat den Vorteil hat, dass neue Datenquellen einfach integriert und wieder entfernt werden können. Die restlichen Datenquellen bleiben davon unberührt.

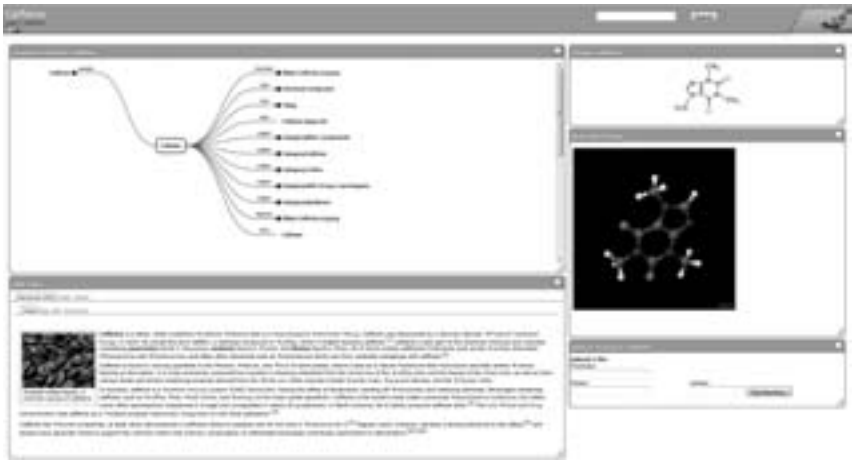


Abbildung 1: Die Entität Koffein

Eine dritte Möglichkeit ist das Einbinden von entfernten Datenquellen über einen SPARQL-Endpunkt. Dies stellt den leichtgewichtigsten Integrationsmechanismus dar, da lokal keinerlei Veränderungen vorzunehmen sind, und das Einbinden und wieder Entfernen aus der Datenbank-Föderation somit sehr schnell und einfach geht. Die Information Workbench erlaubt beliebige Kombinationen dieser drei Integrationsmöglichkeiten. Eine Performance-Analyse der verschiedenen Möglichkeiten der Datenintegration wurde in [HMZ10] durchgeführt. Sowohl zur lokalen Datenspeicherung als auch für die Föderierung wird das Sesame-Framework benutzt[2].

Abbildung 1 zeigt exemplarisch die Detailseite für die Verbindung Koffein. Sie stellt in verschiedenen Widgets die aggregierten Daten aus den Datenquellen integriert dar: Der integrierte Datengraph zeigt strukturierte Daten und Relationen zwischen den Entitäten, ein (semantisches) Wiki-Widget ermöglicht Zugriff auf semi- und unstrukturierte Informationen. Widgets für chemische Verbindungen zeigen automatisch die chemische Struktur der Substanz. Die Auswahl geeigneter Widgets erfolgt automatisch in Abhängigkeit vom Typ der Entität.

4.2 Suchparadigmen

Eine der größten Stärken der Information Workbench ist die Unterstützung sehr unterschiedlicher Suchparadigmen für verschiedene Arten von Informationsbedürfnissen, angefangen von einfacher Volltextsuche über Formular gestützte Suche bis hin zur Unterstützung von SPARQL als strukturierte Anfragesprache[TMH10]. Im Folgenden sollen die Suchparadigmen detailliert vorgestellt und hinsichtlich ihrer Eignung für verschiedene Informationsbedürfnisse bewertet werden.

[2] http://www.openrdf.org

Hybride Suche: Die Information Workbench unterstützt Volltextsuche sowohl auf den strukturierten RDF-Daten, als auch auf den unstrukturierten textuellen Daten, welche für jede Entität gespeichert werden können. Die Volltextsuche arbeitet Entitäts-zentriert, d.h. Ergebnisse einer Volltextsuche sind Entitäten, bei denen die gesuchten Wörter in die Entität beschreibenden Attributen oder im assoziierten unstrukturierten Teil der Daten vorkommen. Die Ergebnisse werden nach Relevanz bezüglich der Suchwörter geordnet. Desweiteren hat der Benutzer die Möglichkeit, das Ergebnis einer solchen Suche mit Hilfe dynamisch aus den strukturierten Daten erzeugter facettierter Suche weiter zu verfeinern. Volltextsuche bietet einen sehr einfachen Zugang zu den Daten, auch für Benutzer die mit analytischen und strukturierten Suchverfahren wenig vertraut sind, oder keinerlei Kenntnis über die Struktur und das Schema der unterliegenden Daten haben. Daher kann die Volltextsuche ein geeigneter Einstiegspunkt beim Arbeiten mit neuen, unbekannten Daten sein, um sich einen schnellen Überblick zu verschaffen und um Einstiegspunkte in den Datengraphen zu finden, von denen aus die Daten weiter exploriert werden können.

Für die Realisierung der Volltextsuche wurde die Apache Lucene Bibliothek, eine Java-basierte Volltext-Engine, benutzt[3]. Sowohl der RDF-Graph als auch alle Wiki-Seiten werden indexiert und mit einer Entität assoziiert.

Ein Beispiel hierfür wäre die Suche nach einem sehr allgemeinen Schlüsselwort wie "drug", um sich einen schnellen Überblick über Medikamente und verwandte Daten zu verschaffen. Der Benutzer kann sich dann einige Ergebnisse anschauen, und sich so einen genaueren Überblick über die verfügbaren Daten und deren Struktur verschaffen, um dann in der Lage zu sein, präzisere Anfragen zu formulieren, oder direkt, unterstützt durch facettierte Suche, die Ergebnismenge auf für ihn interessante Teile der Gesamtmenge einzuschränken.

Expressive Suche ohne Schemakenntnisse: Die Information Workbench unterstützt die RDF-Anfragesprache SPARQL in vollem Umfang. SPARQL ist eine deklarative Anfragesprache, welche in ihrer Struktur Anfragesprachen anderer Datenmodelle, wie zum Beispiel SQL, ähnelt. Diese Art der Suche lässt eine in höchstem Maße präzise Formulierung von Anfragen zu, welche zu sehr exakten Ergebnissen führen. Allerdings kann vom Endnutzer nicht erwartet werden, dass er Erfahrung im Umgang mit komplexen Anfragesprachen hat. Desweiteren ist für das Formulieren korrekter SPARQL-Anfragen detailliertes Wissen über das Schema der Daten notwendig. Auch dies ist bei Benutzern von Informationsmanagement Systemen nicht zwingend der Fall. Daher bietet die Information Workbench verschiedene Möglichkeiten, um den Benutzer bei der Formulierung solcher Anfragen zu unterstützen. Diese reichen von einer formular-gestützten Suche, welche dazu geeignet ist, häufige und somit wahrscheinliche Arten von Anfragen zu formulieren, bis hin zur Interpretation von Schlüsselwörtern, bei der versucht wird eine untrukturierte Eingabe des Benutzers in eine konkrete strukturierte Anfrage zu übersetzen und diese auszuwerten. Hierzu wird ein Strukturindex verwaltet, welcher es ermöglicht Schlüsselwörter auf Knoten des RDF-Graphen zu matchen und durch eine Graph-Exploration verbindende Subgraphen zu finden, welche Datenelemente für alle Schlüsselwörter enthalten. Theoretische Grundlagen dieses Verfahrens können in [TWRC09] nachgelesen werden.

[3] http://lucene.apache.org/java/docs

Auch bei diesen Arten der Suche ermöglicht die Information Workbench zusätzlich das Verfeinern der gefundenen Ergebnisse durch facettierte Suche.

Ein Beispiel hierfür ist die Suche nach allen Entitäten vom Typ "Compound", zusammen mit ihrem Schmelzpunkt und ihrer chemischen Formel. Der Benutzer hat verschiedene Möglichkeiten um zum gesuchten Ergebnis zu gelangen: Entweder er formuliert selbst eine SPARQL-Anfrage, oder er erzeugt mit Hilfe eines Suchformulars durch Eingabe des Typs "Compound", eine Liste aller chemischen Verbindungen und wählt zusätzlich aus der Liste der Eigenschaften die angezeigt werden sollen die gewünschten Attribute "meltingPoint" und "chemicalFormula", oder er gibt die Anfrage "compound melting-point formula" ein und lässt diese als strukturierte Anfrage interpretieren. All diese Wege führen zum gleichen Anfrageergebnis, welches dann weiter verfeinert werden kann, beispielsweise auf Elemente, deren Schmelzpunkt in einem vorgegebenen Intervall liegen.

Suche nach chemischen Verbindungen: Speziell für Anwendungen in Chemie-affinen Domänen wie den Lebenswissenschaften wurde die Information Workbench mit einem Suchinterface für chemische Verbindungen ausgestattet. Hierbei hat der Benutzer die Möglichkeit, ein Molekül zu zeichnen und nach diesem Molekül in der Datenbank zu suchen. Hierzu wird die SMILES-Codierung der gezeichneten Struktur berechnet und mit Einträgen in der Datenbank verglichen.

Dieses Interface ist sehr speziell und domänenspezifisch. Es findet genau dann Anwendung, wenn dem Benutzer die Struktur einer gesuchten chemischen Verbindung bekannt ist, er aber deren Namen nicht kennt. Eine Suche nach Sub- oder Superstrukturen wäre an dieser Stelle ebenfalls denkbar.

5 Schluss

In diesem Beitrag haben wir die Information Workbench als Plattform für die Verwaltung von Informationen in den Lebenswissenschaften vorgestellt, der Fokus lag dabei auf Aspekten der Informationsintegration und neuen Paradigmen der Suche.

Zur Zeit wird die Information Workbench in Fallstudien von Forschungsprojekten (z.B. CollabCloud[4]) eingesetzt. Um die Verbreitung zu erhöhen und einfache Erweiterungen zuzulassen, wird der Kern der Information Workbench als Open Source Projekt entwickelt[5]. Die Erweiterung des Kerns ist über ein SDK einfach möglich, so können neue Datenquellen über zusätzliche Provider oder neue Widgets zur Visualisierung und Interaktion integriert werden. Die Entwicklung von kommerziellen Erweiterungen und Produkten ist möglich und geplant. Ein öffentlicher Demonstrator der Information Workbench mit freien Datensätzen aus den Lebenswissenschaften ist zugänglich unter `http://iwb.fluidops.com/index_ls.html`.

[4] `http://www.collabcloud.de`
[5] `http://code.google.com/p/iwb`

Literatur

[BNT+08] F. Belleau, M.-A. Nolin, N. Tourigny, P. Rigault und J. Morissette. Bio2RDF: Towards a mashup to build bioinformatics knowledge systems. *Journal of Biomedical Informatics*, 41(5):706–716, 2008.

[HMZ10] Peter Haase, Tobias Mathäß und Michael Ziller. An Evaluation of Approaches to Federated Query Processing over Linked Data. In *To appear at the I-SEMANTICS 2010*, September 2010.

[TMH10] Thanh Tran, Tobias Mathäß und Peter Haase. Usability of Keyword-Driven Schema-Agnostic Search. In *Proceedings of the 7th Extended Semantic Web Conference, ESWC 2010*, Seiten 349–364, 2010.

[TWRC09] Thanh Tran, Haofen Wang, Sebastian Rudolph und Philipp Cimiano. Top-k Exploration of Query Candidates for Efficient Keyword Search on Graph-Shaped (RDF) Data. In *ICDE*, Seiten 405–416. IEEE, 2009.

Bringing Semantics to the i2b2 Framework

Matthias Löbe, Sebastian Stäubert

Institute for Medical Informatics, Statistics and Epidemiology (IMISE)
University Leipzig
Härtelstraße 16-18
04107 Leipzig
matthias.loebe@imise.uni-leipzig.de
sebastian.staeubert@imise.uni-leipzig.de

Abstract: i2b2 has proven to be a valuable tool for a variety of tasks in clinical and translational research. But to establish an extended IT infrastructure, several clinical information systems have to be connected seamlessly and routine data integration processes have to be defined. In this position paper, we will outline a number of options that would enhance the existing software stack with regards to semantic precision and maintainability and would expand i2b2's field of application.

1 Introduction

I2b2 is an Open Source framework developed by the i2b2 National Center for Biomedical Computing in Boston. It has a long history going back to 1989 [WM09]. I2b2 takes a modular approach using the analogy of a "hive" of independent modules called "cells" with each having its own database schema that communicate over SOAP/XML Web Services or REST calls. Therefore, they are not tied to a certain programming language and can be distributed over a network of computers. Some cells are required for a basic infrastructure, some are optional. The most important cells are called *clinical research chart* (CRC), which contains all clinical facts and observations and *ontology* (ONT), which contains parameters to query over the facts in the CRC. I2b2 offers a stand-alone client (i2b2 workbench) and an AJAX web client with plug-ins corresponding to each cell.

I2b2 is primarily a free data warehousing / data mining tool. Its intended use includes:

1. Controlling: analyzing typical performance indicators
2. Cohort selection: recruitment of patient for inclusion in clinical trials [DM09],
3. Cohort estimation: prospective approximation of patient cohorts by retrospective queries [DA09]
4. Verification of hypotheses: meta-analyses of clinical trials
5. Aggregating genomic data [MC09]

2 Motivation

We evaluated the i2b2 framework in version 1.4 as part of an effort to use EHR data for clinical research purposes. Our goal was to identify parameters in our local HIS that relate to common inclusion or exclusion criteria which would help to find eligible patients and estimate cohort sizes for clinical trials prospectively.

The first step was to import de-identified patient's data, visits and observations. Even for large record sets of about one million observation facts, this is quite a straight-forward task. For performance reasons, i2b2 is using a star schema [MW10], which means that queries require only a few SQL join operations.

It was more cumbersome to create and maintain an efficient vocabulary. I2b2 doesn't come with any medical or master data vocabulary included, so all relevant variables had to be imported and classified manually. Furthermore, i2b2 doesn't provide any tools for supporting this task. Even for widely-used terminologies like classifications of diseases or procedures, no import is available for common file formats like character separated values (CSV) or the classification markup language (ClaML) [HB06]. As a result, this leads to an extensive use of large SQL scripts. The vocabulary created that way is hard to maintain: all hierarchy levels, concept labels and path names are hard-coded and therefore items are difficult to change or move later on. Vocabularies can not even be shared among different i2b2 instances in case different database systems are used due to SQL syntax variants. The hierarchy tree is parsed by tokenizing the full name as well as most queries are made using the LIKE operator – an error prone approach with regards to real world scenarios.

3 Enhancing Interoperability and Semantic Integration

In this chapter we propose a number of improvements and extensions to the i2b2 architecture and framework to extend its range of application.

3.1 Ontologizing the Ontology Cell

From our point of view, the Ontology cell (ONT) has much potential for improvement. The i2b2 team uses the term "ontology" in its weakest sense. It's not more than a controlled vocabulary. Furthermore, it's ambiguous whether elements of ONT represent terms or concepts. On the one hand, i2b2 defines *concepts* as facts in the clinical research chart [MM07]. On the other hand, *terms* in ONT can have names, label, codes, synonyms, and are group in Schemes – for staying in line with the majority of knowledge organization systems, we will assume that they are concepts as "units of thought". Concepts are arranged in a hierarchy but without enforcing a formal subtype assertion (A is subtype of B => every a instance of A is also instance of B). Concepts in i2b2 have no explicit association to units of measurement. This is precarious if concepts seem to have an obvious unit, e.g. years for an "age" concept, which might not be true for pediatric data sets.

We argue that the ONT cell should be rewritten using more suitable technologies. The Simple Knowledge Organization System (SKOS) [MB09], a W3C recommendation for representing thesauri, classifications and other controlled vocabularies, could be used as a data model for ONT. SKOS is a lightweight RDF vocabulary developed as part of the Semantic Web. The basic building block is the *concept*: an abstract unit of thought. Concepts are represented by an URI that is asserted to be of type skos:Concept. While SKOS has a weaker semantic than ontologies implemented in OWL or higher logics, it is closer to the conceptual model of wide-spread medical terminologies like ICD. SKOS provides a number of modeling features that fit to the existing i2b2 ONT cell model.

- ONT concepts would be modeled as SKOS concepts
- SKOS has a relation for modeling hierarchies called *broader* that would replace the complicated handling of levels and tokenizers in ONT.
- ONT attributes like *name* or *tooltip* would be modeled as SKOS labels
- ONT attributes like *comment* would be modeled as SKOS notes
- ONT attributes like *codes* would be modeled as SKOS notations
- ONT *synonyms* would be modeled as SKOS mappings
- ONT values like *level 1* would be modeled as SKOS Concept Schemes

Some ONT attribute may have no counterpart in SKOS. That is no problem since RDF is easily extendable. Queries to the knowledge base would be realized using the SPARQL Protocol and RDF Query Language. Advantages of that approach include:

- Less error-prone: explicit hierarchy instead of implicit parsing of strings
- Better maintenance: possibility to use existing SKOS tools for creating, editing, removing or moving entries
- Using standards: medical terminologies in SKOS format instead of creating catalogs in SQL
- Federations: Queries over distant i2b2 hives can be made if harmonized concepts schemes are used
- Normalization: no redundant data chunk in a lot of database rows

3.2 Expanding the type of source data repositories in the CRC to EDC systems

The CDISC Operational Data Model (ODM) is the de-facto standard for storing and exchanging clinical trial data. Most current EDC systems are able to import and export ODM, so ODM is a natural source of clinical data as well as of metadata. Therefore, it is reasonable to have an ONT cell parsing ontology concepts from the trial specification (metadata section) and a CRC cell reading facts from subject data sections. This would allow using the aggregation and graphical visualization tools of i2b2 for clinical trial analysis too.

3.3 Sharing the semantic meaning of concepts in a metadata repository

As mentioned before, a clear and detailed description of i2b2 concepts used for instant queries as well as long-term interpretation of results is advantageous. Many vocabulary items can have different meanings in a different context, for instance with regard to precision, measurement units, normal ranges and so on. Currently, there is an effort to establish a metadata repository for data elements in clinical and translational research [SL09]. It would be useful to have an interface for accessing such harmonized, commonly used metadata items.

4 Discussion

In this article, we proposed to enhance the ontology representation in i2b2 and using SKOS as the native data model. First of all, SKOS is an international standard while i2b2's current schema is proprietary. As a consequence, a SKOS implementation could rely on external tools like Protégé for vocabulary construction and maintenance. SKOS concepts have a unique identifier (URI) rather than a lexical string and therefore are referable outside a certain application. SKOS supports explicit hierarchies, even multi-hierarchies and mapping of concepts (exact matches, more specific or more general concepts). Combined with the easy extensibility, one could add a concept for mantle cell lymphoma, classify it as being more specific than the ICD-10-GM concept labeled "Diffuse non-Hodgkin's lymphoma" and assert an exact match on the ICD-11 concept coded "C83.1". Another advantage of SKOS is the opportunity to group concept into collections, which can be labeled and ordered. This is especially necessary for case report forms in clinical trials which are nested collections of data elements in a defined sequence. Finally, SKOS can improve searching in the ontology. Search could benefit from the option to provide alternative labels and even hidden labels for common but deprecated names that should not be displayed but redirected.

The points discussed herein are work in progress and therefore not proven by an implementation or evaluation.

Acknowledgement: Sebastian Mate's thesis [Ma09] on deploying the i2b2 framework gave us a good impression on the internal structure and available feature.

References

[DA09] Dugas, M., Amler, S., Lange, M., Gerss, J., Breil, B., Köpcke, W.: Estimation of patient accrual rates in clinical trials based on routine data from hospital information systems. Methods of information in medicine 48(3), 263-266 (2009).

[DM09] Deshmukh, V., Meystre, S., Mitchell, J.: Evaluating the informatics for integrating biology and the bedside system for clinical research. BMC Medical Research Methodology 9, 70+, October 2009.

[HB06] van der Haring, E. J., Broënhorst, S., ten Napel, H., Weber, S., Schopen, M., Zanstra, P. E.: ClaML: a standard for the electronic publication of classification coding schemes. Studies in health technology and informatics, 2006;124, S. 801-806.

[Ma09] Mate, S.: Evaluation von i2b2 am Universitätsklinikum Erlangen. Studienarbeit. Lehrstuhl für Medizinische Informatik, Friedrich-Alexander-Universität Erlangen-Nürnberg, Augist 2009.

[MC09] Murphy, S., Churchill, S., Bry, L., Chueh, H., Weiss, S., Lazarus, R., Zeng, Q., Dubey, A., Gainer, V., Mendis, M., Glaser, J., Kohane, I.: Instrumenting the health care enterprise for discovery research in the genomic era. Genome Research 19 (9), September 2009, S. 1675–1681.

[MB09] Miles A, Bechhofer S (Editors): SKOS Reference, W3C Recommendation 18 August 2009

[MM07] Murphy, S. N., Mendis, M., Hackett, K., Kuttan, R., Pan, W., Phillips, L. C., Gainer, V., Berkowicz, D., Glaser, J. P., Kohane, I., Chueh, H. C.: Architecture of the open-source clinical research chart from informatics for integrating biology and the bedside. AMIA. Annual Symposium proceedings, 2007, S. 548–552.

[MW10] Murphy, S. N., Weber, G., Mendis, M., Gainer, V., Chueh, H. C., Churchill, S., Kohane, I.: Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2). Journal of the American Medical Informatics Association: JAMIA 17 (2), March 2010, S. 124–130.

[SL09] Stausberg, J., Löbe, M., Verplancke, P., Drepper, J., Herre, H., Löffler, M.: Foundations of a metadata repository for databases of registers and trials. Studies in health technology and informatics, 2009;150, S. 409-413.

[WM09] Weber, G. M., Murphy, S. N., McMurry, A. J., MacFadden, D., Nigrin, D. J., Churchill, S., Kohane, I. S.: The shared health research information network (shrine): A prototype federated query tool for clinical data repositories. Journal of the American Medical Informatics Association 16 (5) , June 2009, S. 624–630.

Eine Umsetzung des IHE Single Source Konzepts für die translationale Forschung bei Knochen- und Weichteilsarkomen

G. Lamla[1], R. Blaser[1], F. Prasser[1], S.H.R. Wurst[1], H. Rechl[2], R. Gradinger[2], K.A. Kuhn[1]

[1] Lehrstuhl für medizinische Informatik
[2] Klinik und Poliklinik für Orthopädie und Unfallchirurgie
Klinikum rechts der Isar der TU München
gregor.lamla@tum.de

Abstract: Informationsintegration spielt für die Unterstützung der translationalen Forschung weltweit eine zentrale Rolle. Am Klinikum rechts der Isar der Technischen Universität München wurde Anfang 2010 eine Therapieeinheit für Knochen- und Weichteilsarkome eingerichtet. Für Forschungszwecke werden über die für die Krankenversorgung erforderlichen Daten hinaus weitere Details zum Krankheits- und Behandlungsverlauf der Patienten hochstrukturiert erfasst. Hierfür werden relevante Teile der Dokumente aus dem Routinesystem in ein Forschungssystem übertragen. Dort erfolgt dann eine strukturierte Dokumentation höherer Granularität; hierfür sind Maßnahmen zur Konsistenzsicherung und Synchronisation vorgesehen. Das hier vorgestellte System ist von genereller Relevanz für die translationale Informationsverarbeitung.

1 Einleitung und Ziele

Informationsintegration, insbesondere die Zusammenführung phänotypischer und genotypischer Daten, spielt für die Unterstützung der translationalen Forschung weltweit eine zentrale Rolle [Na06, Hi07]. Translation beschreibt hier den Kreislauf von genomischer, molekularer und klinischer Datensammlung hin zu individualisierter Behandlung mit Verlaufsbeobachtung und Evaluation bis zur erneuten Hypothesengenerierung. Translationale Medizin eröffnet neue Einblicke in Krankheitsmechanismen und unter stützt so die Ermittlung persönlicher Risiken und die Festlegung personalisierter Therapien [Al08, Ku08].

Am Klinikum rechts der Isar der Technischen Universität München (MRI) wurde eine interdisziplinäre Therapieeinheit für Knochen- und Weichteilsarkome eingerichtet. Diese hat zur Aufgabe, die Diagnose und Therapie von Patienten mit Knochen- und Weichteilsarkomen zu verbessern und in begleitenden translationalen Forschungsprojekten innovative Konzepte für individualisierte Therapien zu entwickeln.

Als System zur Patientendatenverwaltung und als klinisches Arbeitsplatzsystem sind am MRI IS-H und i.s.h.med im Einsatz. SAP IS-H verwaltet u.a. die Patientenbewegungen (Aufnahme, Entlassung, Verlegung) [ISH]. Siemens i.s.h.med erweitert SAP IS-H um medizinische Funktionalitäten und um ein zentrales Repository für klinische Daten. Für

die Erfassung klinischer Daten können generische Dokumente, sog. Parametrisierbare Medizinische Dokumente (PMDs), erstellt werden. [ISM].
Im Verlauf ihrer interdisziplinären Behandlung durchlaufen Tumorpatienten der Therapieeinheit i.a. eine oder mehrere Tumorboardsitzungen, an denen Vertreter der an der Behandlung beteiligten Fachrichtungen teilnehmen. Dabei werden klinische Daten im Tumorboardmodul von i.s.h.med mit Hilfe von PMDs strukturiert erfasst. Die Formulare werden zu Beginn der Sitzung mit relevanten Inhalten aus i.s.h.med vorbefüllt. Dies ermöglicht es den beteiligten Ärzten verschiedener Fachrichtungen, gemeinsam Entscheidungen zu treffen und diese gemäß ihrer Dokumentationspflichten zu dokumentieren. Die in den PMDs erfassten Daten dienen somit der Krankenversorgung; sie sind aber zu einem großen Teil auch für die Beantwortung wissenschaftlicher Fragestellungen relevant. Rechtlich ist die Nutzung von Patientendaten für Forschungszwecke am MRI unter Beachtung der ärztlichen Schweigepflicht und des Datenschutzes durch Art. 27 des Bayerischen Krankenhausgesetzes abgedeckt.
Zur Unterstützung der translationalen Forschungsprojekte sollen in einer forschungsbezogenen Spezialdokumentation die Daten aus der Routinedokumentation um weitere, typischerweise feiner granulare, Details zum Krankheits- und Behandlungsverlauf ergänzt werden. Für die Erstellung einer solchen Spezialdokumentation steht am MRI das Clinical Data Management System (CDMS) MACRO der Firma InferMed zur Verfügung. MACRO wurde für das Datenmanagement in klinischen Studien hinsichtlich der regularischen Anforderungen aus ICH Good Clinical Practice [GCP] und FDA 21 CFR Part 11 [Part11] entworfen und validiert. Es stellt u.a. einen Formulargenerator sowie Datenmanagementfunktionalität (z. B. Plausibilitätsprüfungen) zur Verfügung. [Macro]. In dieser Arbeit wird eine forschungsbezogene Spezialdokumentation für die Therapieeinheit für Knochen- und Weichteilsarkome vorgestellt, die für die Übernahme von Daten aus dem Tumorboardmodul konzipiert wurde. Neben der eigentlichen Datenübernahme sind Konsistenzsicherung bzw. Synchronisation zentrale Herausforderungen.

2 Verwandte Arbeiten

Für die Weiterverwendung von in der Routine erfassten Daten für Forschungszwecke existieren verschiedene Ansätze. Innerhalb des Clinical Data Interchange Standards Consortium (CDISC) hat die Arbeitsgruppe Electronic Source Data Interchange (eSDI) fünf Szenarien für die Datenverarbeitung, Datenübernahme oder Datenweitergabe unter den Anforderungen von Studienregularien entwickelt [CDISC]. Im Szenario „Single Source“ werden Daten vom Prüfarzt einmal in ein elektronisches Formular eingeben und dann den weiterverarbeitenden Systemen zur separaten Verarbeitung zur Verfügung gestellt. Dabei dupliziert eine separate Anwendung die Datenerfassungsfunktionalität eines CDMS und erstellt aus den Eingaben Importdaten für die Empfängerkomponentensysteme. Das Konzeptpapier „Retrieve Form for Data Capture” (RFD) der Integrating the Healthcare Enterprise Initiative (IHE) definiert Komponenten und Abläufe für die Realisierung des „Single Source“ Szenarios. [RFD].
Die Machbarkeit des Ansatzes wurde im Rahmen der STARBRITE Studie [Kus07] evaluiert. Hierbei wurde zunächst die Schnittmenge der Daten ermittelt, die bereits im Behandlungszusammenhang erhoben und in einer Studie wiederverwendet werden. Ziel

von STARBRITE war eine IT-Unterstützung, bei der Daten nur einmal erfasst, aber mehrfach verwendet werden, und zwar in den elektronischen Formularen des Studiensystems und im klinischen Repository. Das durch STARBRITE realisierte System ist in der Lage, die erfassten Daten in CDISC Operational Data Model (ODM) für die Übernahme in das CDMS und in HL7 Clinical Document Architecture (CDA) für die Übernahme in das klinische Repository zu konvertieren [CDISC, HL7].

3 Methode

Der hier vorgestellte Ansatz orientiert sich am oben beschriebenen Single Source Szenario und am RFD-Konzept [RFD]. Die Routinedokumentation erfolgt dabei weiterhin in i.s.h.med. Für jede interdisziplinäre Besprechung eines Patienten wird ein Dokument im Tumorboardsystem von i.s.h.med angelegt. Das Dokument wird mit bereits in IS-H/i.s.h.med vorhandenen Stamm- und Verlaufsdaten vorbefüllt. Während der Besprechung des Boards werden Daten in i.s.h.med strukturiert erfasst. Ein Verantwortlicher redigiert das Tumorboarddokument und gibt es frei; es wird danach nicht mehr modifiziert. Daraufhin werden die Daten aus dem Tumorboard-Dokument von i.s.h.med nach MACRO transferiert und dort in den entsprechenden Formularen gespeichert. Ein Patient kann mehrere Tumorboardsitzungen durchlaufen. Dabei wird jeweils ein neues Dokument in i.s.h.med angelegt, wiederum mit zwischenzeitig neu erfassten oder geänderten Daten aus IS-H/i.s.h.med vorbefüllt und nach Freigabe an MACRO übergeben. In MACRO werden die Daten zum Patienten im Verlauf gespeichert, d. h. jede Tumorboard-Sitzung wird auf einen Tumorboard-Visit in MACRO abgebildet. Ein Visit besteht aus einer Gruppe von Formularen zum entsprechenden Besuch des Patienten. Ergänzend zu diesen Daten können in MACRO forschungsrelevante Daten in zusätzlichen Formularen und auch zeitlich feiner-granularen Follow-Up-Visits erfasst werden.
Neben den bereits genannten Systemen sind weitere Komponenten zur Umsetzung der Anforderungen erforderlich. Aus technischer Sicht ist das Gesamtsystem basierend auf einer serviceorientierten Architektur umgesetzt. Um auf die Systeme zur Datenerfassung zugreifen zu können, wurden Wrapper entwickelt. Auf Ebene der Datenarchitektur ist ein kanonisches Datenmodell erforderlich. Sowohl die Schnittstelle zu i.s.h.med als auch die Schnittstelle zu MACRO verarbeiten Dokumente in systemspezifischen XML-Formaten. Das kanonische Datenmodell ist generisch und basiert auf dem W3C RDF-Standard. Die Grundlage bilden Attribut-Wert-Paare die in einer Graphstruktur miteinander in Verbindung gesetzt werden. Aufgrund seiner Generizität eignet es sich für strukturierte, semistrukturierte und unstrukturierte Daten [Wu09].

4 Ergebnis

Das System wurde implementiert und befindet sich derzeit in Praxistests. Alle Dienste sind als Webservices in Java implementiert. Ein zentraler Steuerungsdienst (Central Service) initiiert und steuert den Übertragungsprozess. Er verwendet dabei Dienste zum Zugriff auf die beteiligten Komponentensysteme und einen Protokollierungsdienst.

Der Dienst zum Zugriff auf i.s.h.med (i.s.h.med Connector) ermöglicht einen Zugriff auf im Tumorboard gespeicherte Dokumente sowie Information zur Freigabe. Ein Dokument wird dabei durch ein Tripel aus Patienten-ID, Fallnummer und Dokumenten-ID (PatID, CaseID, DocID) referenziert. Da jede Fallnummer eindeutig einem Patienten zugeordnet werden kann, ist ein Dokument bereits durch das Tupel (CaseID, DocID) eindeutig identifiziert. Auf technischer Ebene sind die Zugriffe über einen entfernten Prozedur-aufruf implementiert. Vor der Rückgabe an transformiert der Dienst das von i.s.h.med zurückgegebene XML-Dokument in das kanonische Datenmodell. Für die Übernahme von Daten in MACRO wurde ein weiterer Dienst implementiert (MACRO Connector). MACRO identifiziert einen Patienten über eine SubjectID und ordnet jedem Subject Visits zu, die über eine VisitID identifiziert werden. Jedem Visit sind mehrere Formulare zugeordnet, welche ebenfalls über eine ID referenziert werden. Für den Datenimport müssen dem Dienst das Dokument im kanonischen Datenmodell, eine i.s.h.med PatID und eine CaseID übergeben werden. Unter Verwendung des Protokollierungsdienstes werden die PatID und die CaseID in eine SubjectID und eine VisitID übersetzt. Anschließend wird das Dokument auf das von MACRO erwartete XML-Format abgebildet und unter Angabe der bekannten Formular-IDs an MACRO übergeben. Der Dienst implementiert den schreibenden Zugriff auf MACRO durch den Aufruf einer von Infermed entwickelten Windows-DLL. Der Protokollierungsdienst (Tracking Service) überwacht die Übertragung von Dokumenten und speichert die Abbildung von IDs persistent. Für jedes bereits übertragene Dokument speichert der Dienst eine Zuordnung von CaseIDs zu DocIDs. Desweiteren protokolliert er die Abbildung von i.s.h.med PatIDs auf MACRO-spezifische SubjectIDs.
Die Applikationsarchitektur ist in Abbildung 1 dargestellt. In regelmäßigen Abständen startet der Central Service den Transferprozess (1). Er ruft dazu eine Schnittstelle des is.h.med Connectors auf (2/3) um zu überprüfen welche Dokumente des Tumorboards freigegeben wurden. Er leitet die Anfrage an i.s.h.med weiter und liefert dem Central Service eine Liste von Tripeln der Form (PatID, CaseID, DocID) zurück.

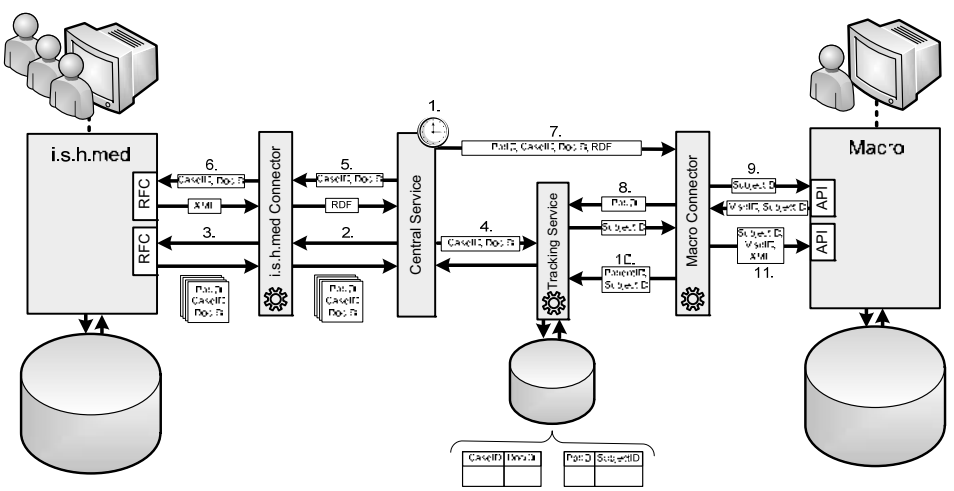


Abbildung 1: Applikationsarchitektur

Für jedes dieser Tripel wird separat der folgende Teilprozess ausgeführt: Zuerst erfolgt ein Zugriff auf den Tracking Service (4), der protokolliert, welche Dokumente für einen Patienten bereits extrahiert wurden. Dazu wird das Tupel (CaseID, DocID) an den Tracking-Service übergeben. Dieser gibt die Information zurück, ob das gegebene Dokument bereits verarbeitet wurde. Falls dies nicht der Fall ist, ruft der Central Service nochmals den i.s.h.med Connector auf und fordert das gewünschte Dokument an (5/6). Der Connector leitet die Anfrage an i.s.h.med weiter, transformiert das erhaltene XML-Dokument in das kanonische Datenmodell und gibt dieses an den Central Service zurück. Dieser leitet das Dokument und das Tripel (PatID, CaseID, DocID) an den MACRO Connector weiter (7). Der MACRO Connector ruft den Tracking Service auf, um den i.s.h.med PatID auf die MACRO-spezifische SubjectID abzubilden (8). Falls der Patient in MACRO bereits vorhanden ist, wird die entsprechende SubjectID zurückgegeben; die leere Menge gibt an, dass der Patient noch nicht angelegt wurde. In diesem Fall legt der MACRO Connector den neuen Patienten an. Desweiteren wird für das gegebene Dokument ein neuer Visit angelegt (9). Als Rückgabe erhält der MACRO Connector nun die aktuell gültige SubjectID und VisitID. Die Abbildung von Patienten-ID auf SubjectID wird zur persistenten Speicherung an den Tracking-Service weitergeleitet (10). In einem letzen Schritt transformiert der MACRO Connector das Dokument in das MACRO-spezifische XML-Format und übergibt dieses zusammen mit SubjectID und VisitID an die MACRO-API (11).

Durch die Replikation der Daten müssen Aspekte der Konsistenz berücksichtigt werden, wobei zwischen Typ- und Instanzebene zu unterscheiden ist. Auf Typebene können betroffene Attribute zugleich in i.s.h.med und in MACRO vorkommen („Schnittmenge“) oder nur in einem der beiden Systeme („Komplement“). Sollen in einem Komponentensystem Änderungen an Attributen aus der Schnittmenge vorgenommen werden, so muss auch das andere Komponentensystem angepasst werden. Eine Vergrößerung der Schnittmenge erfordert eine Anpassung beider Systeme. In beiden Fällen ist eine Abstimmung erforderlich, die sich aber einfach gestaltet, da Änderungen auf Typebene nur von Entwicklern und nicht von Anwendern durchgeführt werden können. Der MACRO Connector muss hierbei angepasst werden. Werden Attribute aus der Schnittmenge gelöscht oder ihre Metainformationen, wie z.B. Wertebereiche, verändert, so erfolgt beim Versuch, diese Attribute aus i.s.h.med zu übernehmen eine Warnmeldung an den Administrator. Das Löschen von Attributen in MACRO führt dazu, dass diese nicht mehr übernommen werden. Das Ändern bzw. Löschen von Attributen aus dem Komplement kann in beiden Systemen unabhängig voneinander durchgeführt werden. Beim Einfügen von neuen Attributen im Komplement muss organisatorisch sichergestellt werden, dass keine partiellen semantischen Überschneidungen mit Attributen aus dem jeweils anderen System vorliegen.

Bei Manipulationen auf Instanzebene lassen sich weitere Fälle identifizieren. Änderungen an Werten von nicht patientenidentifizierenden Attributen in freigegebenen Tumorboarddokumenten in i.s.h.med müssen im klinischen Dokumentationsprozess durch das Anlegen eines neuen Tumorboarddokuments realisiert werden. Dieses wird nach erneuter Freigabe als neuer Visit zum Patienten in MACRO übernommen. Da sich Patienten, für die ein Tumorboarddokument angelegt wird, bereits in interdisziplinärer Behandlung befinden, ist mit einer Änderungen an patientenidentifizierenden Daten in i.s.h.med (Reconciliation) im Regelfall nicht zu rechnen. Im Falle eines Reconciliation-Ereignisses werden entsprechende HL7-Nachrichten von i.s.h.med verschickt. Die darin

enthaltenen Patienteninformationen werden automatisch mit den im Tracking-Service gespeicherten Informationen abgeglichen und ggf. der zuständige MACRO-Administrator benachrichtigt. Das Zusammenführen der betroffenen Patienten in MACRO muss durch den Administrator durchgeführt werden. Freigegebene Tumorboarddokumente können in i.s.h.med nicht gelöscht werden. Das Anlegen neuer Visits in MACRO geschieht bei der Übernahme der Tumorboarddokumente aus i.s.h.med. Es ist nicht vorgesehen, dass Anwender Visits manuell anlegen. In MACRO vorgenommene Änderungen an Daten, die aus i.s.h.med übernommen wurden, müssen ggf. nach i.s.h.med zurück propagiert werden. Dazu kann der MACRO Audit-Trail überwacht werden, um bei Änderungsoperationen an Attributen der Schnittmenge eine Warnmeldung auszulösen. Entsprechende Änderungen in i.s.h.med müssen manuell durchgeführt werden. Das Löschen von Visits durch den Anwender ist nicht möglich.
Im Rahmen der Praxistest wurde die Systemperformanz gemessen. Hierbei ergab sich eine durchschnittliche Gesamtausführungszeiten von weniger als 20 Sekunden für die Übernahme aller freigegebenen Dokumente für einen Patienten. Die Ausführungszeit wird maßgeblich durch den lesenden und schreibenden Zugriff auf die beteiligten Komponentensysteme bestimmt.

5 Diskussion

Die vorgestellte Lösung ist ein Beispiel für die IT-Unterstützung translationaler Forschung durch Informationsintegration. Dazu wurden zwei verbreitete Systeme, das KIS/KAS SAP IS-H/Siemens i.s.h.med und das CDMS InferMed MACRO, basierend auf dem CDISC eSDI „Single Source“-Szenario mit Hilfe des RFD-Konzepts integriert.
Die Übernahme von Daten aus der klinischen Routine in den Forschungskontext ist eine sehr häufige Anforderung. Oft genügt es dabei auf eine separate gemeinsame Eingabeoberfläche zu verzichten. Typisch ist die hier vorgestellte mehrfache Übernahme, die mit der Notwendigkeit zur Konsistenzsicherung verbunden ist. Die rechtliche Grundlage kann variieren. Sehr häufig liegt auch eine explizite Einverständniserklärung vor.
MACRO wird für die Spezialdokumentation verwendet, weil es sich durch den integrierten Formulargenerator und eine einfache Bedienbarkeit gut für die Duplizierung von elektronischen Formularen aus anderen Systemen eignet. Desweiteren verfügt es über Datenmanagementfunktionalität, die für die Qualitätssicherung der Forschungsdaten benötigt wird. Aufgrund der Validierung nach ICH GCP und FDA 21 CFR Part 11 ist bei Verwendung des Eingabepuffers von MACRO auch ein regularienkonformes Datenmanagement möglich.
Der beschriebene Ansatz unterscheidet sich von STARBRITE dadurch, dass zwei sich im Einsatz befindliche Systeme integriert werden und die Lösung in den produktiven Einsatz kommen wird. Auf konzeptioneller Ebene unterscheidet er sich von STARBRITE bzw. CDISC „Single Source“ dadurch, dass auf die Entwicklung eigener Formulare zur Datenerfassung verzichtet wird. Es werden die bereits für die Routinedokumentation eingesetzten Formulare für die Erfassung der Routinedaten verwendet. Die Erweiterung der Formulare um studienspezifische Attribute erfolgt ausschließlich im Forschungssystem. Dies erspart das erneute Anlegen der Dokumente und ermöglicht so eine effizientere Umsetzung des Konzepts. Verglichen mit dem RFD-Konzept der IHE

Initiative werden die Rollen Form Filler, Form Manager, Form Archiver sowie Intermediate Form Receiver durch das klinische Routinesystem repräsentiert.
Die vorgestellte Lösung kann sowohl auf konzeptioneller wie auch auf technischer Ebene für ähnliche Anforderungen wiederverwendet und angepasst werden. Zum Zugriff auf PMDs aus i.s.h.med wurden in Zusammenarbeit mit der Herstellerfirma Schnittstellen entwickelt. InferMed MACRO bringt bereits bei der Auslieferung eine Bibliothek zum externen lesenden und schreibenden Zugriff mit. Durch den Einsatz der Bibliotheken SAP JCO und COM4J können die proprietären Schnittstellen der Komponentensysteme einheitlich als Webservices zugängig gemacht werden.

Literaturverzeichnis

[Al08] Altman, R.B. et al.: Commentaries on "Informatics and Medicine: From Molecules to Populations". Methods Inf Med, 2008;47(4); S. 296-316.

[CDISC] CDISC. http://www.cdisc.org/, letzter Zugriff: 05.05.2010.

[GCP] Good Clinical Practice: Consolidated Guideline. ICH. http://www.ich.org/cache/compo/276-254-1.html, letzter Zugriff 01.05.2010.

[Hi07] Hibbert, M.; Gibbs, P.; O'Brien, T.; Colman, P.; Merriel, R.; Rafael, N.; Georgeff, M.: The Molecular Medicine Informatics Model (MIMM). MEDINFO 2007, K. Kuhn et al. (Eds), IOS Press, 2007.

[HL7] HL7. http://www.hl7.org, letzter Zugriff 01.05.2010.

[Macro] Clinical trial software, remote data entry, clinical decision support, clinical guidelines. InferMed. http://www.infermed.com, letzter Zugriff 02.05.2010.

[ISH] SAP Deutschland - SAP für das Gesundheitswesen – Nutzen. SAP AG. http://www.sap.com/ germany/industries/healthcare/businessbenefits/ index.epx, letzter Zugriff 01.05.2010.

[ISM] i.s.h.med. Siemens Deutschland. http://www.ishmed.de/de/html/index.html, letzter Zugriff 29.04.2010.

[Kus07] Kush, R.; Alschuler, L.; Ruggeri, R.; Cassells, S.; Gupta, N.; Bain, L.; Claise, K.; Shah, M.; Nahm, M.: Implementing Single Source - The STARBRITE Proof-of-Concept Study. J Am Med Inform Assoc. 2007,14(5); S. 662-73.

[Ku07] Kuhn, K. A.; Giuse, D. A.; Lapão, L.; Wurst, S. H. R.: Expanding the Scope of Health Information Systems - From Hospitals to Regional Networks, to National Infrastructures, and Beyond. Methods Inf Med 2007;46; S. 500–502.

[Na06] National Center for Research Resources: Clinical and Translational Science Awards. http://www.ncrr.nih.gov/clinical_research_resources/clinical_and_translational_science_awards/, letzter Zugriff: 15.05.2009.

[Part11] Guidance for Industry. Part 11, Electronic Records; Electronic Signatures. FDA. http://www.fda.gov/downloads/RegulatoryInformation/Guidances/ucm125125.pdf, letzter Zugriff 01.05.2010.

[RFD] IHE Technical Frameworks: Retrieve Form for Data Capture (RFD). http://static.ihe.net/Technical_Framework/upload/IHE_ITI_TF_Supplement_RFD_TI_2007_08_15.pdf, letzter Zugriff 02.05.2010.

[Wu09] Wurst, S. H. R.; Lamla, G.; Prasser, F.; Kemper, A.; Kuhn, K. A.: Einsatz von Dataspaces für die inkrementelle Informationsintegration in der Medizin. GI Jahrestagung, 2009; S. 709-715.

Konzept für ein deutschlandweites Krankheitsnetz am Beispiel von mitoREGISTER

F.M. Kohlmayer [1,2], R.R. Lautenschläger [1], S.H.R Wurst [1], T. Klopstock[3], H. Prokisch[4,5], T. Meitinger[4,5], C. Eckert[2], K.A. Kuhn[1]

[1] Lehrstuhl für medizinische Informatik, Technische Universität München

[2] Lehrstuhl für IT-Sicherheit, Technische Universität München

[3] Friedrich-Baur-Institut, Ludwigs-Maximilian-Universität München

[4] Institut für Humangenetik. Helmholtz Zentrum München

[5] Lehrstuhl für Humangenetik, Technische Universität München

florian.kohlmayer@tum.de

Abstract: Diese Arbeit beschreibt die Architektur eines IT-Systems zur Erfassung und Verwaltung von Patientendaten und Bioproben in einem deutschlandweiten Netzwerk, das im Rahmen des Verbundprojektes mitoNET vom BMBF gefördert wird. Aufgrund der sensiblen Daten liegt ein besonderer Schwerpunkt auf Datenschutz und IT-Sicherheitsaspekten; u.a. werden kryptographische Methoden, Zugriffstokens und getrennte Repositories verwendet. Die Konzepte und das System sind für ähnliche Netzwerke einsetzbar.

1 Einleitung

Nach der Entschlüsselung des menschlichen Genoms hat in der medizinischen Forschung die Verknüpfung genetischer Daten (Genotyp) mit klinischem Erscheinungsbild (Phänotyp) und Verlaufsinformation massiv an Bedeutung gewonnen. Die Analyse von Phänotyp-Genotyp-Beziehungen verbessert das Verständnis von Krankheiten und unterstützt personalisierte diagnostische und therapeutische Ansätze in der Medizin. Detailliert und im Verlauf erfasste medizinische Daten, ausreichende Probanden- bzw. Fallzahlen, sowie annotierte Bioproben sind für diese Analysen entscheidend. Die Erforschung seltener Krankheiten bedarf aufgrund geringer Fallzahlen besonderer Aufmerksamkeit und einer vernetzten Zusammenarbeit, um zu gewährleisten dass möglichst viele Patienten nach definierten Ein- und Ausschlusskriterien erfasst werden. Das hier vorgestellte mitoREGISTER dient der strukturierten longitudinalen Erfassung von Patientendaten sowie zur Bioprobenverwaltung mit dem Ziel einer Erforschung mitochondrialer Erkrankungen.

Ein wesentliches Ziel dieses Projektes liegt darin, unter Wahrung aller datenschutzrechtlichen Vorgaben für die behandelnden und im Projekt mitarbeitenden Zentren (Kliniken und Praxen) ein Register [LG06] aufzubauen, das eine zentrumsübergreifende Erhebung von Patientendaten ermöglicht. Bei jedem Besuch des Patienten werden Bioproben zentral gesammelt und verwaltet. Um die Vergleichbarkeit der Untersuchungen, der Materialen und der Ergebnisse zu gewährleisten, sind standardisierte Terminologien und Standard Operating Procedures (SOPs) erforderlich. Wegen der Langzeiterfassung muss der Fall des Zentrumswechsels eines Patienten berücksichtigt werden. Aufgrund der Komplexität und Heterogenität einerseits (Einbindung existierender Terminologien und Fragebögen, verschiedene Systeme für Datenhaltung und Bioprobenverwaltung, Übernahme von Daten aus Fremdsystemen) und der Schutzwürdigkeit der Daten andererseits (um-

fassend annotierte Bioproben) gibt es derzeit keine allgemein verfügbare oder gar „off-the-shelf“ Lösung, die den Anforderungen genügt.
Diese Arbeit beschreibt ein Sicherheitskonzept und eine Architektur (mit den Schwerpunkten Geschäftsprozess und Anwendung) für eine sichere und datenschutzkonforme Sammlung von Patientendaten und Bioproben. Zwei Aspekte waren bei der Entwicklung wesentlich: die Möglichkeit der Anpassung an die Anforderungen anderer nationaler oder internationaler Forschungsnetze und die einfache Umsetzbarkeit.
In Deutschland werden derzeit 16 verschiedene Forschungsverbünde für seltene Erkrankungen vom BMBF finanziert [PD08]. Diese sind von den Kompetenznetzen in der Medizin zu unterscheiden, in denen weit verbreitete Krankheiten (wie etwa Diabetes mellitus) von ausgewählten Kompetenzzentren untersucht werden. Datenbanken zur Erfassung klinischer Phänotypen verbunden mit Bioprobensammlungen werden in zunehmender Zahl für epidemiologische Fragestellungen, in der Translationsforschung und im Zusammenhang mit klinischen Studien eingesetzt [BBMRI]. Dabei spielt auch die sichere Übernahme freigegebener Daten zwischen Systemen eine wesentliche Rolle [eSDI]. Ein generisches Datenschutzkonzept für medizinische Datensammlungen und Biobanken wurde in [PO07] beschrieben; in dieser Arbeit wird eine Weiterentwicklung vorgestellt.

2 Methoden

Die Datenerhebung erfolgt verlaufsbezogen nach Vorliegen einer Einverständniserklärung; dazu werden standardisierte Formulare und kontrollierte Vokabulare [LG06] verwendet. Zudem werden Bioproben entnommen. Zur Eingabe der medizinisch-phänotypischen Daten in mitoREGISTER werden Fragebögen eingesetzt, die von den Ärzten während oder nach einer Untersuchung ausgefüllt werden. Das System verwendet sowohl neuentwickelte als auch bereits etablierte Fragebögen aus [NMDS], [SARA] und [SF12]. Zur Sammlung von Biomaterialen werden Kits („mitoKITs“) von einem zentralen Labor zur Verfügung gestellt. Sie enthalten mit einem Barcode versehene Probenbehälter, die Einverständniserklärung zur Probenentnahme, eine eidesstattliche Versicherung und einen Dokumentationsbogen. Die Einverständniserklärung eines Patienten dient als elementarer, vertraglicher Bestandteil für die Teilnahme an diesem Forschungsvorhaben. Die eidesstattliche Versicherung dokumentiert, dass die Einverständniserklärung des Patienten korrekt ausgefüllt und archiviert wurde. Sie wird vom Arzt unterschrieben und beim Versand der Proben beigelegt. Die im mitoREGISTER umgesetzten Konzepte orientieren sich an den Vorgaben des BDSG zur Datensicherheit in Forschungsnetzen [BDSG] sowie an [PO07]. Für ein IT-System ergeben sich dabei u.a. die folgenden Anforderungen bzw. Schutzziele [EC09]. Ein Webbrowser soll als Client zum Einsatz kommen, da keine Installation zusätzlicher Software in den beteiligten Kliniken erfolgen darf. Dies gewährleistet Plattformunabhängigkeit, geringen Wartungsaufwand und breite Verfügbarkeit. Die Möglichkeit zur Langzeiterfassung von Patientendaten muss gegeben sein. Eine Anbindung eines zentralen Labors mit Lagerung, Verwaltung und Auswertung der Bioproben ist gefordert. Eine Site-based-View, d.h. nur das Behandlungsteam darf auf die direkt identifizierenden Daten wie Name und Anschrift der eigenen Patienten Zugriff haben. Ebenso ist eine Rechteverwaltung mittels RBAC (Role-based Access Control), mit den Rollen: Administrator/in, Labormitarbeiter/in, Arzt/Ärztin, Forscher/in notwendig. Die Authentifizierung am System soll keine

weitere zu installierende Hardware erfordern (Kartenlesegerät, Fingerprint etc.). Es muss eine Trennung der direkt identifizierenden Daten von den medizinischen und den Analysedaten (Pseudonymisierung) erfolgen. Die Vertraulichkeit und Integrität der Patientendaten und Nachvollziehbarkeit der Änderungen im System (Verbindlichkeit) sind sicher zu stellen.

Das System setzt sich aus verschiedenen Teilsystemen zusammen, die alle mit einem Java-Server-Faces-basierten Framework entwickelt wurden. Als Datenbank kommt MySQL mit InnoDB Engine zum Einsatz, als Web-/Applikationsserver wird Tomcat mit Apache als Frontend eingesetzt. Die einzelnen Teilkomponenten sind jeweils mittels einer Firewall geschützt. Zwecks Ausfallsicherheit werden virtuelle Server verwendet. Backups der virtuellen Server wie auch der Daten werden nur verschlüsselt übertragen und gespeichert. Verwendet wurde hierbei ausschließlich Open-Source-Software, wodurch auch in Zukunft keine Lizenzkosten oder Einschränkungen aufgrund proprietärer Technologie zu erwarten sind. Dies ist wichtig für den Fall, dass die Software anderen Arbeitsgruppen bzw. Netzen zur Verfügung gestellt wird. Durch das Model-View-Controller-Architekturmuster ist das System leicht änder- und erweiterbar. Dadurch ist eine größtmögliche Flexibilität gegenüber sich ändernden Anforderungen vorhanden und Systeme mit ähnlichen Anforderungen können mit geringem Aufwand umgesetzt werden. Das System unterstützt den gesamten Prozess der Patientenaufnahme und Patientenidentifikation, der Datenerhebung, der Probensammlung und der Datenauswertung. Jedem Benutzer werden vom Anwendungsadministrator eine Rolle und ein Zentrum zugeordnet. Wichtig ist eine strikte organisatorische und räumliche Trennung zwischen direkt identifizierenden und medizinischen Daten; sie werden auf unterschiedlichen Teilsystemen jeweils zentral gehalten. Zugriffe zwischen den Systemen werden mittels Einmal-Zugriffs-Tokens geschützt. Die direkt identifizierenden Daten der Patienten sind mittels asymmetrischer Kryptographie geschützt. Ein Audittrail auf Datenbankebene protokolliert alle Änderungen des Systems. HTTPS wird für die Kommunikation genutzt. Erste, automatisierte Penetrationstests wurden durchgeführt. Die Anwendung der o.g. Konzepte resultiert in der in Abbildung 1 dargestellten Architektur.

Das Teilsystem Patientenliste speichert die direkt identifizierenden Daten (IdentDaten) der Patienten und ordnet diesen einen eindeutigen Patientenidentifikator (PatID) zu. Dieser Patientenidentifikator ist eine zufällig generierte alphanumerische Zeichenkette und stellt ein Pseudonym dar. Die Aufgabe der ProbIDListe ist es die auf den Probenbehältern sichtbaren eindeutigen Barcode-Identifikatoren Pseudonyme zweiter Stufe zuzuordnen. Der Barcode-Identifikator wird durch den Arzt in der ProbIDListe registriert (und indirekt einem Patienten über die Med-Datenbank zugeordnet), dabei werden eigene systeminterne Probenidentifikatoren (ProbID für Med-Datenbank, ProbID‘ für Analysedatenbank) zugeordnet. Einem Patienten können mehrere Probenbehälter zugeordnet sein. Diese Probenidentifikatoren sind zufällig generierte alphanumerische Zeichenketten und stellen Pseudonyme für den physischen Probenbehälter, analog zum Patientenidentifikator (PatID), dar. Mit Hilfe der ProbIDListe ist die ProbID und ProbID‘ mit dem als Barcode-Label sichtbaren Identifikator der Probe und den Daten der Med-Datenbank und Analysedatenbank verknüpft. Die Aufbewahrung der Bioproben erfolgt im zentralen Bioprobenlager. Die Verwaltung der Bioproben erfolgt mit dem in Abbildung 1 dargestellten Labor-Informations- und Management- System (LIMS) anhand des Barcode-Identifikators auf den Probenbehältern. Lokal werden Daten der Eidesstattlichen Versicherung und Angaben zum Lagerort der Proben gespeichert. Die Med-Datenbank dient

der Speicherung medizinischer Daten. Die Analysedatenbank dient der Speicherung der Analysedaten. Diese Daten sind durch die Pseudonyme ProbID‘, ProbID und PatID den direkt identifizierenden Daten des Patienten in der Patientenliste zugeordnet.

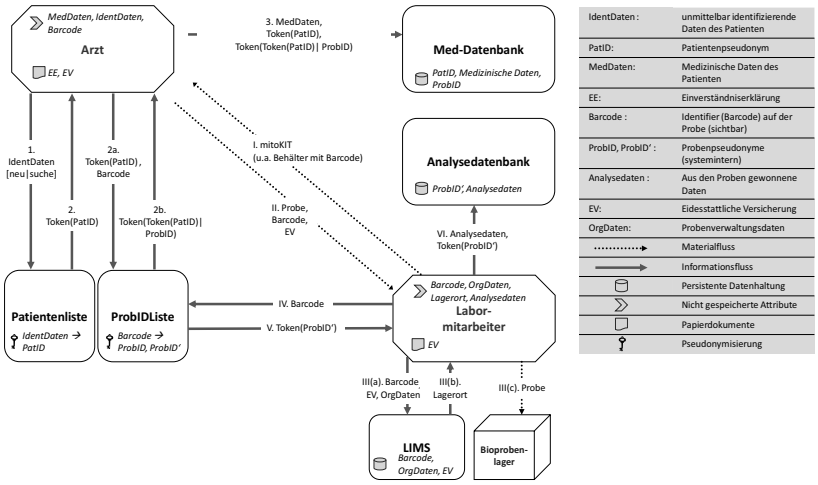


Abbildung 1: Sicherheitsarchitektur

Ein möglicher Zugriff auf das System durch eine(n) Arzt/Ärztin gestaltet sich wie folgt: Nach Authentifizierung des Arztes an der Patientenliste bekommt der Arzt die direkt identifizierenden Daten aller seiner bereits eingegebenen Patienten angezeigt. Der Arzt kann nun einen bereits vorhandenen Patienteneintrag auswählen oder einen neuen Eintrag hinzufügen. Wurden aktuell Bioproben entnommen, kann der Arzt dem Patienten diese mithilfe der ProbIDListe und Med-Datenbank zuordnen. Dies geschieht anhand des auf der Bioprobe stehenden Barcodes (Schritte 1-2b). Zum Einsehen bzw. Editieren medizinischer Daten bekommt der Arzt ein verschlüsseltes Zugriffstoken von der Patientenliste (Schritt 2) mit welchem er automatisch an der Med-Datenbank angemeldet wird (Schritt 3). Der Arzt sieht danach eine Übersicht der in der Med-Datenbank gespeicherten Daten des ausgewählten Patienten, gegliedert nach Behandlungsterminen. Die entnommenen Bioproben werden vom Arzt an das zentrale Labor geschickt (Schritt II). Das Labor kann nun die angeforderten Analysen durchführen. Anschließend werden die Proben eingelagert (Schritte IIIa-c). Die Analysedaten werden in die Analysedatenbank übertragen (Schritte IV-VI). Die einmal gültigen Zugriffstoken für den Zugriff von Patientenliste und ProbIDListe auf Med-Datenbank, sind AES-verschlüsselte und integritätsgeschützte Zeichenketten. Auf den Schlüssel haben jeweils nur die die Patientenliste, ProbIDListe und die Med-Datenbank Zugriff. Der Inhalt des Tokens setzt sich zusammen aus Benutzernamen, Passwort, fortlaufendem Zähler, Patientenpseudonym (PatID) und im Falle einer neu zugeordneten Bioprobe, dem systeminternen Probenidentifikator (ProbID), um die Zuordnung von ProbID und PatID in der Med-Datenbank zu ermöglichen. Der Zähler dient zum Schutz vor Wiedereinspielungsattacken. Durch das verschlüsselte Zugriffstoken ist keine direkte Kommunikation zwischen Patientenliste und Med-Datenbank notwendig.

3 Ergebnisse

Das mitoREGISTER-System wurde am Klinikum rechts der Isar in München entwickelt und befindet sich in einer ersten Ausbaustufe im produktiven Betrieb. Die Patientenliste und ProbIDListe ist organisatorisch wie auch räumlich von der Med-Datenbank getrennt. Die Analysedatenbank und Med-Datenbank sind momentan kombiniert (ProbID und ProbID‘ sind identisch), da nur ausgewählte Analysedaten erhoben werden. Die Patientenliste wird unter Obhut des Datenschutzbeauftragten des Klinikums betrieben, die Med-Datenbank am Institut für medizinische Statistik und Epidemiologie. Im Falle einer Kompromittierung eines der Systeme (Patientenliste, ProbIDListe oder Med-Datenbank) verhindert die pseudonymisierte Speicherung und Trennung der Systeme die Zuordnung von direkt identifizierenden zu medizinischen Daten. Die Lösung mit dem Zugriffstoken ermöglicht eine Integration der Systeme auf Präsentationsebene mithilfe von Standard-HTML-Framesets. Bei der Authentifizierung der Patientenliste an der Med-Datenbank werden in einem weiteren Frame unterhalb der direkt identifizierenden Daten die medizinischen Daten eines Patienten eingeblendet. Hierdurch ist der Zugriff auf die Systeme für den Benutzer transparent. Eine zusätzliche Sicherheitsmaßnahme ist die asymmetrische Verschlüsselung der direkt identifizierenden Patientendaten in der Patientenliste. Jedes Zentrum erhält sein eigenes Schlüsselpaar, und alle Patientendaten eines Zentrums werden mithilfe des öffentlichen Schlüssels eines Zentrums verschlüsselt und können nur durch den privaten Schlüssels des Zentrums entschlüsselt werden (site-based view). Der private Schlüssel eines Zentrums ist mit dem Passwort des Nutzers geschützt und in der Patientenliste gespeichert. Hierdurch wird das Risiko minimiert, dass Patientendaten einsehbar sind, die nicht zum Zentrum des aktuell angemeldeten Benutzers (Arzt/Ärztin) gehören. Der Einsatz asymmetrischer Kryptographie unterstützt auch den Fall eines Zentrumswechslers. In diesem Fall werden die direkt identifizierenden Daten des Patienten mit dem privaten Schlüssel des Zentrums des aktuell angemeldeten Benutzers entschlüsselt und mit dem öffentlichen Schlüssel des neuen Zentrums verschlüsselt. Somit findet eine „Übergabe“ der Daten im System an das neue Zentrum statt.

4 Diskussion und Ausblick

Das präsentierte Konzept stellt eine Erweiterung gegenüber [PO07] dar. Durch das verschlüsselte Zugriffstoken entfällt eine direkte Kommunikation zwischen Patientenliste und Med-Datenbank. Indem der Rechner des Arztes dieses Zugriffstoken weiterleitet, besteht ein logischer, aber vertraulicher und integritätsgeschützter Kanal zwischen Patientenliste und Med-Datenbank, und die Sicherheit wird dadurch gewährleistet, dass der Zugriffstoken-Inhalt auf dem Arztrechner ohne Kenntnis des geheimen Schlüssels nicht ausgelesen werden kann. Eine Veränderung würde durch den Integritätsschutz erkannt und hätte eine Ablehnung durch die Med-Datenbank zur Folge. Durch die site-based view, durch die jeder Nutzer nur Zugriff auf die Patienten seines Zentrums hat, hätte auch eine Kompromittierung eines Benutzerkontos (wie etwa durch Phishing oder Brute-Force) keinen vollständigen Verlust der Vertraulichkeit aller Daten zur Folge. Es wären nur die Daten des betreffenden Zentrums zugreifbar bzw. entschlüsselbar. Auch wird hierdurch eine Wartung durch einen Dritten i.S. des Datenschutzgesetzes (z.B. Systemadministrator) erleichtert, da dieser nur Zugriff auf verschlüsselte Patientendaten hat. Durch die Speicherung personenbezogener Daten in der Patientenliste entsteht ein erhöhter Schutzbedarf. Dem wird u.a. durch asymmetrische Verschlüsselung der direkt identi-

fizierenden Daten der Patienten Rechnung getragen. Wichtig ist, dass eine Ersetzung des Barcode-Identifikators des Probenbehälters im Labor nicht notwendig ist, da die Daten unter einem systeminternen Probenidentifikator (ProbID) in der Med-Datenbank gespeichert werden und die Einverständniserklärung mit den direkt identifizierenden Daten (wie etwa Name, Unterschrift) beim jeweiligen Arzt/Ärztin verbleibt. Damit sind im Labor keine direkt identifizierenden Daten zur Probe bekannt, wodurch ein generelles Problem des Probenversands hinsichtlich des Datenschutzes gelöst wurde. Die Teilsysteme übernehmen Aufgaben wie die Verwaltung von Benutzern, Bioproben, direkt identifizierenden Patientendaten, medizinischen Patientendaten, statistische Erhebung und grafische Aufbereitung. Der Zugriff auf das System bzw. dessen Teilsysteme erfolgt für den Benutzer transparent, da eine vollständige Integration der Präsentationebene realisiert wurde. Dezentrale Labore, analog zu dem hier vorgestellten Zentrallabor, lassen sich einfach in die Architektur integrieren. Auch ein weiterer Pseudonymisierungsschritt zwischen Patientenliste, ProbIDListe und Med-Datenbank ist möglich. Durch die symmetrische Integration von Labordaten und Eingabe der medizinischen Daten können beliebige weitere Datenquellen integriert werden, insbesondere falls eine kontinuierliche Übernahme von freigegebenen klinischen Daten in die Med-Datenbank notwendig wird. Sollten Daten exportiert werden, sind Maßnahmen zur Gewährleistung starker Anonymität (z.B. k-Anonymität) der Daten vorgesehen. Ein großes Sicherheitsrisiko im vorgestellten Konzept bleibt der Rechner des Arztes/Ärztin. Durch die dezentrale und heterogene IT-Landschaft in den beteiligten Zentren kann keine Garantie für die Schadsoftware-Freiheit der Rechner gegeben werden: hier sind die Administratoren der jeweiligen Zentren gefordert. Sicherheitsfortbildungen für Ärzte werden vorgeschlagen, um das Risiko zu minimieren. Auch weitere technische Maßnahmen wie Zwei-Faktor-Authentifizierung der Ärzte sind angedacht. Hierzu sind Machbarkeitsanalysen vorgesehen.

Danksagung Das diesem Bericht zugrundeliegende Vorhaben wurde mit Mitteln des Bundesministeriums für Bildung, und Forschung unter dem Förderkennzeichen 01GM0862 gefördert und wird unterstützt von der Graduate School of Information Science in Health (GSISH) der TU München. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei den Autoren.

5 Literaturverzeichnis

[BBMRI] www.bbmri.eu, letzter Zugriff 30. April 2010.
[BDSG] Bundesdatenschutzgesetz, BDSG 2009.
[EC09] C. Eckert: IT-Sicherheit: Konzepte - Verfahren - Protokolle, Oldenbourg, 2009.
[eSDI] CDISC eSDI Group: Leveraging the CDISC Standards to Facilitate the use of Electronic Source Data within Clinical Trials, 2005.
[LG06] F. Leiner, et al.: Medizinische Dokumentation: Grundlagen einer qualitätsgesicherten integrierten Krankenversorgung. Lehrbuch und Leitfaden, Schattauer, 2006.
[NMDS] Schaefer et al.: Mitochondrial disease in adults: A scale to monitor progression and treatment, Neurology, 2006.
[PD08] K. Pommerening, et al.: Register zu seltenen Krankheiten, Bundesgesundheitsblatt, 51/2008, S. 491-499.
[PO07] K. Pommerening, Das Datenschutzkonzept der TMF für Biomaterialbanken, it - Information Technology, 49/2007, S. 352-359.
[SARA] Schmitz-Hübsch et al.: Scale for the assessment and rating of ataxia: Development of a new clinical scale, Neurology, 2006.
[SF12] www.sf-36.org/tools/sf12.shtml, letzter Zugriff 30. April 2010.

Herausforderungen bei der Extraktion von biochemischen Daten aus der Literatur

Ulrike Wittig, Renate Kania, Isabel Rojas, Wolfgang Müller

Scientific Databases and Visualization Group, Heidelberg Institute for Theoretical Studies (HITS), Schloss-Wolfsbrunnenweg 35, 69118 Heidelberg
Email: Ulrike.Wittig@h-its.org

Abstract: Biochemische Daten in der wissenschaftlichen Literatur liegen in einem nur wenig strukturierten und standardisierten Format vor. Um diese Informationen nutzen und automatisieren zu können, entstand eine Vielzahl von Datenbanken, für die publizierte Daten größtenteils manuell aus der Literatur extrahiert werden, um sie Biowissenschaftlern zur Nutzung zur Verfügung zu stellen. Die Herausforderung bei dem Betrieb solcher Datenbanken besteht unter anderem in der Sicherung der Qualität der Daten. Dies bedeutet, dass viel Zeit von biologischen Experten investiert werden muss, um die Daten aus der Literatur zu extrahieren und für die Eingabe in die Datenbank vorzubereiten, um sie bestehenden Standards anzupassen. Dies erzeugt einen Großteil der zum Betrieb erforderlichen Kosten und beeinflusst damit direkt die Machbarkeit von Projekten. In dieser Publikation beschreiben wir anhand der Datenbank SABIO-RK, welche Probleme von eventuellen automatischen Methoden gelöst werden müssten, um menschliche Arbeitskraft zu ersetzen.

1 Einleitung

Obwohl eine Vielzahl wissenschaftlicher Publikationen in den Lebenswissenschaften heutzutage elektronisch erreichbar ist, ist die Art, in der Informationen publiziert werden, immer noch traditionell innerhalb eines Textes, in Tabellen oder als Abbildungen. Es gibt nur wenige Zeitschriften, die von ihren Autoren die Publikation von Ergebnissen in strukturierter und standardisierter Form verlangen. Um Wissenschaftler bei der Suche nach relevanten Daten für ihre Arbeit zu unterstützen, arbeiten mehrere Gruppen an der Entwicklung von wissenschaftlichen Datenbanken, die Literaturdaten in suchbarer und strukturierter Form enthalten. Allerdings tun sie dies zu recht hohen Kosten, da menschliche Arbeitskraft nötig ist, um die benötigte Information in hoher Qualität herauszufiltern. In dieser Publikation geben wir einen Überblick über einige Probleme, die es zu lösen gilt, um die zur Zeit von Menschen durchgeführten Prozesse wirksam mit Rechnern zu unterstützen. Wir werden am Beispiel der SABIO-RK Datenbank, die in unserer Gruppe entwickelt wird, die Herausforderungen bei der Extraktion von Daten aus der Literatur aufzeigen. SABIO-RK (http://sabio.villa-bosch.de/SABIORK) [1,2], ist eine kuratierte, online zugängliche Datenbank für biochemische Reaktionen und deren detaillierten Informationen zu Reaktionsgeschwindigkeiten. Derzeit wird die Mehrzahl der notwendigen Daten manuell aus der Literatur extrahiert. Neue experimentelle Daten können auch direkt aus dem Experiment heraus eingepflegt werden. Die für die Eingabe in die SABIO-RK Datenbank wichtigen Daten werden dabei aus der wissenschaftlichen

Literatur der letzten 50 Jahre gesammelt. SABIO-RK-Datenbanknutzer sind überwiegend Biowissenschaftler, die diese so genannten enzymkinetischen Daten zur Erstellung von Computermodellen und zur Simulation von biochemischen Reaktionen, Stoffwechselwegen und komplexen biologischen Netzwerken verwenden. Die Enzymkinetik beschreibt dabei die Konzentrationsabhängigkeit der Reaktionsgeschwindigkeit einer biochemischen Reaktion und bestimmt Parameter für ein einzelnes Protein, welches als Enzym, einem so genannten Biokatalysator fungiert. Da Enzyme biochemische Reaktionen beschleunigen und kontrollieren, ist die enzymkinetische Analyse zum Verständnis von Enzymfunktionen unerlässlich. Die relevanten Daten liegen aber in der Literatur in einer Form vor, die die Weiterverwendung (z.B. zur Modellbildung) erschwert. SABIO-RK ist nun ein Dienst, der das Auffinden der wichtigen Parameter und somit die Nachnutzung der Daten erleichtert.

2 Problemstellung

In diesem Abschnitt geben wir einen Überblick über die wichtigsten Probleme, mit denen wir bei der Extraktion von relevanten Daten für SABIO-RK konfrontiert werden. Diese Probleme treten aber nicht nur exklusiv bei SABIO-RK auf, sondern lassen sich auch auf viele andere wissenschaftliche Datenbanken übertragen. Um die Probleme darzustellen, haben wir in Abbildung 1A vier Seiten einer insgesamt 7-seitigen Publikation ausgewählt. Ohne detaillierte Textstellen zu analysieren, soll die Abbildung anhand der Farbmarkierung die räumliche Verteilung verschiedener Daten der gleichen Entität innerhalb einer Publikation aufzeigen. Abbildung 1B zeigt demgegenüber die strukturierte Darstellung eines SABIO-RK Eintrages in der webbasierten Benutzeroberfläche der Datenbank, der die extrahierten Informationen aus der Publikation aus Abbildung 1A enthält, Aus dieser Publikation resultieren insgesamt 6 unterschiedliche Einträge in der SABIO-RK Datenbank.

Räumliche Verteilung in der Publikation: Publikationen experimenteller biologischer Daten folgen in der Regel einem klassischen Schema, bei dem nach einer Einleitung, die das Experiment beschreibt, also Metadaten und Hintergrundinformationen enthält, der Methodenteil mit Angaben über Versuchsabläufe und Messmethoden anschließt. Darauf folgt die Auflistung der eigentlichen Ergebnisse, die dann wiederum in einem vom Ergebnisteil unabhängigen getrennten Abschnitt interpretiert und diskutiert werden. Diese strikte Trennung von Informationen innerhalb einer Publikation bedingt Probleme bei einer möglichen automatischen Zuordnung von experimentellen Ergebnissen. Zusätzlich werden experimentelle Ergebnisse entweder im Fließtext beschrieben und können weit über den Text verstreut vorkommen oder sie werden in Tabellen oder Graphiken dargestellt.

Wenn im Methodenteil z.B. beschrieben wird, dass die Parameter durch die Anwendung einer Gleichung bestimmt wurden, die den kinetischen Mechanismus der ausgewählten Reaktion beschreibt, ist es wichtig, alle zur Gleichung gehörenden Parameter zu finden und einzufügen. In unserem Beispiel enthalten die *grün* markierten Textstellen kinetische Parameter und Informationen über mathematische Gleichungen, mit denen die

Parameter kalkuliert wurden. Diese befinden sich in dieser Beispielpublikation im Text über die gesamte Publikation verteilt, aber auch in einer Tabelle und in einer graphischen Darstellung. Dass, wie in unserem Beispiel, nicht alle Parameter einer Gleichung zusammen in einer Tabelle dargestellt werden, ist leider kein Einzelfall.

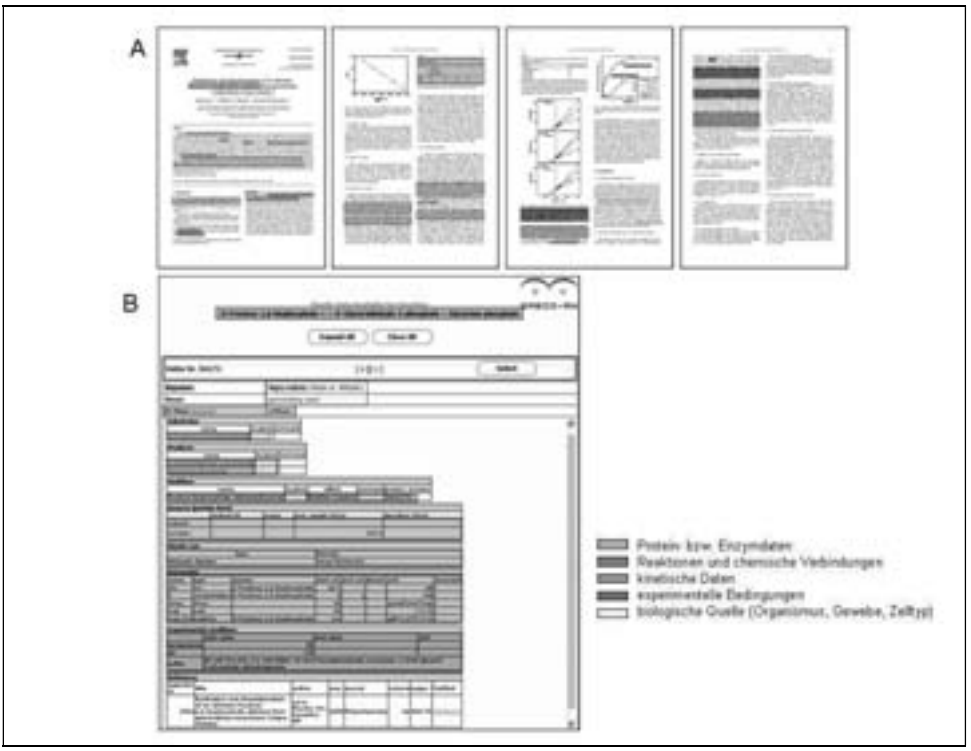


Abbildung 1: Farbmarkierung relevanter Daten in einer Publikation für die Eingabe in SABIO-RK (A) und deren strukturierte Darstellung in einem Eintrag im SABIO-RK Webinterface (B)

Fehlende Daten: Ein Teil der Unvollständigkeit der Daten beruht auf dem Alter der Artikel, z.B. fehlen nähere Angaben zu Proteinen/Enzymen, da zum Zeitpunkt der Publikation nicht bekannt war, dass ein bestimmtes Enzym als Isoform vorliegen kann. Dies bedeutet, ein Organismus besitzt ein oder mehrere Proteine, die der gleichen Enzymklasse angehören und somit die gleiche biochemische Reaktion katalysieren. Diese Proteine unterscheiden sich aber z.B. in der Aminosäurezusammensetzung und besitzen somit andere Molekulargewichte. In ca. 85% der Publikationen gibt es keine Referenzen zu Einträgen von Proteinsequenzen in der Standard-Proteindatenbank UniProt [4]. Diese Datenbank wurde Ende der 1980iger Jahre entwickelt und gilt heute als Standardreferenz für Proteindaten. Aber auch die Mehrzahl der neueren Publikationen enthalten keine Referenzen zu UniProt. Hier sollten die Autoren von

Seiten der wissenschaftlichen Zeitschriften mehr angehalten werden, solch wichtige Referenzen anzugeben.

Biochemische Reaktionen werden durch chemische Verbindungen beschrieben, die als Ausgangssubstrate, Endprodukte und Modifikatoren fungieren. Modifikatoren sind Verbindungen, die die Reaktionsgeschwindigkeit einer Reaktion beeinflussen können, wie z.B. Katalysatoren (Enzyme), Inhibitoren oder Aktivatoren. Für die Eingabe von Reaktionen in SABIO-RK fehlen allerdings in ca. 14% der Publikationen die vollständigen Reaktionsgleichungen. Es werden häufig nur die Substrate genannt, die für die Messungen verwendet wurden und eine Angabe von Reaktionsprodukten fehlt.

Die Wirkungsweise eines Enzyms ist stark abhängig von Temperatur und pH-Wert. Ein großer Vorteil von SABIO-RK gegenüber anderen Enzymdatenquellen ist daher nicht nur die Darstellung der biochemischen Reaktionen und ihrer Reaktionsgeschwindigkeits-relevanten Daten, sondern auch die zusätzlich gespeicherte Information zu den experimentellen Bedingungen wie Temperatur, pH-Wert und Pufferzusammensetzung, unter denen die Daten gemessen wurden. Aber auch hier sind wir mit dem Fehlen oder der Ungenauigkeit von wichtigen Angaben konfrontiert In. 12% der Publikationen gibt es überhaupt keine Angabe zur Temperatur und, bei etwa 6% steht die nicht sehr präzise Angabe Zimmertemperatur (room temperature). Eine mit etwa 10% des Öfteren praktizierte Art der Angabe von experimentellen Bedingungen, ist der Verweis auf eine andere Publikation in der Liste des Literaturverzeichnisses. In diesen Fällen ist eine zeitaufwändige Recherche in den Referenzpublikationen und zum Teil in weiterführenden Referenzen erforderlich. In ca. 20% der Publikationen werden die zu den experimentellen Bedingungen gehörenden Zusammensetzungen des Puffers und die Stoffmengenangaben für Reaktionszusätze nicht in mol/l als Standardeinheit angeben sondern müssen erst manuell auf ein Standardvolumen umgerechnet werden.

Inkonsistenzen: Experimentelle Bedingungen im Methodenteil eines Artikels müssen nicht zwangsläufig übereinstimmen mit den experimentellen Bedingungen in der Legende von Tabellen oder Graphiken, in denen kinetische Parameter angegeben werden. Dies trifft auf etwa 6% der Publikationen zu. Darüber hinaus werden häufig im Methodenteil mehrere verschiedene experimentelle Bedingungen angegeben. Ein Beispiel hierfür ist, dass der Km-Wert unter anderen Bedingungen und mit einer anderen Methode bestimmt wurde als der Kd-Wert. Beide Werte können aber in einer Tabelle angegeben sein. Dann erfordert eine automatische Zuordnung der Methoden die Trennung der Parameter in der Tabelle. Auch innerhalb von SABIO-RK kommt es zur Trennung dieser Parameter und es werden aus diesen Daten unterschiedliche Einträge generiert, da unterschiedliche experimentelle Bedingungen vorliegen.

Vergleiche über Artikel hinweg: Teilweise werden Parameter aus einem ausgewählten Artikel verglichen mit Parametern, die aus der Literatur stammen. Diese werden dann in die Ergebnistabellen mit aufgenommen und nur in der Legende der Tabelle als Referenzwerte ausgewiesen. Hier besteht die Herausforderung darin, die Referenzwerte herauszufiltern und der entsprechenden referenzierten Publikation zuzuordnen.

Die ausgewählten Beispiele zeigen die Komplexität der Problematik und verdeutlichen die Anforderungen, die notwendig sind, um eine sinnvolle manuelle und/oder automatische Informationsextraktion biochemischer Daten aus der Literatur durchführen zu können.

3 Lösungsmöglichkeiten

Die oben genannten Beispiele für Probleme bei der Extraktion von Daten aus der Literatur zeigen, dass derzeit sehr viel manuelle Arbeit von Seiten der Biowissenschaftler bei der Eingabe der Daten in SABIO-RK notwendig ist. Weit verbreitet sind Software-Anwendungen, die die fehlerarme Zuordnung von Namen zu Einträgen in Ontologien erleichtern. Die Möglichkeiten reichen von Auswahllisten zu der linguistischen Analyse von Entitätennamen [5] zu Ansätzen, Nutzer mittels Machine Learning zu unterstützen [6,7].

Für *zukünftige* Publikationen sind in erster Linie die Autoren gefordert, ihre Daten vollständig und standardisiert in der Literatur zu beschreiben. Hierfür sollten die Zeitschriftenverlage idealerweise Formate und Standards vorgeben oder standardisierte Eingabemöglichkeiten anbieten. In Zukunft sind die Autoren der Publikationen gefordert, sich an vorgegebene Standards zu halten. Aber auch die Gutachter und Herausgeber der Zeitschriften sollten darauf achten, dass alle notwendigen Informationen angegeben wurden. Ideal wäre eine enge Zusammenarbeit von Datenbanken und Herausgebern um z.B. eine Eingabemaske bereitzustellen und die Autoren anzuhalten, ihre Daten an die entsprechende Datenbank weiterzuleiten. Ein Schritt in diese Richtung wird von der STRENDA Initiative [8] unternommen, die eine Liste von Anforderungen mit Vorschlägen für die Publikation von Enzymdaten erstellt hat. Diese Liste enthält ein Minimum an Informationen, das notwendig ist, um ein Enzym und seine relevanten Daten vollständig zu charakterisieren. Einige wissenschaftliche Zeitschriften haben bereits zugestimmt, ihren Autoren die Anforderungsliste der STRENDA Initiative als Grundlage für das Publizieren von Enzymdaten nahe zu legen. Auch die Gründung der International Society for Biocuration im Jahre 2009 [9] zeigt die Wichtigkeit der Sicherung der Datenqualität und die Notwendigkeit der Definition von Standards durch die stetig wachsende Anzahl von Daten in den Biowissenschaften aufgrund der Anwendung immer modernerer Meßmethoden [10].

Aber auch wenn mehr und mehr Daten neu sind, bleibt immer noch die Erschließung bereits erfolgter Messungen für die Speicherung in Datenbanken, also *Legacy-Publikationen.* Bei den im Abschnitt Problemstellung genannten Beispielen ergibt sich die Frage, inwieweit es möglich ist, die Extraktion der Informationen aus der Literatur durch Nutzung von Softwareanwendungen zu automatisieren. Für einen Großteil der fehlenden und ungenauen Angaben in der Literatur gibt es kaum eine Möglichkeit, die zusätzlichen Informationen automatisch zu generieren.

Dagegen wäre bei Problemen der Zuordnung von Daten und von über den gesamten Text der Publikation verteilten Informationen eine automatische Lösung zumindest ein

Ziel. Am Beispiel der Extraktion von kinetischen Daten aus der Literatur besteht eine Herausforderung darin, die richtigen experimentellen Bedingungen den kinetischen Parametern zuzuordnen, fehlende Informationen aus der Referenzliteratur zu extrahieren oder Zusammenhänge zwischen Parametern herzustellen. Letzteres bezieht sich z.B. auf die Analyse, welche Parameter unter identischen experimentellen Bedingungen bestimmt wurden und somit gemeinsam in einer mathematischen Gleichung repräsentiert werden können. Ein Beispiel hierfür ist die Suche nach experimentellen Ergebnissen, die z.B. als Parameterwerte Km und Vmax angegeben werden, die beide gemeinsam in der mathematischen Gleichung für die Michaelis-Menten-Kinetik dargestellt werden können. In über 90%der Publikationen werden die Ergebnisse im Text nicht zusammen mit der mathematischen Gleichung und den experimentellen Bedingungen, unter denen sie ermittelt wurden, angegeben. Stand der Technik ist die automatische Markierung von relevanten Informationen in einer Publikation. Aus unserer Sicht muss diese mit Methoden komplementiert werden, wenige mögliche Zuordnungen von Daten unterschiedlicher Entitäten vorzuschlagen. Diese Kombination von Markierungen und möglichen Zuordnungen würde die Wissenschaftler bei der Suche und Zuordnung von Daten unterstützen und die Extraktion von Daten erheblich erleichtern.

Danksagung

Das SABIO-RK Projekt wird durch die Klaus-Tschira-Stiftung und das Bundesministerium für Bildung und Forschung (SysMO/ Virtuelle Leber) finanziert. Die Autoren danken Henriette Slogsnat für anregende Diskussionen.

Literaturverzeichnis

[1] Wittig U *et al.* (2006) SABIO-RK: integration and curation of reaction kinetics data. Lecture Notes in Computer Science, 4075, 94-103

[2] Rojas I *et al.* (2007).Storing and annotating of kinetic data. In Silico Biol., 7 (Suppl 2) (17822389):37-44

[3] Hucka M *et al.* (2003) The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. Bioinformatics. 19, 524-31

[4] The UniProt Consortium. (2008) The Universal Protein Resource (UniProt)**.** Nucleic Acids Res. 36:D190-D195

[5] Engelken H *et al.* (2009) Flache und semantische Verarbeitung von Namen biochemischer Verbindungen. In INFORMATIK 2009 - Im Focus das Leben, Beiträge der 39. Jahrestagung der Gesellschaft für Informatik e.V. (GI), Lübeck, Germany, LNI

[6] Van Auken K *et al*. (2009) Semi-automated curation of protein subcellular localization: a text mining-based approach to Gene Ontology (GO) Cellular Component curation. BMC Bioinformatics. 10:228

[7] Doms A, Schroeder M (2005) GoPubMed: exploring PubMed with the Gene Ontology. Nucleic Acids Res. 33:W783-6

[8] STRENDA: http://strenda.org/

[9] International Society for Biocuration: http://biocurator.org/

[10] Howe D *et al.* (2008) Big data: The future of biocuration, Nature. 455(7209):47-50.

Ranked Set Search in Medline Documents

Kristin Mittag[1], Alexander Hinneburg[2]

[1]OntoChem GmbH, Halle
[2]Informatik, Martin-Luther-University, Halle-Wittenberg, 06099 Halle
{mittag,hinneburg}@informatik.uni-halle.de

Abstract: Information needs like searching scientific literature that involve high recall rates are difficult to satisfy with ad hoc keyword search. We propose to state queries implicitly by the specification of a set of query documents. The result of such a query is a set of answer documents that are ranked within the answer set. We describe efficient techniques to process such queries. Preliminary experiments using data from the TREC Genomics track 2005 are reported.

1 Introduction

Searching literature in life sciences for relevant publications is a difficult task. The diversity of the scientific vocabulary and ambiguous naming conventions for entities make boolean keyword search less effective in retrieving all papers relevant to a given information need. Controlled vocabularies like MESH-terms and meta-information mitigate the problem. However, constructing a boolean query to retrieve for example all randomized clinical trials on a particular subject that are present in Medline still requires a lot of manual tuning by search experts. Thus, potential knowledge hidden in publication databases will be difficult to access for life science researchers in case search experts for the particular problem at hand are not available.

A typical situation in searching scientific literature is that some papers on a particular subject are easy to find. Those papers may serve as starting point for further searching. However, it is difficult to find all relevant publications on the subject. One strategy is to use the known relevant papers and search for similar papers for each of them, maybe by using the PubMed's related article feature [LW07]. It precalculates the probability that a user wants to see a document given interest in another through a probabilistic topic model based on Poisson distributions over term frequencies. This strategy is used by [LA98] where they use the related article feature to update a bibliography. But it still requires much manual effort – they run their algorithm several iterations and in every round the bibliography gets updated by user-selected relevant articles.

Searching a document collection based on a given set of relevant papers had been formulated as a two-class document classification problem[SA05, FBSS$^+$09, GvdL05]. The papers that are found in the beginning are used as examples to train a classifier. Random documents from the document collection pose as background class. The trained classifier is then applied to all documents in the collection to distinguish relevant documents

from the background, thus, it is a computationally expensive operation. Such a classifier may be a set of discriminating words, derived by comparing probability distributions of training set and background collection. The problem is that many training documents (several thousands) and a well defined background are required. Thus, this technique is best applicable to large document classes and can for example be used to support document annotation with MESH-terms.

We propose a new search technique that does not require manual query construction and it needs a relatively small set of query documents (5 to 50). In addition, it works on top of standard information retrieval technology thus query processing is very efficient. Our new approach requires the user to specify the search query as a few query documents relevant to the information need at hand. Literally spoken, such type of query could be interpreted as the task: Read the given documents, find the subset of the document collection containing all related documents and rank them by relevance. Stating a query for a complex information need by picking a few example documents is convenient and relieves the user of difficult tasks like iteratively constructing lengthy boolean keyword queries. Instead, the information need is given implicitly by the combination of selected query documents. The result of the search is not only a ranking of all documents in the collection like in standard information retrieval systems but our method also decides how to limit the answer to a reasonable subset of documents. Thus, it delivers a ranked subset of documents as an answer. We call this problem *Ranked Set Search (RSS)*.

Our contributions are that we first propose a formal definition of the new RSS problem. Second, we develop efficient techniques to compute a solution. Finally, we report preliminary experiments on real data from the TREC Genomics evaluation track 2005 that demonstrate superior effectiveness of our approach over several baselines. In the remainder of the paper, first, we develop the formal problem definition as well as efficient search techniques in Section 2. In Section 3, we present preliminary experiments and discuss the result.

2 Ranked Document Set Retrieval

In the sequel, we describe the new framework of ranked set search. The given data consists of a collection of documents $D = \{d_1, \ldots, d_N\}$. A query is a set of query documents $D_q = \{q_1, \ldots, q_R\}$ that may or may not be part of the document collection D. The output is a set of documents $D_a \subset D$ that is a subset of the document collection. The individual documents in D_a are ordered according to a distance function, called **document distance** $dist_{doc}$, that determines the ranking of the documents within the answer set. The document distance function defines a distance between a document $d \in D$ and the query D_q. Thus, the document distance also induces a ranking of documents over the entire collection D.

The answer set D_a is usually a relatively small subset of D. The answer set is chosen among the possible subsets of D such that a second distance function, called **set distance** $dist_{set}$, becomes minimal. In order to avoid to search among all possible subsets of D, which might be very time consuming in case of general set distances, we restrict the search for an answer set with minimal set distance to the top-k documents of the ranking induced by document distance. The definition of ranked set search summarizes our concept.

Definition 1 *Given a collection $D = \{d_1, \ldots, d_N\}$ and a set of query documents $D_q =$*

$\{q_1, \ldots, q_R\}$, ***Ranked Set Search (RSS)*** *determines an answer set* $D_a \subset D$ *such that*

$$dist_{set}(D_q, D_a) = \min_{1 \leq k \leq k_{\max}} \left\{ dist_{set}\big(D_q, top_k(D, D_q, dist_{doc})\big) \right\}$$

with $dist_{set}(\cdot, \cdot)$ *is a distance function between sets of documents,* $dist_{doc}(d, D_q)$ *is a distance function between a document* $d \in D$ *and the set of query documents and* $top_k(D, D_q, dist_{doc})$ *delivers the subset of the top-*k *documents of* D *with respect to the ranking induced by* $dist_{doc}(d, D_q)$.

Ranked set search does not require the specification of a parameter like k that limit a ranking to the subset of the top-k documents. Instead, the size of the answer set is controlled by the set distance function. How to choose a meaningful set distance? Computing set distance as the minimum, maximum or the average of the document distances of the top-k documents is not a promising way. The set distance would be either constant (minimum) or monotonically increasing with k (maximum, average). Thus, the minimization of such set distances would not yield interesting solutions.

A class of set distance functions, which does not lead to trivial solutions, is to compare the word probability distributions of the query set D_q and the top-k subsets of the ranking of D. A word probability distribution in the simplest form is a multinomial distribution over the vocabulary W, which assigns a probability to every word $w \in W$ and the probabilities of all words sum up to one.

Kullback-Leibler-Divergence between two probability distributions x and y over the vocabulary is a measure with sound information-theoretic basis:

$$KL(x||y) = -\sum_{w \in W} x(w) \ln y(w) - x(w) \ln x(w)$$

KL-divergence has a nice information-theoretic interpretation. It measures the additional average amount of bits required to code a document following a word distribution x with a coding scheme designed for documents following a distribution y instead of x. We denote the word probability distribution of the query document set D_q by P_q and the distribution of the answer set D_a by P_a. KL-divergence is zero, when both distributions are identical, otherwise it is larger zero. All word distributions are estimated as multinomial distributions. To compensate for the word burstiness phenomenon[1], we use idf-transformation of the multinomials [MKE05].

However, KL-divergence is not symmetric, i.e. $KL(P_q||P_a) \neq KL(P_a||P_q)$. What version should be used? To understand the impacts of both versions, we look at the active vocabularies W_q and W_a that belong to D_q and D_a respectively. $KL(P_q||P_a)$ is small when all words of W_q are also included in W_a. Words of W_q that are missing in W_a create large overhead, because they are considered very infrequent according to distribution P_a. Vice versa, words that are in W_a but not in W_q do not increase $KL(P_q||P_a)$ by much. Thus, minimizing $KL(P_q||P_a)$ favors answer sets that obey $W_q \subseteq W_a$. The opposite version $KL(P_a||P_q)$ does not care much about words of W_q that are missing in W_a. It is much more sensitive to new words that do not appear in W_q but in W_a and tries to avoid such new words. Thus, minimizing $KL(P_a||P_q)$ favors answer sets that obey $W_q \supseteq W_a$.

[1] Burstiness says that a word that occurs once in a document has higher probability to appear again [Kat96].

method	greedy	TF-IDF	KL 1	KL 2	KL 3	KL 4
MAP	0.92	0.45	0.39	0.56	0.58	0.59
R-Precision	0.81	0.44	0.49	0.59	0.60	0.60
size	77	24735	24	96	269	678

Table 1: Mean Average Precision (MAP), R-Precision and size of result-set for different methods using 100 samples in all experiments.

We make use of both versions. First, we minimize the set distance $KL(P_q||P_a)$. Thus, W_a contains as many as possible words of W_q and a few additional related words. We use the answer set as *pseudo relevance feedback* and expand the query with it $D_{q'} = D_q \cup D_a$ to generalize the vocabulary of the original query D_q. In a second step, we minimize the opposite version, the set distance $KL(P_a||P_{q'})$ with respect to the expanded query $D_{q'}$. Using the answer set of the second step as well as pseudo relevance feedback, whole two-step procedure can be repeated. In our experiments, we compare KL-based RSS methods that run the two steps one and two rounds respectively.

Finally, we discuss the document-distance that compares a single document with the whole set of query documents. A document $d \in D$ is close to the query set D_q, when d contains most of the words in W_q. Therefore, we use the first version of the KL-divergence $KL(P_q||P_d)$ as document distance that defines the ranking. P_d is the word distribution of the single document d. Because, standard information retrieval systems do not support that kind of distance by an index structure, we use sampling and standard TF-IDF rankings to approximate the KL-based ranking. For the approximation, we draw several samples from the word distribution P_q. Each sample consists of a set of randomly drawn words and is used as keyword query that is submitted to a standard retrieval system. The system delivers a TF-IDF ranking of documents for each keyword sample. The final KL-based ranking is computed by merging the TF-IDF rankings with respect to the document distance $KL(P_q||P_d)$. An example is given at the end of Section 3. Notice that the document distance changes after pseudo relevance because D_q does change. In our preliminary experiments, we did not recompute the keyword samples and the TF-IDF rankings but just continued the merging with the changed document distance.

3 Experiments

We evaluate KL-based RSS on the TREC Genomics collection (ad hoc retrieval task 2004+2005) [HCY+05]. That is a 10-year subset of Medline (1994-2003) with about 4.5 million documents. In preliminary experiments we used a single topic (ID 131) out of 50 topics, from TREC Genomics 2005 ad hoc retrieval task. It contains 42 documents, which are judged relevant by human evaluators. For each document, abstract and title were preprocessed using *Apache Lucene* (`http://lucene.apache.org`). In the preprocessing, we used stemming and filtered stopwords as well as infrequent words.

We evaluate our method against a baseline search method that uses TF-IDF rankings. The method merges the TF-IDF rankings of the sampled keyword queries using the TF-IDF score of the retrieved documents. For efficiency reasons we restricted the TF-IDF ranking

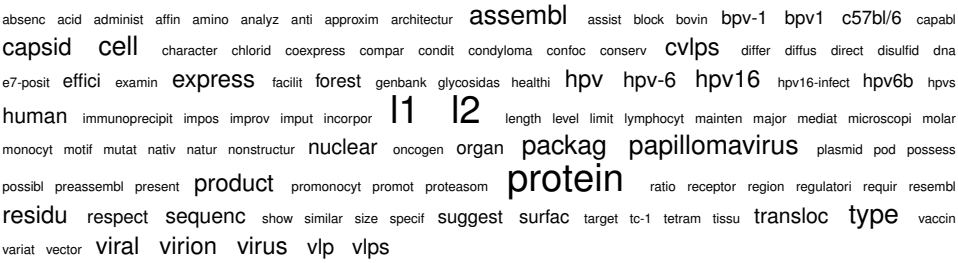


Figure 1: Keyword cloud for our example run.

of each query to 1000 documents. When a document occurs in more than one TF-IDF ranking, the highest score is used. The result is a single ranking of all documents from the TF-IDF rankings.

We also developed an algorithm to compute an upper bound of the search performance to check the open potential of our method. This algorithm needs to know the relevant documents of a RSS query beforehand, thus, it can be run only on the evaluation data set. The problem is to cover all relevant documents in the TF-IDF rankings of the sample keyword queries by prefixes of minimal lengths. We map that problem to the set cover problem, a standard NP-complete problem, and compute an approximate solution by an greedy algorithm.

In each test run, we randomly selected 10 documents as query set from our test topic 131 and took the remaining 32 documents as relevant documents. Performance is measured by *Mean Average Precision* (MAP) and *R-Precision*, which are standard IR performance measures and produce consistent results [BV00]. MAP averages the precisions at each relevant document in the result set and takes the mean of the average precisions over all queries. R-Precision takes the precision at position $R = number\ of\ relevant\ documents$ (32 in our case) that is the first position where recall 1 is possible. All results are averages over 50 repetitions.

Table 1 shows results of the baselines and two KL-based RSS-algorithms. The KL-based RSS-algorithm that minimize the two versions of KL-Divergence one time (KL 1) has comparable R-Precision as the baseline (TF-IDF) but small MAP. The KL-based RSS-algorithm that repeat the minimization two times (KL 2), however, achieves a significant improvement over the baseline in both R-Precision as well as MAP. Additional repetitions (KL 3, KL 4) only marginally improve MAP and R-Precision. On the other hand, the result of the upper bound (optimal) shows that there is still room for improvement based on new techniques.

The task of topic 131 is to find documents about genes "*L1* and *L2* in the *HPV11* virus" and the "role of L2 in the viral capsid". HPV is an abbreviation for *human papillomavirus* and L1 and L2 are the names of two proteins of the viral capsid and their respective genes. Our method to specify queries by documents instead of keywords also helps to automatically identify selective keywords. The keyword cloud (Figure 1) illustrates the frequency of keywords in the keyword query samples that contribute to the final result. Words with bigger font size are found more often in those samples and could be relevant for the topic. For example *papillomavirus* is part of the acronym HPV11. HPV6 and HPV16 are dif-

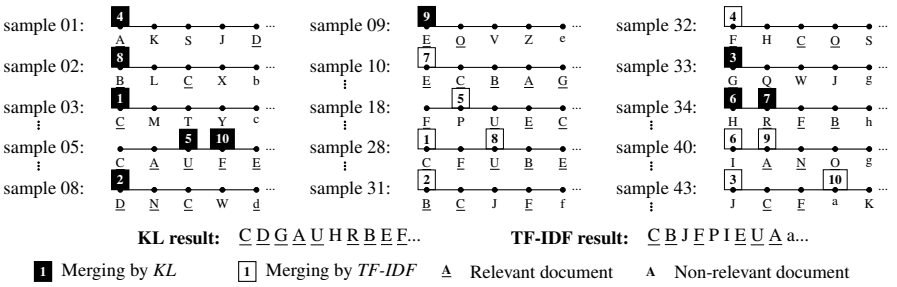


Figure 2: Top-5 documents of the TF-IDF rankings of the keyword query samples are shown that contribute to the final results of *KL* and *TF-IDF*. Documents are identified by letters.

ferent types of HPV, as well as HPV11. The two proteins L1 and L2 are expressed in the infected cell, in which new viruses are assembled.

Figure 2 illustrates the merging of the TF-IDF rankings for the first 10 documents of the final results of KL-based RSS and the baseline TF-IDF on a real example. Different rankings contribute to the results of KL-based RSS and TF-IDF. KL-based RSS has only a single non-relevant document among the top 10, while TF-IDF selected four non-relevant documents.

We conclude that stating a query implicitly by a set of query documents helps to take variations of terms as well as related terms into account. Our preliminary experiments show that RSS is a promising approach to searching life science literature.

References

[BV00] C.Buckley and E. M. Voorhees. Evaluating evaluation measure stability. In *SIGIR '00*, pp. 33–40, 2000.

[FBSS+09] J.-F. Fontaine, A. Barbosa-Silva, M. Schaefer, M. R. Huska, E. M. Muro, and M. A. Andrade-Navarro. MedlineRanker: flexible ranking of biomedical literature. *Nucl. Acids Res.*, 37(suppl_2):W141–146, 2009.

[GvdL05] T. Goetz and C. W. von der Lieth. PubFinder: a tool for improving retrieval rate of relevant PubMed abstracts. *Nucleic Acids Res*, 33(Web Server issue), 2005.

[HCY+05] W. Hersh, A. Cohen, J. Yang, R.T. Bhupatiraju, P. Roberts, and M. Hearst. TREC 2005 Genomics Track Overview. In *TREC 2005*, 2005.

[Kat96] S. M. Katz. Distribution of content words and phrases in text and language modelling. *Nat. Lang. Eng.*, 2(1):15–59, 1996.

[LA98] X. Liu and R. B. Altman. Updating a Bibliography Using the RELATED ARTICLES function within PubMed. In *Proc. AMIA Symp*, pp. 750–754, 1998.

[LW07] J. Lin and J. W. Wilbur. PubMed related articles: a probabilistic topic-based model for content similarity. *BMC Bioinformatics*, 8(1), 2007.

[MKE05] R. E. Madsen, D. Kauchak, and C. Elkan. Modeling word burstiness using the Dirichlet distribution. In *ICML '05*, pp. 545–552, 2005.

[SA05] B. P. Suomela and M. A. Andrade. Ranking the whole MEDLINE database according to a large training set using text indexing. *BMC Bioinformatics*, 6(1), 2005.

Workshop

5th German Workshop SDF 2010 - Sensor Data Fusion: Trends, Solutions, Applications

Wolfgang Koch

Multi-Target Tracking using Signal Strength Measurements

Michael Mertens and Martin Ulmke
Dept. Sensor Data and Information Fusion
Fraunhofer FKIE
Wachtberg, Germany
michael.mertens[martin.ulmke]@fkie.fraunhofer.de

Abstract: Many surveillance tasks rely on the generation of stable, continuous tracks of objects of interest. Often, track continuity has a higher priority than track accuracy, as valuable information on the identity or the origin of an object is lost by each track drop. The main source of track fragmentation are missing detections due to a limited field of view or technical or topographical masking. In the multi-target case, in addition, individual tracks can be interchanged by unknown data assignments. Therefore, the exploitation of additional sensor data is required in order to discriminate individual objects and to associate track fragments. As signal strength measurements are standard output of modern radar systems, the integration of this information into a Bayesian tracking scheme is discussed in the present paper. In contrast to previous approaches, the knowledge on the target's signal strength is not only used for an improved calculation of the association probabilities, but it enters into the algorithm as a random variable which is estimated sequentially. By this approach it is not only possible to discriminate closely-spaced targets and improve the track continuity, but also to support possible classification and identification tasks. The signal strength fluctuations of the target returns are modeled by the Swerling-I and Swerling-III cases. As a first performance evaluation, numerical results are presented based on a two-target simulation scenario.

1 Introduction

In ground surveillance with airborne Ground Moving Target Indication (GMTI) radar, the main task of establishing and maintaining tracks of relevant moving objects is challenged not only by imprecise, uncertain, and ambiguous measurements. To a large degree, the main difficulty arises from complex target dynamics, e.g. stop & go behavior or strong maneuvers, masking due to the sensors Doppler blind zone, nontrivial topography causing terrain obscuration, situations with closely-spaced targets, a strong false alarm background, etc. In general, these factors quickly lead to a strong performance degradation or even track loss. To counterbalance these factors, it is beneficial to incorporate additional sources of information into the tracking process.
Being a standard output of a modern radar system, the target amplitude or the corresponding target signal strength has already been used in the past to facilitate the problem of associating a track with its correct measurement. But in contrast to previous approaches, e.g. [LBS90, vK96, DCV10], in this work the knowledge on the target's signal strength is not only used for an improved calculation of the association probabilities, but it enters into

the algorithm as a random variable which is estimated sequentially. With this approach it is not only possible to discriminate closely-spaced targets and improve the track continuity, but also to support possible subsequent classification and identification tasks. In addition, for phased array antennas the signal strength estimates could also be used to contribute to the radar resource management: depending on the estimated target SNR, the detection threshold could be adapted to allow better target detections.

2 Signal strength model

Due to its complexity, a realistic modeling of a real target's back-scattering characteristics is in general impossible. Therefore statistical models are used instead which can be handled analytically. In this work it is assumed that the fluctuations of the input signal at the detector, resulting from fluctuations of the target cross section, can be described by the Swerling models [Swe60].

The complex target signal $\mathbf{v} = (v_1, v_2)$ with orthogonal and statistically independent components is added by white Gaussian noise within the receiver unit. The detector uses the total signal $\mathbf{u} = (u_1, u_2)$ to form the signal strength $||\mathbf{u}||^2 = (u_1)^2 + (u_2)^2$ with the probability density given by the Rice distribution. In connection with the above mentioned fluctuation model and a detection threshold λ, this leads to explicit equations for the detection probability $\mathbf{P}_D$ and the detected signal probability density $p(s)$ as functions of target strength and λ. The functions $p(s)$ are given in terms of Gamma functions (see [Koc06]) and will play the role of the likelihood function in the Bayesian update below.

3 Incorporation of signal strength information

In Bayesian target tracking, e.g. [BSL93, BSLK01], the probability density $p(\mathbf{x}_k|\mathcal{Z}^k)$, which describes the target state $\mathbf{x}_k$ at time step t_k, conditioned on the measurement sequence $\mathcal{Z}^k = \{\mathbf{z}_1, \mathbf{z}_2, ..., \mathbf{z}_k\}$ with $\mathbf{z}_k = \{z_k^i\}_{i=1}^{n_k}$, is sequentially updated by

$$p(\mathbf{x}_k|\mathcal{Z}^k) = \frac{p(\mathbf{z}_k|\mathbf{x}_k)\, p(\mathbf{x}_k|\mathcal{Z}^{k-1})}{\int d\mathbf{x}_k\; p(\mathbf{z}_k|\mathbf{x}_k)\, p(\mathbf{x}_k|\mathcal{Z}^{k-1})} \tag{1}$$

To incorporate signal strength, the kinematic target state $\mathbf{x}_k$ is augmented with a target strength variable $\mathbf{s}_k$ and, thus, becomes $\mathbf{X}_k = (\mathbf{x}_k, \mathbf{s}_k)$. The signal strength measurements κ_k^i, on the other hand, are introduced into the measurement set: $\mathbf{Z}_k = \{z_k^i, \kappa_k^i\}_{i=1}^{n_k}$. We assume that the target density factorizes in $\mathbf{x}_k$ and $\mathbf{s}_k$. As usual in the tracking literature, the density on $\mathbf{x}_k$ is described by normal distributions, predicted and updated, for given target–detection assignments, with the Kalman filter. Because of the structure of the likelihood function, the density on $\mathbf{s}_k$ is well described by an inverse Gamma function

which keep its structure under successive application of the Bayes update:

$$p(\mathbf{s}_k|\mathcal{Z}^{k-1}) \quad = \quad \mathcal{I}(\mathbf{s}_k;\hat{\mathbf{s}}_{k-1},\mu_{k-1}) \;=\; \mathrm{N}_{\mu_{k-1}}\mathbf{s}_k^{-\mu_{k-1}-1}e^{-\frac{(\mu_{k-1}-1)\hat{\mathbf{s}}_{k-1}}{\mathbf{s}_k}} \tag{2}$$

with a normalization constant $\mathrm{N}_{\mu_{k-1}}$. The probability density $\mathcal{I}(\mathbf{s};\hat{\mathbf{s}},\mu)$ has the expectation value $\mathrm{E}[\mathbf{s}] = \hat{\mathbf{s}}$. If the parameter $\mu > 2$, then the variance exists with $\mathrm{Var}[\mathbf{s}] = \frac{\mathbf{s}^2}{\mu-2}$. Because of $\frac{d}{dt}\mathrm{SNR}_0 = 0$, no "dynamics" is needed for the signal strength, thus the prior density at t_k is identical to the posterior density at t_{k-1}.

3.1 Combined likelihood

The combined likelihood function $p(\mathbf{Z}_k|\mathbf{X}_k) = p(\mathbf{Z}_k|\mathbf{x}_k,\mathbf{s}_k)$ is the probability density of the measurements and comprises all possibilities how the given sensor output $\mathbf{Z}_k$ can be interpreted, given the true target state $\mathbf{X}_k$. Assuming independent, identically distributed false alarm measurements with the number of false alarms determined by the Poisson distribution, the likelihood function, up to a factor which does not depend on $\mathbf{X}_k$, can be written as

$$p(\mathbf{Z}_k|\mathbf{X}_k) \quad \propto \quad (1-\mathbf{P}_D(\mathbf{s}_k))\rho_F + \mathbf{P}_D(\mathbf{s}_k)\sum_{i=1}^{n_k}\mathcal{N}(z_k^i;\mathbf{H}_k\mathbf{x}_k,\mathbf{R}_k)\,\mathcal{LR}_S^i \tag{3}$$

where ρ_F is the false alarm density and $\mathcal{N}(z_k^i;\mathbf{H}_k\mathbf{x}_k,\mathbf{R}_k)$ is the normally distributed single measurement likelihood, which results from a linear measurement model with additive white Gaussian noise: $z_k^i = \mathbf{H}_k\mathbf{x}_k + v_k$, $\mathbf{v}_k \propto \mathcal{N}(\mathbf{v}_k;0,\mathbf{R}_k)$ with measurement matrix $\mathbf{H}_k$ and measurement covariance $\mathbf{R}_k$. The signal strength likelihood ratio $\mathcal{LR}_S^i$ in (3) is given by the above mentioned signal densities $p(s)$.

3.2 Filter update step

Assuming strong targets, i.e. $1+\mathbf{s}_k \approx 2+\mathbf{s}_k \approx \mathbf{s}_k$, the posterior density $p(\mathbf{X}_k|\mathcal{Z}^k) = p(\mathbf{x}_k,\mathbf{s}_k|\mathcal{Z}^k)$ can be written as the sum of the product of a Gaussian with an inverse Gamma density, describing the target's kinematic state and signal strength, respectively:

$$p(\mathbf{X}_k|\mathcal{Z}^k) = \sum_{i=0}^{n_k} w_k^i\,\mathcal{N}(\mathbf{x}_k;\mathbf{x}_{k|k}^i,\mathbf{P}_{k|k}^i)\,\mathcal{I}(\mathbf{s}_k;\mathbf{s}_{k|k}^i,\mu_{k|k}^i)\,. \tag{4}$$

where the estimates $\mathbf{x}_{k|k}^i$ and covariances $\mathbf{P}_{k|k}^i$ are calculated by the known Kalman filter update equations, and weights w_k^i, signal strength estimates $\mathbf{s}_{k|k}^i$ and $\mu_{k|k}^i$ are given in [Koc06].

3.3 Algorithmic implementation

The presented update scheme is implemented into the single-target PDAF [BSL93] and the multi-target JPDAF [BSLK01] tracking algorithms. In the PDAF algorithm, n_k measurements lead to $n_k + 1$ hypotheses which are merged to a single main hypothesis by second-order moment matching at the end of the filter update step. In the JPDAF algorithm, the complete set of possible global hypotheses is processed, i.e. all possible combinations to associate measurements to tracks including missed detections and false alarms are considered. Thus it avoids the association of a certain measurement to more than one track, which is a shortcoming of the simple single-target PDAF.

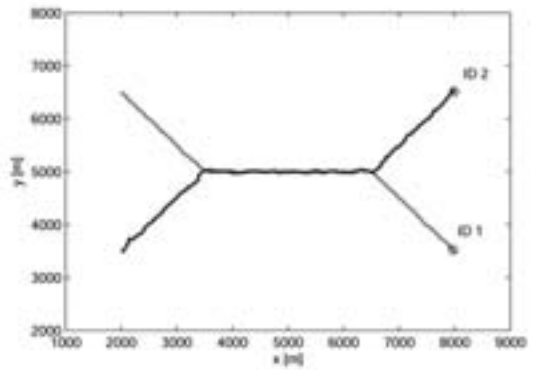


Figure 1: Two-target simulation scenario. Shown is the snapshot at the final revisit time.

4 Simulation scenario & results

We consider a simulation with two targets approaching and moving along the same trajectory for a considerable period of time (see Fig. 1), leading to a loss of identity when traditional tracking algorithms are used. The objective is to determine the capability of target discrimination with signal strength information in the final stage of the scenario. Based on the JPDAF algorithm and the simulation parameters given in Tab. 1, the results are plotted in Fig. 2 for Swerling-I and III fluctuations, respectively, of the two targets. The figure depicts the probability for correct (solid lines) and incorrect (dashed lines) association of tracks with true targets at the final revisit for different signal strength combinations, based on the two fluctuation models. The signal strength of one target is fixed, corresponding to the minimum value of each line. The upper plots correspond to target tracking without signal strength information. In this case, the signal strength only affects the true detection probability and thus the occurrence of target detections. As expected, for all combinations of the two targets' signal strength the probability for correct association amounts to 50%. The lower plots illustrate the advantage of taking signal strength information into account:

Table 1: Simulation parameters

Monte Carlo runs	N_{MC}	=	1000
Target velocity	v_{target}	=	15 m/s
Process noise	σ_p	=	0.5 m/s^2
Range error	σ_r	=	10 m
Azimuth error	σ_φ	=	0.25
Mean false alarms	$\bar{n}_{FA}$	=	1
Field of view	$\lvert V \rvert$	=	10 km $\times$ 10 km
Sensor position	$\vec{r}_{sensor}$	=	$[-1, 5, 10]^\top$ km
Revisit rate	Δ_T	=	2 s
Detector threshold	λ	=	4
Mean clutter SNR	CNR_0	=	10

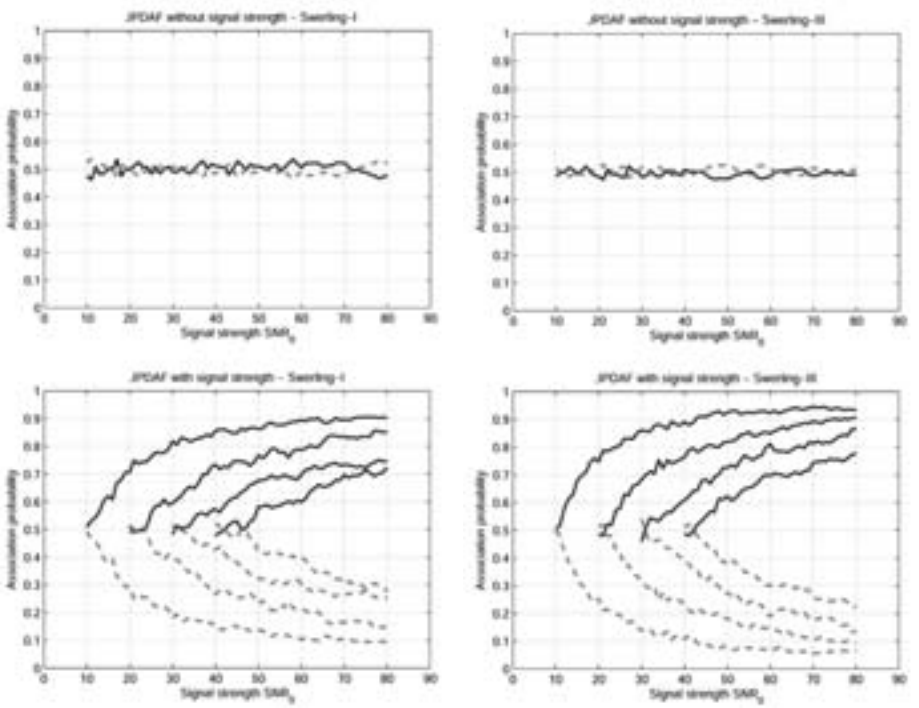


Figure 2: Probability for correct (solid lines) and incorrect (dashed lines) association of tracks with true targets at final revisit for different signal strength combinations, based on fluctuation model Swerling-I (left) and Swerling-III (right). The signal strength of one target is fixed, corresponding to the minimum value of each line. Shown are the results for JPDAF without (above) and with (below) exploitation of signal strength information.

A clear performance gain in target discrimination is visible for both fluctuation models, with a stronger separation capability for larger differences in signal strength. Comparing the results in Fig. 2 (lower plots), apparently the discrimination performance of the algorithm is slightly higher in case of Swerling-III. This is probably due to the distinct peak in the probability density of Swerling-III fluctuations, leading to a more obvious discrimination of the two signal strength distributions.

5 Conclusions

In this paper we developed a tracking algorithm which incorporates signal strength information. In contrast to previous approaches, the knowledge on the target's signal strength is not only used for an improved calculation of the association probabilities, but it enters into the algorithm as a random variable which is sequentially estimated. Based on simulation scenarios, the performance of the presented algorithm is evaluated. As expected, the exploitation of signal strength leads to a performance gain in target discrimination in a multi-target scenario.
Further work will focus on the incorporation of signal strength information into the CPHD algorithm [Mah07] and on a multiple model approach for different target fluctuation models.

References

[BSL93] Yaakov Bar-Shalom and Xiao-Rong Li. *Estimation and Tracking: Principles, Techniques, and Software*. Artech House, Boston, MA, 1993.

[BSLK01] Yaakov Bar-Shalom, Xiao-Rong Li, and T. Kirubarajon. *Estimation with Applications to Tracking and Navigation*. Wiley and Sons, Boston, MA, 2001.

[DCV10] Ba-Ngu Vo D. Clark, B. Ristic and B.T. Vo. Bayesian Multi-Object Filtering With Amplitude Feature Likelihood for Unknown Object SNR. *IEEE Transactions on Signal Processing*, 58(1):26–37, 2010.

[Koc06] Wolfgang Koch. Tracking and Data Fusion Applications. In *Advanced Radar Signal and Data Processing, Educational Notes RTO-EN-SET-086*, number 9. NATO-RTO, 2006.

[LBS90] D. Lerro and Y. Bar-Shalom. Automated Tracking With Target Amplitude Information. In *Proceedings of the American Control Conference*, 1990.

[Mah07] R.P.S. Mahler. *Statistical Multisource-Multitarget Information Fusion*. Artech House, 2007.

[Swe60] P. Swerling. Probability of Detection for Fluctuating Targets. *IRE Trans. on Information Theory*, IT-6:269–308, 1960.

[vK96] G. van Keuk. Multihypothesis Tracking Using Incoherent Signal-Strength Information. *IEEE Transactions on Aerospace and Electronic Systems*, 32(3):1164–1168, 1996.

First Order Multiple Hypothesis Testing for the Global Nearest Neighbor Data Correlation Approach

Amichai Painsky

4D Security Solutions
107D Corporate Boulevard
South Plainsfield, New Jersey, 07080
United States
Amichai_painsky@4-dsecurity.com

Abstract: The growing necessity in multiple targets tracking (MTT) in surveillance systems, with the recent dramatic increase in computational capabilities, has lead to a major interest in improving the performance of classical methods, such as the Global Nearest Neighbor (GNN), to enhanced schemes of Data Correlation. Today, the Multiple Hypothesis Testing (MHT) is generally accepted as the preferred approach for MTT systems, as it demonstrates better results in more complicated and uncertain environments. However, embedding such a mechanism to a deployed GNN-based system requires an extensive software change, and may introduce a major engineering risk to the working environment. Moreover, in a system that is deployed at different sites, which addresses operational environments of different complexities, such a change may be too costly and even superfluous. In this paper we will present a method which will address the challenge of multiple targets tracking in changing environments through a First Order Multiple Hypothesis Testing for a Global Nearest Neighbor engine. We will start with presenting the basics of multiple targets tracking, followed by a review of the proposed solution and conclude with simulations to verify its performance in different scenarios.

1 Introduction

Estimating a single target's state from a noisy set of observations is one of the fundamental problems in estimation theory. Several solutions have been suggested to minimize different error measures. Focusing on the mean square error (MSE) measure, and under the constraint of a discrete time linear estimator, the Kalman filter provides an efficient, recursive, real-time, optimal solution, which minimizes the MSE and converges to the non-linear optimal solution if the noise is known to be normally distributed. Assuming a simplified dynamic true state model:

$$\underline{X}_k = F \cdot \underline{X}_{k-1} + B \cdot a_k$$

$$\underline{Y}_k = \underline{X}_k + \underline{V}_k$$

Where $\underline{X}_k$ is the target's state vector (position and velocity), F is the state transition model, B and a_k are the acceleration update model, and V_k is a normally distributed measurement noise:

$$\underline{X}_k = \begin{bmatrix} X_k \\ \dot{X}_k \end{bmatrix}, \text{F}=\begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix}, \text{B}=\begin{bmatrix} \frac{\Delta t^2}{2} \\ \Delta t \end{bmatrix}, a_k \sim N(0, \sigma_a^2), \underline{V}_k \sim N(0, R_k).$$

We define:

$$w_k = B \cdot a_k \sim N(0, Q_k)$$

$$Q_k \triangleq Var(B \cdot a_k) = \sigma_a^2 B B^T$$

$\underline{\hat{X}}_{\text{k|k}}$ - The a-posteriori state estimate at time *k* given observations up to and including time *k*.

$\text{P}_{\text{k|k}}$ - The a-posteriori error covariance matrix at time k, given observations up to and including time *k*.

Then, the Kalman Filter solution consists of two phases, as follows:

Predict

Predicted (a priori) state:

$$\underline{\hat{X}}_{\text{k|k}-1} = F \cdot \underline{\hat{X}}_{\text{k}-1|\text{k}-1}$$

Predicted (a priori) estimate covariance:

$$\text{P}_{\text{k|k}-1} = F \cdot \text{P}_{\text{k}-1|\text{k}-1} \cdot F^T + Q_k$$

Update

Innovation or measurement residual:

$$\underline{\tilde{\text{Y}}}_{\text{k}} = \underline{\text{Y}}_{\text{k}} - \underline{\hat{\text{X}}}_{\text{k|k}-1}$$

Innovation (or residual) covariance:

$$\text{S}_{\text{k}} = \text{R}_{\text{k}} + \text{P}_{\text{k|k}-1}$$

Optimal Kalman gain:

$$\text{K}_{\text{k}} = \text{P}_{\text{k|k}-1}\text{S}_{\text{k}}^{-1}$$

Updated (a posteriori) state estimate:

$$\underline{\hat{\text{X}}}_{\text{k|k}} = \underline{\hat{\text{X}}}_{\text{k|k}-1} + \text{K}_{\text{k}}\underline{\tilde{\text{Y}}}_{\text{k}}$$

Updated (a posteriori) estimate covariance:

$$\text{P}_{\text{k|k}} = (\text{I} - K_k)\text{P}_{\text{k|k}-1}$$

However, in a surveillance system where multiple targets have to be tracked and monitored simultaneously, the problem becomes more complicated. Assuming we are given the association between each target and its observation, multiple targets tracking through independent Kalman filters will independently minimize each estimated target's MSE, thus minimize the overall MSE. Therefore, using the structure of the Kalman Filter, an optimal solution for the MTT problem can be described as in Figure 1.

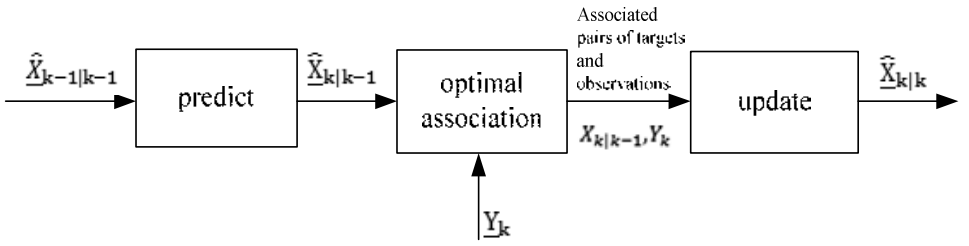


Figure 1: Block diagram for a multi-target estimation process

Based on this approach, the better the performance of the association phase is, the closer we get to the optimal solution.

Several methods were introduced to reflect the association's sub-optimality in the model and by that improve the estimation performance. In [Na69], [SS71] it was proposed to reflect the association uncertainty in the targets' estimate covariance matrix. A different method, the Joint Probabilistic Data Association (JDPA) [Fo80], suggests allowing weighted sum association of a single observation to multiple targets in uncertain association conditions (dense and cluttered environment, maneuvering targets). This means that a single observation may contribute to the update of several targets at a time, based on a probabilistic approach.

However, forcing an increased Kalman Filter estimate covariance matrix to reflect errors in different modules of the system may actually result with a reduced overall performance of the system [FV87], while associating a single observation to several targets may cause extremely similar targets in a dense environment, where several close observations contribute to several close targets.

In this paper we will tackle the sub-optimality of the association phase by taking into consideration a number of feasible associations and choosing the one which achieves the best association result, given a set of incoming observations. By that, we will address the possibility of false association by allowing an immediate recovery followed by corrected continuous tracking

2 Association Phase

As illustrated above, the association phase receives a predicted targets vector $\underline{\hat{X}}_{k|k-1}$ and an observations vector $\underline{Y}_k$, and generates pairs of predicted targets and corresponding observations. The association is calculated so the global price of association is minimized, where the price is defined as a distance measure between a predicted target and its associated observation. For example:

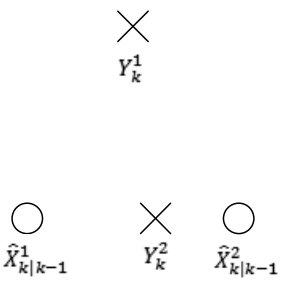


Figure 2: Observation-to-target association

Where $\hat{X}_{k|k-1}^1$, $\hat{X}_{k|k-1}^2$ are the positions of the predicted targets at time k, and Y_k^1, Y_k^2 are the positions of the observations at time k.

In this example, we chose the distance measure to be:

$$d^T S^{-1} d + \log(\det(S))$$

Where d is the geographical distance vector between the observation and the target and S is the observation's covariance matrix. The logarithm of the determinant of the covariance matrix is a value which corresponds to the volume of the covariance matrix, thus the uncertainty of the observation.

Therefore, the association of $\hat{X}_{k|k-1}^1$with Y_k^1 and $\hat{X}_{k|k-1}^2$with Y_k^2 will minimize the global statistical distance, assuming both observations' covariance matrices are equal.

The two-dimensional association problem is a well studied problem in the optimization field which can be solved using the generalized Auction algorithm [Be98] or a network flow algorithm [Go91] in (pseudo) polynomial time. Since these algorithms are well known they will not be described in this paper. The reader is referred to [Be98] and [Go91] for further details.

In order to reduce the computational load of the association phase, and define the minimal distance above which a target and an observation are not considered as candidates for association, we use the concept of Gating. In the pre-association gating phase the observations' and targets' positions are compared to each other to generate for each target its list of observations that are candidates for association. Graphically:

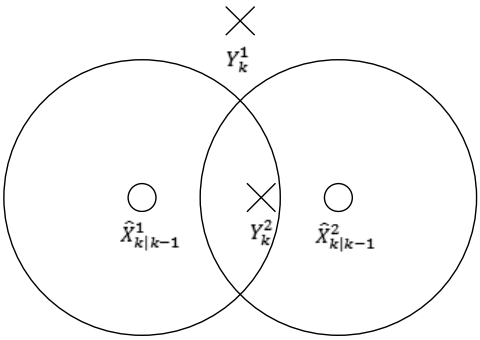


Figure 3: Gating the observation-to-track association

Where the circles represent the distance measure above which observations will not be considered as candidates for association for that target.

3 Dense Environment and Maneuvering Targets

In a well-spread target environment, with slow maneuvering targets, the number of candidate observations for each target is limited and result with a certain and undoubtedly preferred observation-to-target association. However, in a dense environment with fast maneuvering targets, the preferred association may only be slightly better than other feasible ones. Moreover, the specific environment which resulted several preferred associations may become clearer when future observation are available, thus making the decision easier. Therefore, by choosing only the best association at any given time we may even cause a propagation of error, when the environment is ambiguous.

For example:

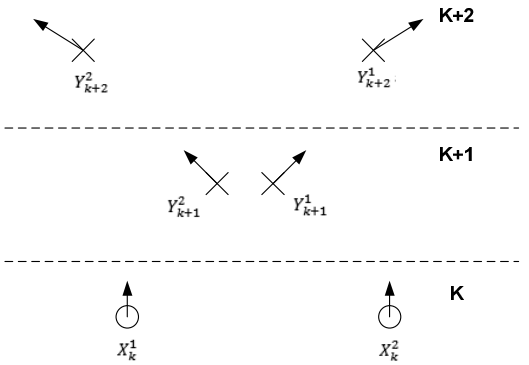


Figure 4: Crossing targets and corresponding observations

Where X_k^m and Y_k^m are target *m* and observation *m* at time *k*, and the arrows represent velocity vectors for each target and observation. In this example, the two targets cross each other between time stamps *k* and *k+1*. However, the best association at time *k+1* will be $X^1_{k+1|k}$, Y^2_{k+1} and $X^2_{k+1|k}$, Y^1_{k+1} which will result with an increasing MSE and falsely target tracking at time *k+2* and beyond. To overcome such hard decisions, we introduce the concept of First Order Multiple Hypothesis Testing for the Global Nearest Neighbor Data Correlation Approach.

4 Multiple Hypothesis Testing for the Global Nearest Neighbor Approach

As described above, making a hard decision at any time interval may lead to a propagation of error in targets tracking. Therefore, we shall regard not only the best association, but also store a configurable number of possible target sets that were generated from the N best (or better) feasible associations. We will then compare it against the new set of incoming observations. The target set which will result with the best association with the new observations will be updated with those observations and displayed to the user. The N-1 best (or better) associations shall be updated with the observations and stored for the next iteration.

In a pseudo code:

1) **First set of observations introduced to the system (initialization):**
   All observations generate new targets (no existing targets yet)

2) **Second set of observations:**
   a. Acquire sensor's observations.
   b. Predict targets (compute $\underline{\hat{X}}_{k|k-1}$)

c. Generate observations candidate list for association, for each target (Gating phase).
d. Find N best (better) observation-to-target associations.
e. Update targets ($\underline{\hat{X}}_{k|k}$) according to the N associations to generate N hypotheses (target sets), and store in memory.
f. Display the best hypothesis.

3) For the following sets of observations:

a. Acquire sensor's observations
b. Predict all stored target sets (compute $\underline{\hat{X}}_{k|k-1}$)
c. For each of the N stored hypotheses (targets sets):
   - Generate observations candidate list for association, for each target (Gating phase).
   - Find N best (better) observation-to-target associations.

d. Find the N best (better) observation-to-target associations from the associations calculated in the previous step.
e. Update targets ($\underline{\hat{X}}_{k|k}$) according to these N associations to generate N hypotheses, and store in memory.
f. Display the best hypothesis.

To guarantee that the system will not overwork in non-ambiguous scenarios, in which the observations are well-spread and easily associated, leading to a single association which achieves significantly better results, we introduce the Quality Factor. The Quality Factor is a threshold which defines a relative value of the best association below which an association will not be considered as possible hypothesis and will not be stored.

Therefore, by setting the Quality Factor to a high enough value, we make sure the system will only generate multiple hypotheses in an ambiguous environment, and converge to a standard Global Nearest Neighbor solution in spread and easily analyzed scenarios.

A great deal of work has been done in the area of calculating N-best associations [CM95], [DN93], [Mi95], [Mu68], in the context of solving an S-D assignment problem [DP92a], [De97]. These methods were used as guidelines in our algorithm's implementation, where their sub-optimality was handled by the Quality Factor threshold, as described above.

5 Simulations and Performances

In order to verify the performance of the First Order Hypothesis Testing for the Global Nearest Neighbor algorithm and quantify its behavior in different environments, we randomly generated targets in different scenarios and observed them through real deployed sensors models. We then injected the observations to the system and compared the estimated targets we got with their corresponding simulated ones, in terms of MSE. In all simulations the sensors models generated observations according to time-constant covariance matrices, providing a set of observations every two seconds. The area of experiment is 0.5 square kilometer.

- Simulation no. 1: Increasing Velocities

  Quality Factor: 0.7

  Number of simulated targets: 20

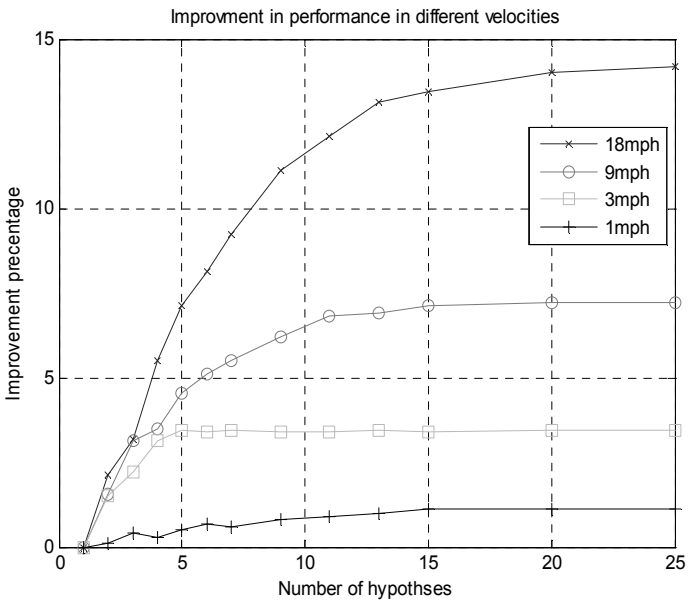


Figure 5: Improvement percentage as a function of maximal number of hypotheses

The following chart describes the relative improvement in the MSE of the simulated and the estimated targets at different average targets' velocities. The X axis describes the maximal number of hypotheses the system may generate and store, while the Y axis describes the percentile MSE improvement, compared to a single hypothesis testing (GNN).

We can first notice that the relative improvement increases, on average, as we allow more hypotheses testing to be made for a given scenario. This happens up to a saturation level above which increasing the number of hypotheses does not result with a significant improvement. The reason is that for a given scenario, the algorithm tends to converge to a preferred target tracking (in the described sense) for a certain number of hypotheses. From this point on, increasing the number of allowed hypotheses will not show a significant increase in the relative improvement.

Moreover, we can see the relative improvement in the MSE becomes more significant as the velocity of the simulated targets increases. The greater velocity makes the tracking of the simulated targets more complicated, which results with more uncertainty in the association phase. Therefore, in higher velocities, the Multiple Hypothesis Testing demonstrates greater improvement in the performance of the system.

- Simulation no. 2: increasing number of targets

  Quality Factor: 0.7

  Targets average velocity: 9 mph.

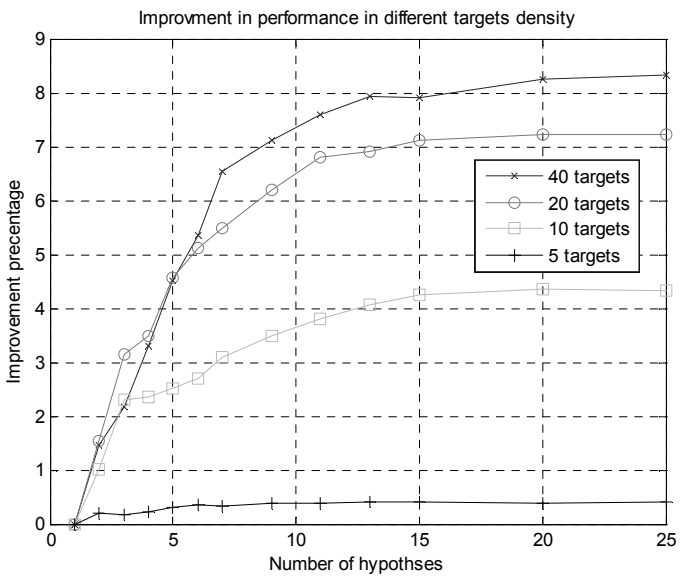


Figure 6: Improvement percentage as a function of maximal number of hypotheses

In this simulation we can see the effect of the targets density on the performance of the algorithm. Increasing the number of targets at a given area results in a more ambiguous association making, which is reflected in greater and closer association results. Therefore, allowing more hypotheses increase the performance in higher density environments.

- Simulation no. 3: decreasing Quality Factors

  Number of targets: 20

  Targets average velocity: 9 mph

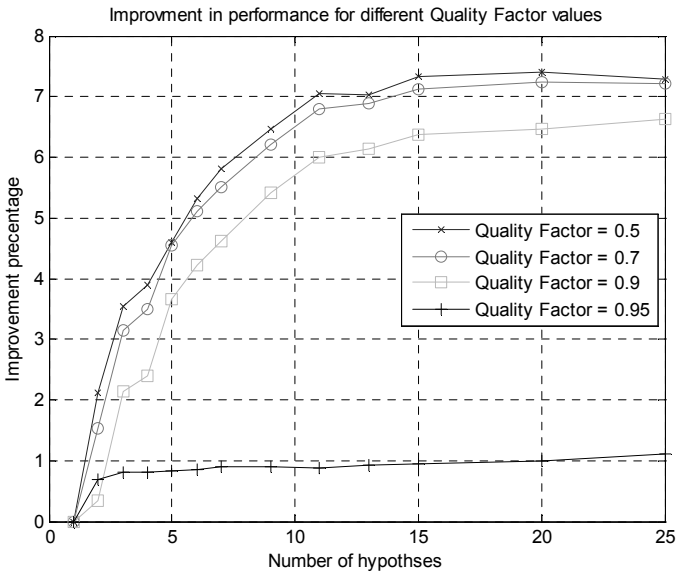


Figure 7: Improvement percentage as a function of Quality Factor values

Here we can see that in uncertain environments, decreasing the quality factor improves the performance of the system to a certain point above which it tends to saturate. The reason for this behavior is that decreasing the Quality Factor allows more, but less valuable, hypotheses testing. These hypotheses may improve the performance only in very ambiguous situations, in which the environment was so hard to track that a very unlikely association ended up being the better one to choose.

6 Conclusions

In this paper we introduced the concept of First Order Multiple Hypothesis Testing for a Global Nearest Neighbor Data Correlation Approach. Generating and storing multiple target hypotheses allows an immediate recovery in case of a false decision in uncertain association environments. In addition, it improves the system's ability to handle multiple target tracking, and creates a more accurate situational picture for the system's operator.

Introducing the Quality Factor, and a configurable number of maximum hypotheses testing, assures the system can be easily adjusted to different environments to meet the exact necessary tradeoff between its estimation accuracy and computational load.

The simulations we held demonstrated significant improvement comparing to the classical Global Nearest Neighbor Approach in more ambiguous scenarios, such as dense and cluttered environment or fast maneuvering targets.

Implementing a First Order Multiple Hypothesis testing for the Global Nearest Neighbor Approach does not necessitate an extensive software change; it is easily adjustable to meet different performance requirements and does not entail sophisticated or complex pruning mechanisms as other MHT-based methods. All these qualities make the upgrade from the classical GNN to an enhanced performance engine less complicated, in terms of code change and risk mitigation, for real-time deployed Command and Control surveillance systems.

7 References

[BL95] Y. Bar-Shalom and X. Li, Multi-target Multi-sensor Tracking: Principles and Techniques. Storrs, CT: YBS Publishing, 1995.

[BL93] Y. Bar-Shalom and X. Li, Estimation and Tracking: Principles, Techniques and Software. Boston: Artech House, 1993.

[Na69] N. Nahi, Optimal recursive estimation with uncertain observation. IEEE Transactions on Information theory, IT-15, 4, July 1969, 457-462

[SS71] R. A. Singer and J. J. Stein, An optimal tracking filter for processing sensor data of imprecisely determined origin in surveillance system. In Proceedings of 1971 IEEE conference on Decision and Control, Miami Breach, FL, Dec, 1971, 171-175.

[Bl86] S. Blackman, Multiple target tracking with Radar Applications. Dedham, MA: Artech House, 1986.

[FV87] W. Felskes and G. van Keuk, On single target tracking in dense clutter environment-quantitative results. In Proceedings of 1987 International Radar Conference, 130-134.

[Be98] D. Bertsekas, The auction algorithm: A distributed relaxation method for the assignment problem. Annals of Operations Research: Special Issue on Parallel Optimization, 14 (1998), 105-123

[Dr90] O. Drummond, D. Castanon, and M. Bellovin, Comparison of 2-D assignment algorithms for sparse rectangular, floating point, cost matrices. Journal of SDI Panels on Tracking, Institute of Defense, Analyses, Alexandria, VA, 2/1990 (Dec. 1990), 81-97

[BJ14] F. Bourgeois and L. Jean-Claude, An extension of the Munkres algorithm for the assignment problem to rectangular matrices. Communications of the ACM, 14(12).

[BT75] Y. Bar-Shalom, and E. Tse, Tracking in a cluttered environment with probabilistic data association. Automatica, vol. 2, pp. 451-460, Sept. 1975.

[Fo80] T. E. Fortmann, Y. Bar-Shalom and M. Scheffe, Multi-target tracking using joint probabilistic data association. In Proceedings with 1980 IEEE Conf. on Decision and Contr., Dec. 1980, pp. 807-812.

[Po94] A. Poore, Multi-dimensional assignment formulation of data association problems arising from multi-target and multi-sensor tracking. Computational Optimization Application., vol. 3, pp. 27-57, 1994.

[SL93] R. L Streit and T. E. Luginbuhl, A probabilistic multi-hypothesis tracking algorithm without enumeration and pruning. In Proceedings with 6th Joint Device Data Fusion Symp., Laurel, MD, 1993.

[St94] R. L. Streit, Maximum likelihood method for probabilistic multi-hypothesis tracking. SPIE Proc., vol. 2234-5, pp. 394-406, 1994.

[Ma87] R. Mahler, A unified foundation for data fusion. In Proceedings with 1994 Data Fusion Syst. Conf., Applied Phys. Lab., Johns Hopkins Univ., June 1987.

[Bl04] S. Blackman, Multiple hypothesis tracking for multiple target tracking. IEEE A & E Systems Magazine, vol. 19, no. 1, part 2, pp. 5-18, 2004

[PR93] A. Poore and N. Rijavec, A Lagrangian relaxation algorithm for multidimensional assignment problems arising from multi-target tracking. SIAM Journal of Optimizations 3, pp. 544-563, Aug. 1993.

[Br89] W. Brogan, Algorithm for ranked assignments with applications to multi-object tracking. Journal of Guidance, 12, 3 (May 1989), 357-364, 1989.

[CM95] I. Cox and M. Miller, On finding ranked assignments with application to multi-target tracking and motion correspondence. IEEE Transactions on Aerospace and Electronic Systems, 32, 1, 486-489, Jan. 1995

[DN93] R. Danchick and R. Newnam, A fast method for finding the exact N-best hypotheses for multi-target tracking. IEEE Transactions on Aerospace and Electronic Systems, 29, 2 (Apr. 1993), 555-560.

[DP92a] S. Deb, K. Pattipati and Y. Bar-Shalom, An S-dimensional assignment algorithm for track initiation. In Proceedings of IEEE Int'l. Conf. on Engrg. Systems, Kobe, Japan, Sept. 1992.

[DP92b] S. Deb, K. Pattipati, and Y. Bar-Shalom, A new algorithm for the generalized multidimensional assignment problem. In Proceedings of IEEE Int'l. Conf. on Systems, Man and Cybernetics, Chicago, IL, Nov. 1992.

[DP93] S. Deb, K. Pattipati , and Y. Bar-Shalom, A multi-sensor multi-target data association algorithm for heterogeneous sensors. IEEE Transactions on Aerospace and Electronic Systems, 29, 2, 560- 568. Apr. 1993

[De97] S. Deb, M. Yeddanapudi, K. Pattipati and Y. Bar-Shalom, A generalized S-D assignment algorithm for multi-sensor multi-target state estimation. IEEE Transactions on Aerospace and Electronic Systems, 33, 2, 523-538, Apr. 1997

[Mi95] M. Miller, H. Stone, and I. Cox, Optimization to Murty's algorithm for determining a ranked set of assignments. Technical Report, NEC Research Institute, 1995

[Mu68] K. Murty, An algorithm for ranking all the assignments in order of increasing cost. Operations Research, 16 (1968), 682-687

[Po99] R. Popp, K. Pattipati, and Y. Bar-Shalom, Dynamically adaptable m-Best 2D assignment algorithm and multi-level parallelization. IEEE Transactions on Aerospace and Electronic Systems, 35, 4, 1145-1160, Oct. 1999.

[Be79] D. P. Bertsekas, A distributed Algorithm for the Assignment Problem. Lab. for Information and Decision Systems Working Paper, M.I.T., March 1979.

[Be81] D. P. Bertsekas, A new algorithm for the assignment problem. Math Programming, vol. 21, 1981, pp. 152-171.

[GH88] K. G. Gilbertand and R. B. Hofstra, Multidimensional assignment problems. Decision Sciences, 19 (1988), 306-321.

[Go91] A. V. Goldberg, S. A. Plotkin and E. Tardos, Combinatorial algorithms for the generalized circulation problem. Mathematics in Operations Research, 16 (1991), 351-381.

Track Heading and Speed Estimation

Dr. Wolfgang Konle

ISEG 46
EADS Germany
Clude-Dornier-Str.
88090 Immenstaad
wolfgang.konle@eads.com

Abstract: A new method to enhance the performance of multi radar tracking systems is presented in this paper. In parallel to an IMM model type tracker a second tracker estimates track heading and speed based upon differential measurements from multiple sensors. A differential measurement contains the differences between two subsequent measurements from one sensor. This new method accepts information from multiple sensors but completely avoids sensor registration problems.
While a common multi radar tracker must take into account at least residual registration errors, the heading/speed tracker accepts the full accuracy of each individual sensor. This leads to high manoeuvre susceptibility and also to a high stability in the non manoeuvre case.
The results demonstrate the tracking performance enhancement achieved by the combination of the new method with the standard tracking method.

1 Introduction

A major problem in tracking of multiple targets with multiple sensors is sensor registration. Even after application of high sophisticated sensor alignment procedures a residual systematic error remains. Taking into account this non Gaussian error, forces to apply more restrictive filters. The side effect of restrictive filters is a reduced tracking reactivity.

A method to avoid the effect of sensor misalignment on the track speed and heading is presented in this paper. Key to overcome the sensor registration problems is the usage of differential information from subsequent measurements instead of absolute values.
The differential azimuth ($d\alpha$) and the horizontal component of the differential range (dr) from all available sensors are used to derive the horizontal speed (v) and the heading (φ) as well as the acceleration (a, dv/dt) and the turn rate (ω, $d\varphi/dt$)

The basic idea is to establish an additional separate tracking filter for the estimation of speed and heading, which is stable and robust in respect to misalignment and highly sensitive to target manoeuvres.

The results presented in the paper show, that the separate speed and heading filter provides higher accuracy of speed and heading values as the operational multi sensor tracking system.

1.1 Combination of Filtering Methods

The track heading/speed filter provides estimation values for horizontal speed, heading, acceleration, and turn rate. This information is not sufficient to solve multi target association problems. Therefore only a combination of both methods leads to a complete tracking system. In the combination, both filters are fed with the same measurements and operate in parallel. As soon as the new filter provides a state vector, it can be mixed with the according values derived by the common filter or even completely replace the common filter's track state components horizontal speed, heading, acceleration or turn rate.

2 The Speed and Heading Filter

The Speed and Heading Filter is a common extended Kalman Filter [1] for the estimation of horizontal speed v, heading φ, acceleration a, and turn rate ω. The measurements comprise the distance difference dr and the azimuth difference dα of two subsequent radar measurements (r_1, α_1), (r_2, α_2).

2.1 System model

The system model comprises simple relaxation and maneuver states and is expressed by the following relations:

$$v(t+dt) = v(t) + a\,dt\,e^{-dt/\tau_1} \quad (1)$$

$$\varphi(t+dt) = \varphi(t) + \omega\,dt\,e^{-dt/\tau_2} \quad (2)$$

The relaxation times τ_1 for the acceleration relaxation and τ_2 for the turn relaxation are different and depend on the maneuver state.
In matrix notation it reads as

$$x_k(-) = \Phi_{k-1} x_{k-1}(+) \quad (3)$$

2.2 Measurement model

The measurement model describes the relation between the measurement vector z=(dr, dα) and the state vector x=(v, φ, a, ω) and is expressed by the function z=h(x).

Figure 2-1 shows the geometrical relation between the state vector (v, φ) and the raw measurement vectors (r_1,α_1), (r_2,α_2).

After some algebraic reformatting we finally get:

$$dr = vdt\cos(\varphi-\alpha_2) - r_1(1-\cos(\alpha_2-\alpha_1)) \quad (4)$$

$$d\alpha = \arcsin(vdt\sin(\varphi-\alpha_1)/r_2) \quad (5)$$

for the relation between (dr, dα) and (v, φ, a, ω).
The relation shows, that the model is complete, which means that the measurement values z can be expressed by a function h(x) with the state variables and given constants.

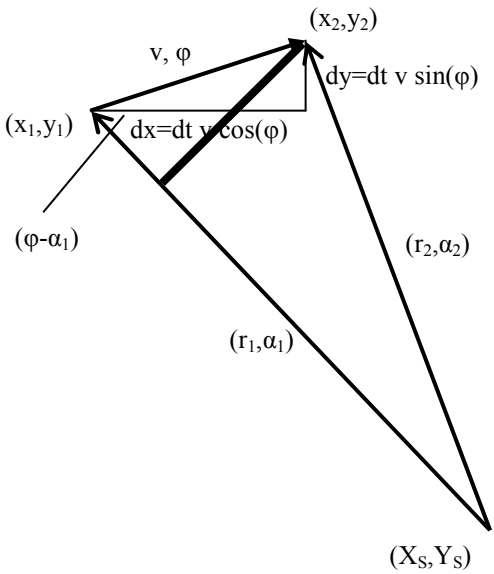


Figure 2-1 Measurement Geometry

2.3 Differential measurement model

The differential measurement model is the Jacobian defined as the partial derivatives of the measurement variables in respect to the system variables: $H_{ij}= \partial h_i/\partial x_j$.
We get:

$$H_{11}: \partial dr/\partial v = dt\cos(\varphi-\alpha_2) \quad (6)$$

$$H_{12}: \partial dr/\partial \varphi = -vdt\sin(\varphi-\alpha_2) \quad (7)$$

$$H_{21}: \partial d\alpha/\partial v = \sin(\varphi-\alpha_1)r_1dt/r_2^2 \quad (8)$$

$$H_{22}: \partial d\alpha/\partial \varphi = vdt(vdt+r_1\cos(\varphi-\alpha_1))/r_2^2 \quad (9)$$

2.4 Measurement error model

In accordance to the measurement definition, the difference between two subsequent measurements, we get an error covariance matrix R with doubled variance contributions:

$$R = \begin{matrix} 2\sigma_r^2 & \\ & 2\sigma_\alpha^2 \end{matrix}, \quad (10)$$

2.5 System noise model, manoeuvre detection

The system noise model expresses the variation of the system covariance P due to possible target manoeuvres. Possible target manoeuvres may begin at time t_{k-1} and lead to an increment Q in the system covariance. With this incremental term we get for the error covariance extrapolation:

$$P_k(-)=\Phi^c_{k-1}(P_{k-1}(+)+Q_{k-1})\Phi^{c\,T}_{k-1} \quad (11)$$

The "c" in the propagation matrix denotes, that the system propagation matrix Φ used in (3) differs from the error propagation matrix Φ^c used in (11).

The system noise matrix Q contains the two components Q_{33} and Q_{44} which express the possible acceleration and turn rate changes. The numerical values are derived from the currently considered measurement triple with the current measurement and the last two additional measurements from the same sensor. The instantaneous adaptation of the system noise to the geometrical shape of a set of observations is very efficient maneuver detection. The opposite, "non maneuver detection," if the acceleration and turn rate is below a threshold, is a not less important criterion to detect maneuver termination or to avoid maneuver initiation upon noise.

2.6 Extended Kalman Filter Definition

With the equations (3..12) the definition of an extended Kalman Filter is complete. The additional equations for the Kalman Gain Matrix (12), the state estimate update (13) and the error covariance update (14) are directly derived from equations (3..12) [1,Table 4.2-1]

$$K_k=P_k(-)H_k^T[\,H_k]P_k(-)H_k^T+R_k]^{-1} \quad (12)$$

$$x_k(+)=x_k(-)+K_k[z_k-h_k(x_k(-))] \quad (13)$$

$$P_k(+)=[\,I-K_kH_k]P_k(-)] \quad (14)$$

The Extended Kalman Filter defined by the equations (1...14) is able to update tracks with subsequent measurements z = (dr, dα). Due to the composite nature of the measurements a problem with the time order arises.

The validity time of a tuple m of measurements $m_1(t_1),m_2(t_2)$ with the measurement values dr,dα is valid at time

$$t_m = (t_1+t_2)/2=(t_1+dt)/2. \quad (15)$$

The effect of this validity time t_m is that the time stream t_i of measurements $m_i=(t_i,dr_i,d\alpha_i)$ is not monotonous. Nevertheless, the Extended Kalman Filter algorithm requires a processing of the information in a strictly monotonous time order. Let us assume that the slowest radar sensor has a turn time T_{max} and the fastest sensor has a

turn time T_{min}. Let us further assume that after one miss of the slowest sensor the resulting measurement tuple with $dt=2T_{max}$ shall be taken into account immediately after a tuple of the fastest sensor has been processed. After processing a tuple of the fastest sensor available at time t_{now}, the minimum time of validity is

$$t_{val/minl} = t_{now} - T_{min}/2. \quad (16)$$

After processing a measurement tupel with one miss(!) from the slowest sensor, the maximum time of validity is

$$t_{val/max} = t_{now} - T_{max}. \quad (17)$$

The maximum possible time difference dt_{max} is therefore

$$dt_{max} = t_{val/minl} - t_{val/max} = T_{max} - T_{min}/2. \quad (18)$$

If as an example T_{max} is 12 seconds and T_{min} is 3 seconds, the validity time becomes rolled back for 10.5 seconds after the reception of the measurement from the slowest turning sensor after one miss. In order to be able to process the incoming information from the slow sensor within the correct time order, the processing since $t_{val/actual}$-dt_{max} must be repeated with all sets of measurement in the correct time order.

In order to be able to accomplish this, a set of state information back in time with a delay of dt_{max} must be kept in mind for all tracks in combination with all measurement tuples $m_i = (t_{val}, dr, d\alpha)_i$ within the validity time interval I_{time}.

$$I_{time} = t_{val/actual} > t_{vali} > t_{val/actual} - dt_{max} \quad (19)$$

2.7 Initiation model

Initial values x(~) for the system state x based upon one set of measurements z are given by the inverse measurement model function $x(\sim)=h^{-1}(z)$.

But instead of inverting the equations (4,5) it is more easy to derive the inverse measurement model function directly from the geometry. In order to be able to derive the full state vector including a and ω, it is necessary to consider three subsequent measurements of the target position P_0, P_1, P_2.

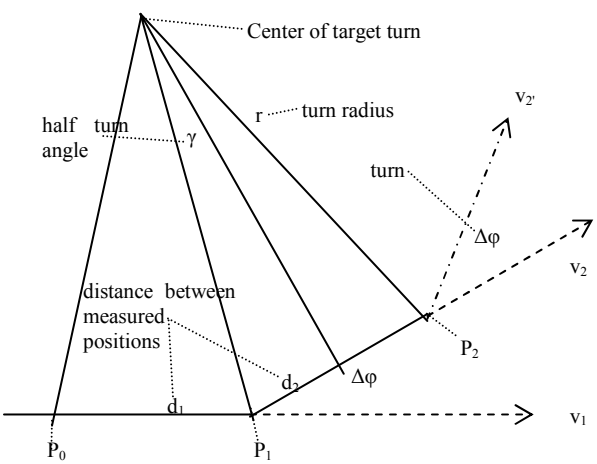


Figure 2-2 Target Movement Geometry

Figure 2-2 shows the geometrical relations based upon three subsequently measured positions $P_0(t_0)$, $P_1(t_1)$, $P_2(t_2)$

We get:

$$v_1((t_0+t_1)/2) = d_1/(t_1-t_0) \quad (15)$$

$$\varphi_1((t_0+t_1)/2) = \arctan((P_1(y)-P_0(y))/(P_1(x)-P_0(x))) \quad (16)$$

$$v_2((t_1+t_2)/2) = d_2/(t_2-t_1) \quad (17)$$

$$\varphi_2((t_1+t_2)/2) = \arctan((P_2(y)-P_1(y))/(P_2(x)-P_1(x))) \quad (18)$$

$$a((t_1+t_2)/2) = \frac{v_2((t_1+t_2)/2)-v_1((t_0+t_1)/2)}{(t_1+t_2)/2 - (t_0+t_1)/2} \quad (19)$$

$$\omega((t_1+t_2)/2) = \frac{\varphi_2((t_1+t_2)/2)- \varphi_1((t_0+t_1)/2)}{(t_1+t_2)/2 - (t_0+t_1)/2} \quad (20)$$

The initial covariance P(~) is given by

$$P(\sim)=(P_0^{-1} +(H^T R^{-1} H))^{-1} \quad (21)$$

with a default covariance P_0 which represents default knowledge about possible state values which might be not observable with the available set of measurements. Without the default covariance part, the matrix inversion may lead to non realistic values.

3 Integration of the Heading Speed Filter

Like a common filter, the heading speed filter accepts plots from different sensors for each track and provides an estimate for speed, heading, acceleration, and turn rate. But the results are not suitable to solve multi target association problems. Therefore for multi target tracking this filter only can be used in combination with a common tracking filter which estimates the full state vector including position. The combination of the filters implies that state values of the common tracking filter are modified. These modifications however must be consistent. As an example if the speed estimate becomes modified, the filters capability to estimate acceleration is affected. Therefore the information transfer from the speed/heading filter to the complete tracking filter must comprise the full state vector (v, φ, a, ω). Another aspect is the decision, to which extend the estimate of the complete filter is replaced. A fixed rate of 75% has been selected in the examples which are discussed below.

Figure 3-1shows the result using the speed heading estimate in a tracking system. The white path with speed heading display represents the original tracking result, while the yellow path results in a 75%acceptance of the heading, speed estimate in the tracking system. The result shows a reduction of the track displacement of about 30Figure 3-1b shows the same comparison as shown in Figure3-1a for a noisy part of a straight flight path. The red path corresponds to the original tracking result. The figure shows that there is no significant reaction on the noise, if the heading filter is applied.

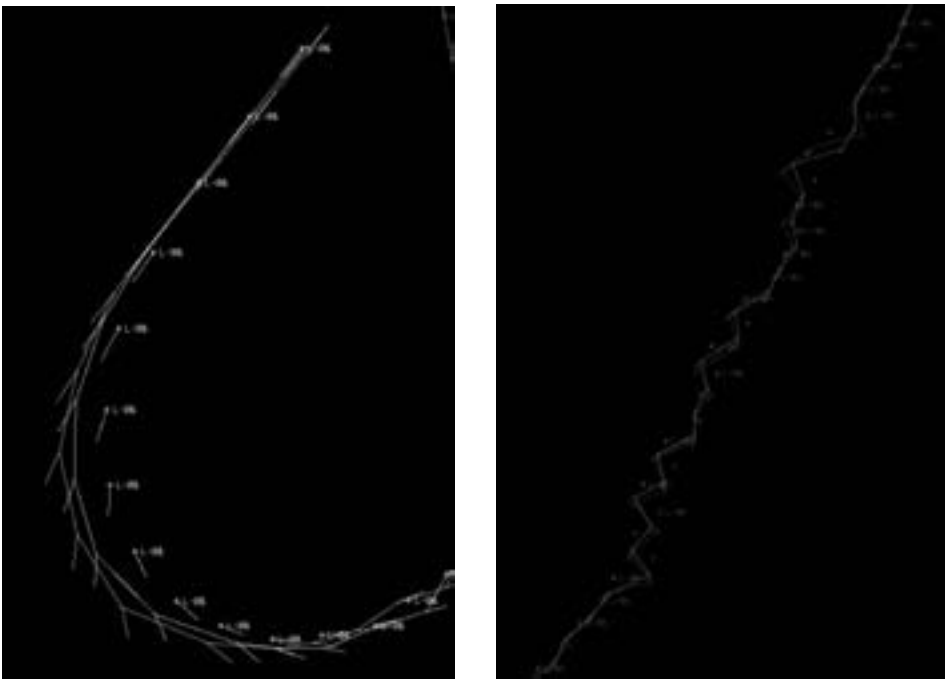


Figure 3-1 a Maneuver and b StraightTracking, one Sensor

4 Conclusion

The solution presented in this paper addresses the problem of multi sensor tracking taking into account sensor registration problems. A standard method to avoid sensor registration problems is to apply single sensor tracking and track fusion. A more stable method is direct sensor data fusion supported by automated and highly accurate sensor registration. But due to still present residual systematic sensor errors a stable tracking requires to underestimate the sensor accuracy. This restriction could be removed with the new method presented in this paper.
A new supplementary multi sensor tracking method using single sensor plot tuples as differential measurements has been presented. The new method estimates speed, heading, acceleration and turn rate of targets. With simulated data, it could be shown that the method is appropriate to generate estimation values which represent the given trajectory values accurately. Applied to tracking with recorded live data, it could be demonstrated that for maneuvering targets the tracking performance can be enhanced significantly, without loss of tracking stability.

Literaturverzeichnis

[1] Arthur Gelb, *Applied optimal estimation*

Connections Between Multitarget Intensity Filters and Medical Imaging

Roy L. Streit

r.streit@ieee.org ; streit@metsci.com

Abstract:

The discussion of multitarget intensity tracking filters is greatly enlivened by exploring the close connections that exist between these filters and two medical imaging methods whose clinical use is well established. These connections arise because nonhomogeneous Poisson point processes (PPPs) play a central role in both applications as models of the spatial variability of the crucial quantity of interest. In tracking applications this quantity is the expected target density per unit state space, while in medical imaging applications it is the density per unit volume of radioisotopes absorbed by the body. Understanding the many connections between these fields is potentially beneficial to both; however, the immediate payoff is in the tracking application — the connection significantly clarifies the interpretation and meaning of the multitarget state model, and it validates the multisensor intensity filter as an averaging filter.

1 Introduction

Both single and multisensor multitarget tracking intensity filters are based on PPP models of the multitarget state. The output of these filters is a Bayesian information update of the defining parameter of the PPP – its intensity function. To update the intensity is to update the PPP, hence the name, intensity filter. Their close connection with two well established medical imaging methodologies is discussed. The single sensor multitarget intensity filter is the dynamic analog of PET (positron emission tomography). It is shown here that the multisensor multitarget intensity filter is the analog of SPECT (single photon emission computed tomography).

In both PET and SPECT, the PPP model is an excellent match to the underlying physics and, moreover, both the intensity function and the realizations of the PPP are physically meaningful. In contrast, in multitarget tracking, the intensity function is physically meaningful but, as is seen below, the points of the realizations of a target PPP are not.

The paper is organized as follows. The discussion begins with the single sensor intensity filter and the tracking application. PET and the Shepp-Vardi maximum a posteriori (MAP) image estimation algorithm are then described, and the connections with the intensity filter are explored. In the multisensor case, however, the discussion is reversed, that is, the multicamera imaging method called SPECT is discussed first and subsequently the connections with the multisensor intensity filter are described.

The emphasis throughout is on the PPP model and the central role it plays in certain kinds of tracking and medical imaging applications. Understanding these imaging applications greatly improves insight into the multitarget model and intensity filtering. The connections presented here are discussed in [1].

2 Intensity Filters As Imaging Filters

The intuitive notion behind intensity filters is that the points of a PPP realization are an approximate model for the states of the various targets; however, the points of a realization of a PPP on the target state space are a poor representation of the physical reality of a multiple target state. To see this, denote the target state space by S and recall that the random number N of points (i.e., targets) in a realization of a PPP with intensity function $\lambda(x)$ on S is Poisson distributed with parameter $\mu = E[N] = \int_S \lambda(x)\, dx$. Thus,

$$p_N(n\,;\,\mu) = e^{-\mu}\frac{\mu^n}{n!} \tag{1}$$

is the probability that a PPP realization has n points in S. Suppose exactly one target is present, so the expected number of targets is $\mu = \int_S \lambda(x)\, dx = 1$. The probability that a realization of the PPP has *exactly* one point target is

$$p_N(N = 1\,;\,\mu = 1) = e^{-1} \approx 37\%\,. \tag{2}$$

Hence, 63% of all realizations have either no target or two or more targets. Evidently, realizations of the PPP seriously mismodel this simple tracking problem.

The problem worsens with increasing target count: if exactly n targets are present, then the probability that a realization has exactly n points is

$$p_N(N = n\,;\,\mu = n) = e^{-n}\, n^n \,/\, n! \approx (2\,\pi\, n)^{-1/2} \to 0, \qquad as\ n \to \infty\,. \tag{3}$$

Evidently, PPP realizations are poor models of any fixed number of real targets.

A more refined intuitive interpretation of the PPP approximation is that the crucial element of the multitarget model is the intensity function, *not* the points of the PPP realizations. The shift of perspective means that the intensity function is the physically meaningful quantity. Said another way, the concept of expectation, or ensemble average over realizations, corresponds more closely to physical target reality than do the PPP realizations.

A huge benefit comes from accepting the PPP approximation to the multiple target state – exponential numbers of assignments of measurements to targets are eliminated and replaced by the concept of target density. The PPP approximation finesses the data assignment problem by replacing it with a stochastic imaging problem, and the imaging problem is easier to solve. To see this, it is only necessary to explore the connections with PET medical imaging.

3 PET and the Single Sensor Intensity Filter

In PET imaging, a short-lived radioisotope is attached to a sugar and injected into the blood stream. The sugar is subsequently metabolized. The degree of absorption varies by tissue type, but pre-cancerous cells typically show heightened metabolic activity. Body tissues are imaged by estimating the spatial density of the radioisotope. This density is directly proportional to the intensity, $\lambda(x)$, of radioisotope decay at the point x.

The radioisotope undergoes beta decay and emits positrons. Positrons encounter electrons (within a few millimeters), and they annihilate. Annihilation events emit pairs of (gamma) photons that move in opposite directions. Due to conservation of momentum, these directions are essentially collinear if the positrons and electrons are effectively zero velocity.

The raw measurement data are the arrival times of photons at an array of detectors that comprise scintillator crystals and photomultipliers. A pair of photons arriving within a short time window (measured in nanoseconds) at two appropriately sited detectors determine that an annihilation event occurred: the event lies on the chord connecting the detectors, and the specific location on the chord is determined by the time difference of arrivals. Photons without partners within the time window are discarded.

A form of the "at most one measurement per target" assignment rule for tracking also holds in PET because of the basic physics – every positron-electron annihilation generates at most one measurement. Thus, it is not fortuitous that the PET imaging problem is essentially the same problem that arises in intensity filtering, albeit without the dynamic time evolution concerns of tracking.

The tissue being imaged is represented by a grid of pixels (or voxels). The spatial distribution of radioisotope is assumed to be constant in each pixel. The mean number of annihilations occurring in each pixel is estimated from the measured data, and these pixel-level estimates constitute the image. The MAP estimator is the Shepp-Vardi algorithm (1982). It is an iterative algorithm derived using Expectation-Maximization. (An identical algorithm for image deconvolution dates to 1972/1974. This algorithm is known in the larger image processing community as the Richardson-Lucy algorithm.)

The limiting case of the Shepp-Vardi algorithm as the pixel size goes to zero is of interest here. Let $\lambda(x)$ denote the spatial density of the radioisotope at the point $x \in S$, where S is in this case the tissue volume being imaged. An initial intensity function $\lambda^{(0)}(x) > 0$ is given. Assuming that no positron-electron annihilations are missed, the small pixel limit yields a recursion whose first iteration, initialized appropriately, is identical to the information update of the detected target intensity.

A simple but extraordinarily insightful interpretation of the Shepp-Vardi algorithm is adapted to give the information updated detected target intensity filter.

4 SPECT and the Multisensor Intensity Filter

In SPECT, a radioisotope is introduced into the body. As the isotope decays, gamma photons are emitted in all directions. A gamma camera takes a two-dimensional "snapshot" of the photons emitted in the direction of the camera. The camera is then moved to a fixed number of different viewing positions around the body, and a snapshot is taken in each position. These snapshots are used to construct the SPECT image, namely, the estimated intensity of radioisotope decay within the three dimensional volume of the imaged object.

The three-dimensional estimate of the spatial distribution of the radioisotope based on two dimensional data from a *single* camera lacks depth information. Surprisingly, the MAP estimate of the radioisotope intensity is the sum of the single-camera intensity estimates. The single-camera intensity estimates are added, not multiplied, because of the physics.

Connecting SPECT to the multisensor intensity filter requires noticing that the different camera positions are conceptually equivalent to different sensors. The multisensor multitarget tracking intensity filter is derived under the reasonable assumption that the PPP multiple target model holds regardless of the number of sensors; in SPECT, this means that all the cameras see the same radioisotope distribution. Because of the PPP model, the various sensors are *independent* not conditionally independent. In SPECT, this property follows from the physics. The analogy is faulty in one regard only — in SPECT the camera-level intensity estimates are added, but in the multisensor intensity filter they are averaged. Averaging is justified by a maximum likelihood argument. Averaging also reduces the variance of the estimated target count by a factor equal to the number of sensors. The variance reduction property of the multisensor intensity filter is potentially important, especially for large sensor fields.

5 Concluding Remarks

Intimate connections between single and multiple sensor intensity filters for multitarget tracking and medical imaging methods arise because PPPs are used to model the crucial quantities of interest. This quantity is the target density in the case of the intensity filters, and it is the radioisotope distribution in the case of PET and SPECT imaging. These connections greatly clarify the meaning of the multitarget PPP model. This improved understanding, and the connection of the multisensor intensity filter to SPECT, presented here for the first time, explains simply why the multisensor intensity filter is an average of the sensor-level intensity filters. Averaging is significant because it reduces the variance of the multisensor fused intensity function.

References

[1] R. L. Streit, *Poisson Point Processes: Imaging, Tracking, and Sensing*, Springer, New York, 2010.

Knowledge-aided Multi-Sensor Data Processing for Maritime Surveillance

Giulia Battistello, Wolfgang Koch

Sensor Data and Information Fusion
Fraunhofer Institute for Communication, Information Processing and Ergonomics
Neuenahrer Straße 20
53343 Wachtberg
giulia.battistello@fkie.fraunhofer.de
wolfgang.koch@fkie.fraunhofer.de

Abstract: Maritime Situational Awareness (MSA) is founded on the collection and processing of information from heterogeneous sources, including radars, Navigation Aids (Vessel Traffic Monitoring and Information Systems, Automatic Identification System, Long Range Identification and Tracking System, etc), air-based and space-based monitoring services (Earth Observation), and recently-conceived passive sensors for maritime surveillance (SIGINT, passive HF/VHF/X band radars). Several strategies for optimally fusing two or more of these information data flows have been proposed for MSA applications. However, the exploitation of context information – i.e. the so-called Knowledge Base – has not yet been thoroughly formulated. This paper provides a preliminary overview of the potential strategies for "knowledge-based" Maritime Multi-Sensor Data Fusion.

1 Introduction

The necessity of tracking the elements of a dynamic system usually requires combining information from heterogeneous and complementary data sources in order to overcome the limitations of each sensor. The gathered information might be related to the target status (position, kinematics), its physical features (shape, size, composition) or intentions (route plan, friend/foe, engaged sensor modes, etc). The combination of such heterogeneous information proved to benefit from the availability of *context* information: static and dynamic features of the scenario represent the Knowledge Base (KB) for efficient multi-sensor data processing. Data fusion processing is commonly described by the Joint Directors of Laboratories (JDL) Model ([HM04]) that takes into account the possible architectures (centralized versus autonomous) and the level of inference achieved for the sensed targets (Level 1 to 5). In the MSA frame the exploitation of the *knowledge* is proposed for multi-sensor tracking (JDL Level 1) to multi-sensor data interpretation for target intention prediction (JDL Level 2) and anomaly detection.

For ground moving target tracking (GMTI) applications ([Kl04]), *a-priori* geographic information has been efficiently encapsulated in the tracking process: road information automatically provides constraints to the prediction of target motion by a single sensor or multiple cooperative sensors. Performance is thus significantly enhanced.

Knowledge based techniques are potentially valuable in maritime surveillance applications, as well. Preliminary results are reported in [Ve08] and [BCF06]. Maritime vessel dynamics largely differ from GMTI targets: most maritime targets exhibit slow quasi linear trajectories (bounded to known origin and destination) while only few low-RCS, rapidly moving, bounded-free boats stem out from the "*normal*" scenario (e.g., conventional sea routes). Scenario knowledge is thus to be built up from static and dynamic factors in order to support the multi sensors data processing at different levels.

2 Maritime Surveillance Sensors

Maritime traffic monitoring is based on the exploitation of cooperative and non cooperative systems. In compliance with the International Maritime Organization (IMO) regulation [IMO02], ships (300 gross tonnage and upwards in international voyages, 500 and upwards for cargoes not in international waters and passenger vessels) are obliged to be fitted with **Automatic Identification System** (AIS) equipment. AIS will also be required for fishing vessels with a length of more than 15 m and sailing in water under the jurisdiction of Member States of the European Union [EU09]. AIS is based on broadcasting over VHF channels of GNSS-based position reports from cooperative vessels. This system is usually augmented with scenario-sensing sensors (radars, cameras, electro-optical sensors). Radars potentially detect all vessels in the covered area, such as incoherent X/S-band coastal radars – adopted in the frame of **Vessel Traffic Services** (VTS) implementation. The traffic picture provided jointly by AIS and VTS radars covers the territorial waters (up to 12 NM) for law enforcement.

Recently, the spreading of transnational and illegal activities led to the investigation of wide area surveillance means in order to increment early warning capabilities ([IMO06]). Systems such as **Long Range Identification and Tracking** (LRIT), **Satellite based AIS** (Sat-AIS), Satellite based **Earth Observation** and **Airborne Remote Sensing** spread in the latest years and are currently operated by national and international maritime authorities. Finally, the idea of fusing active sensing outputs with the information collected by **passive sensors** has been proposed. It is clear that one of the major achievements to be sought for in future MSA systems is the effective integration of all data sources. However, the heterogeneity of the available data sources poses stringent requirements on data fusion algorithms. Ancillary information or higher level of abstraction are to be used in order to combine information "bits" with common origin but different in nature, time and/or space. In this perspective the Base of Knowledge has a significant role.

3 Maritime Knowledge Based Data Fusion & Tracking

JDL Level 1 (Object Refinement) for maritime data fusion has been widely addressed from theoretical and implementation points of view, ([DL06], [VSB08]). Addressed data sources are mostly coastal radars, AIS and satellite-based earth observation, while the extension to all the above-described sources is still under investigation.

So far, decentralized or autonomous data fusion architectures have been considered, due to the operational constraints and unavailability of raw datasets at central level. Theoretical advantages of centralized approaches – used in multi-sensor tracking – are still to be demonstrated for MSA, to the Authors knowledge. Moreover, hybrid data fusion architectures have been developed in [CC10], where the *track-break-fuse* solution exploits both measurements and track information in a multiple hypotheses frame.

Several "un-sensed" elements have influence on the evolution of the maritime traffic; these represent sources of *a-priori* information that can be used in target inference. This information can be collected either statically (once before the observation period) or dynamically (before each traffic picture update). Examples of a-priori information are road maps along which moving targets are bounded. This knowledge is effectively used in GMTI applications ([UK06]). Following the considerations reported by the Authors in [Ve08], the sources of a-priori information for MSA could be ports, the coastline, sea corridors, Aid to Navigation and declared/preferred routes. Also target dimensions, target behavioural models and available maritime target signatures could contribute to the base of knowledge. Different degrees of inference might exploit different KBs.

KB exploitation for enhanced tracking has been formulated mainly for ground moving targets. Few examples do exist for maritime surveillance. In [DL06], littoral target tracking is aided by the knowledge on geographical maps in order to reduce false tracks and speed up true track confirmation. In [Ve08], the Authors propose an innovative approach: a-priori information dynamically concurs in creating spatial maps of "potential". Vector forces affect locally the target motion as in a force field (Navigation Field). In this research work the navigation field-based approach is encapsulated in the mathematical models presented in [UK06] and [SC02], where knowledge based and constrained tracking filters for moving targets are described. A set of "atomic" maritime scenarios will be presented: the performance improvement for Knowledge-based Level 1 data fusion is under evaluation. The heterogeneity of the presented scenarios clearly stresses the potential added value offered by KB processing. Similar considerations can be drawn for Level 2 to Level 5 processing.

4 Conclusions

This paper introduces the framework for enhanced multi-sensor data processing for MSA applications. The data fusion problem is presented with reference to state-of-the-art data models and sensors features. Knowledge-based algorithms are then assessed in the frame of common data fusion architectures and MSA reference scenarios. Results of the KB exploitation on simulated scenarios will be further shown.

References

[HM04] Hall, D.L.; McMullen, S.A.H.: Mathematical Techniques in Multisensor Data Fusion, Artech House, 2004.

[Kl04] Klemm, R.: Applications of Space-Time Adaptive Processing (IEE Radar, Sonar, Navigation and Avionics), Klemm, 2004.
[Ve08] Vespe, M. et.al.: Maritime Multi-Sensor Data Association Based on Geographic and Navigational Knowledge: Proc. of IEEE Radar Conference, 2008.
[BCF06] Benavoli, A.; Chisci, L.; Farina, A. et al.: Knowledge-Based System for Multi-Target Tracking in a Littoral Environment, IEEE Transaction on Aerospace and Electronics Systems, Vol. 42, Issue 3, pp. 1100–1119, 2006.
[IMO02] International Maritime Organisation, SOLAS convention, Chapter V, Regulation 19, 2002.
[EU09] Official Journal of the European Union, Directive 2009/17/EC of the European Parliament and of the Council, 23 April 2009.
[IMO06] International Maritime Organisation, SOLAS convention, Chapter V, Regulation 19, 2006.
[DL06] Di Lallo, A. et.al: Real Time Test Bed for 2D and 3D Multi-Radar Tracking and Data Fusion with Application to Border Control, Proc. of CIE International Conference of Radar, 2006.
[VSB08] Vespe, M.; Sciotti, M.; Battistello, G.: Multi-Sensor Autonomous Tracking for Maritime Surveillance, Proc of International Conference on Radar, 2008.
[CC10] Carthel, C.; Coraluppi, S.; Bryan, K.; Arcieri, G.: Wide-Area Feature-Aided Tracking with Intermittent Multi-Sensor Data, Proc. of SPIE Defense, Security, and Sensing, 2010.
[Ko08] Koch, W.: On Bayesian Tracking and Data Fusion: a Tutoruial Introduction with Examples, 2008.
[UK06] Ulmke, M.: Koch, W.: Road-Map Assisted Ground Moving Target Tracking, IEEE Transaction on Aerospace and Electronics Systems, Vol. 42, Issue 4, pp. 1264–1274, 2006.
[SC02] Simon, D.; Chia, T.L.: Kalman Filtering with State Equality Constraints, IEEE Transactions on Aerospace and Electronic Systems, Vol. 39, pp. 128-136, January 2002.

Benefits of a Multi-GNSS Receiver in an Interference Environment

Ulrich Engel
Fraunhofer Institute for Communication, Information Processing and Ergonomics FKIE
Department Sensor Data and Information Fusion (SDF)
Neuenahrer Str. 20, 53343 Wachtberg, Germany
ulrich.engel@fkie.fraunhofer.de

Abstract: In this paper, we investigate the benefits of a MULTI-GNSS receiver in an interference environment. The global satellite navigation systems GPS, GALILEO, GLONASS and COMPASS will be considered. The regional navigation systems QZSS and IRNSS are not part of the investigation. We assume that a future Multi-GNSS receiver will use only CDMA signals. Therefore, the FDMA signals of the GLONASS system are not included in the simulations. Precise almanac data are used to simulate the satellite constellations of the GNSS systems. The link budget for the received signal power includes all major path losses and antenna gain factors. Finally, the results of the interference calculations are presented for the Earth surface.

1 Introduction

Currently, most civil GNSS receivers use the C/A-Code for navigation. However, during the design phase of GALILEO, COMPASS and the modernization plans of the GLONASS system interoperability[1] and compatibility[2] have been important issues. With more navigation systems and interoperable and compatible signals soon available, it should be beneficial to use these additional information for navigation.
We investigate the benefits of such a Multi-GNSS receiver in an interference environment. In our analyses we will concentrate on CDMA signals which are open for civil users. GPS is introducing three new civil signals L1C, L2C and L5. Most of the GALILEO signals on E1, E6 and E5 will be open signals. COMPASS will have only two open signals which are similar to the GALILEO system. And GLONASS introduces also two new CDMA signals on L1 and L5. We will simulate a realistic interference szenario and compare the results for Single- and Multi-GNSS receivers. The influence of the atmosphere, the gain pattern of the satellite and receiver antenna, the transmitted satellite power, the polarisation mismatch, the degradation due to the interference signal and the free-space-path loss will be taken into account.

[1] *Interoperability refers to the ability of civil space-based PNT services to be used together to provide better capabilities at the user level than would be achieved by relying solely on one service or signal.*

[2] *Compatibility refers to the ability of space-based PNT services to be used separately or together without interfering with each individual service or signal, and without adversely affecting navigation warfare.*

2 Link budget

The link budget between the satellite and the receiver is an important factor in the simulations. The exact knowledge of link gains and losses is necessary to determine the received signal power levels at the receicer, which is calculated by equation (1). In the following simulations we assume that the receiver is always located on the Earth surface.

$$C_i = P_{sat,i} + G_{sat,i} - L_{dist} - L_{trop} - L_{iono} - L_{pola} + G_{rec} \tag{1}$$

with i = satellite index, C = received signal power, P_{sat} = Transmitted satellite power, G_{sat} = satellite antenna gain, L_{dist} = free-space-path loss, L_{trop} = tropospheric loss, L_{iono} = ionospheric loss, L_{pola} = polarisation loss and G_{rec} = receiver antenna gain.

A high-level diagram of the simulation toolkit, which has been developed for this analysis, is shown in Figure 1. The main functions are all written in C/C++. Numerical complex functions have been implemented in Matlab.

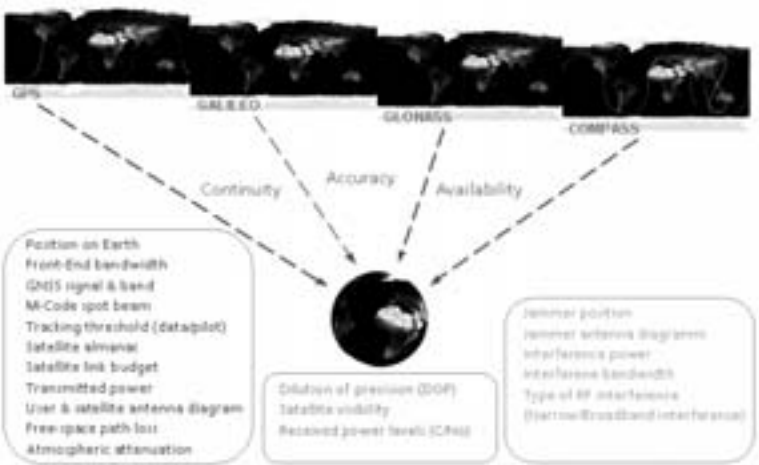


Figure 1: Simulation Toolkit "Multi-GNSS Engine"

3 Interference Calculations

The interference calculations between the satellite signals and the interference signal are based on [3]. We do not consider multiple interference signals in this paper. Then, the effective carrier-to-noise-density ratio $C_s/N_{o,eff}$ can be estimated by equation (2) [9]:

$$(C_s/N_o)_{eff} = \frac{1}{\frac{1}{(C_s/N_o)} + \frac{C_i/C_s}{\frac{\int |H_r(f)|^2 G_s(f) df}{\int |H_r(f)|^2 G_i(f) G_s(f) df}}} \tag{2}$$

with $H_r(f)$ = filter transfer function at the transmitter and receiver, $G_s(f)$ = power density function of desired signal and $G_i(f)$ = power density function of interference. The expression in the denominator is known as the spectral seperation coefficient (SSC).

$$\text{SSC} = \int |H_r(f)|^2 G_i(f) G_s(f) df \tag{3}$$

The SSC depends on the spectrum of the desired signal as well as the spectrum of the interference. From equation (2) and (3) it is obvious that the SSC must be nonzero in order to effect the received $(C_s/N_o)_{eff}$. The interference signal must pass the filter transfer function at the receiver and overlap with the spectrum of the desired signal.

4 GNSS Signal Characteristics

Each space constellation has slightly different orbital plane parameters. The most important parameters of the corresponding MEO satellites for GPS, GLONASS, GALILEO and COMPASS are summarized in Table 1. The constellation size is specified in brackets.

Table 1: Orbital plane parameters for MEO satellites

Constellation	Eccentricity	Inclination	Semi-major axis	Period
GPS (30)	0	55°	26561750 m	43082.05 s
GLONASS (24)	0	64°	25507550 m	40542.83 s
GALILEO (27)	0	56°	29993707 m	50682.21 s
COMPASS (27)	0	55°	27840000 m	46229.03 s

We consider only open signals for the simulations. The characteristics of these signals in the L1-band are described in Table 2. Important information for the following simulations are the spreading modulation, the sub-carrier frequency, the chip rate and the transmitted signal power. However, the latter is no official data. It is the result of a link calibration based on the individual minimum received signal power levels.

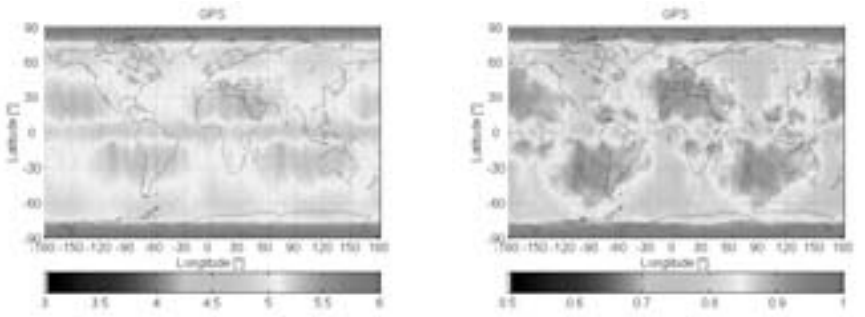
Table 2: Open signal characteristics for GPS, GLONASS, COMPASS and GALILEO in L1-Band

GNSS System	GPS		GLONASS	GALILEO	COMPASS
Service Name	C/A	L1C data	N/A	E1 OS data	B1C data
Service Type	Open	Open	Open	Open	Open
Center Frequency	1575.42 MHz				
Frequency Band	L1				
Access Technique	CDMA				
Spreading Modulation	BPSK(1)	BOC(1,1)	BOC(2,2)	MBOC(6,1,1/11)	
Sub-carrier Frequency	-	1.023 MHz	2.046 MHz	1.023 / 6.138 MHz	
Chip Rate	1.023 MHz	1.023 MHz	2.046 MHz	1.023 MHz	1.023 MHz
Signal Component	Data	Data	Data	Data	Data
Primary PRN Code Length	1023	10230	N/A	4092	N/A
Secondary PRN Code Length	-	-	-	-	-
Data / Symbol Rate	50 / N/A	N/A	N/A	N/A / 250	N/A
Minimum Received Power	-158 dBW	-162.5 dBW	-158 dBW	-160 dBW	-160 dBW
Elevation	5°	5°	5°	10°	10°
Transmitted Power	11.7 dBW	11.95 dBW	11.6 dBW	9.7 dBW	9.3 dBW

5 Simulation Results

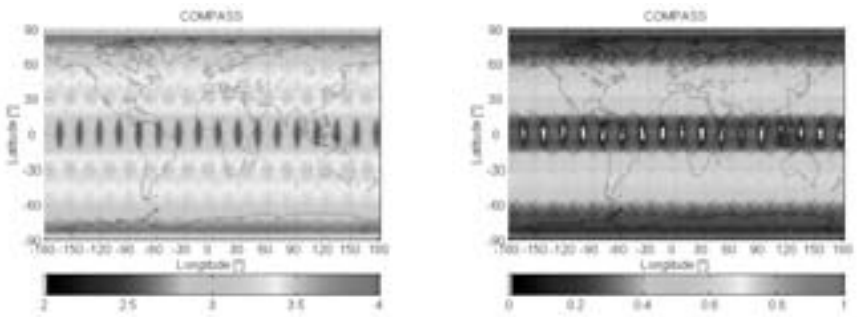
For the following computations we assume a noise density N_o = -201.0 dBW/Hz and a cable loss of $L_c = 1$ dB. The frontend bandwidth of the GNSS receiver is 10.23 MHz. The bandwidth of the bandlimited white gaussian noise interference is choosen to be 4.092 MHz. The simulation period is 1 day with a temporal resolution of 864 seconds and a spatial resolution of 1°. The minimum elevation angle is $\varphi = 5^\circ$ for all satellites. The lower bound of the tracking threshold is assumed to be $C_s/N_o = 15$ dBHz. Without interference the C_s/N_o lies in the range $35 - 50$ dBHz. We present the results for a GPS receiver using the GPS C/A-Code and a COMPASS receiver using the B1C data signal. The characteristics of the GALILEO E1 OS data signal are very similar to the COMPASS B1C data signal (see Table 2). The interference power is C_i = -105.0 dB.

5.1 Single-GNSS



(a) Global visibility (b) Probability for number of satellites ≥ 4

Figure 2: Interference computations for the GPS C/A-Code signal

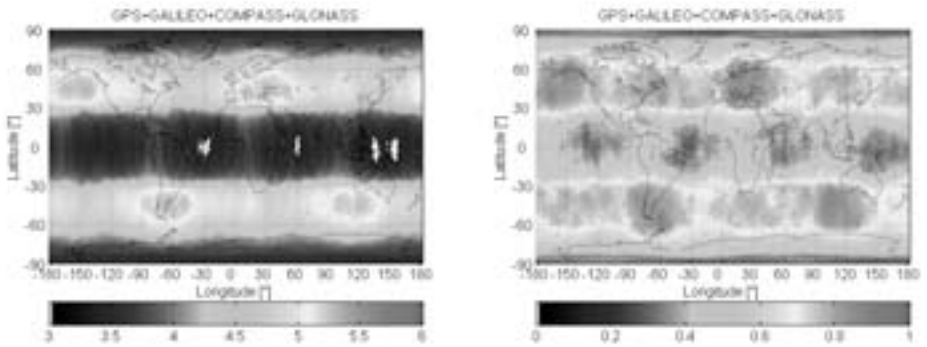


(a) Global visibility (b) Probability for number of satellites ≥ 4

Figure 3: Interference computations for the COMPASS B1C data signal

5.2 Multi-GNSS

When all global satellite navigation systems are considered the number of satellites increases and is equal to 108. As each satellite system uses different orbital planes, the satellites should be well distributed over the hemisphere. Therefore the interference power for the Multi-GNSS szenario is assumed to be C_i = -100.0 dB.



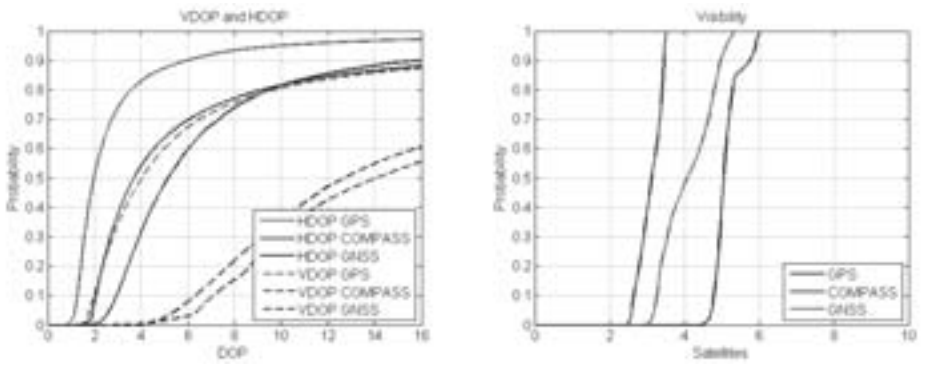
(a) Global visibility

(b) Probability for number of satellites ≥ 4

Figure 4: Interference computations for a MULTI-GNSS receiver

6 Conclusions

In this paper, we have investigated the benefits of a single-frequency (L1-Band) Multi-GNSS receiver in an interference environment. For slightly higher interference power the Multi-GNSS constellation shows similar performance with respect to satellite visibility (see Figure 5). The performance would be even better for the same interference power.
For the selected szenarios the DOP values of the Multi-GNSS receiver are significantly worse than the DOP values for GPS and COMPASS alone. Although more than 4 satellites are mostly visible, the satellite geometry constantly degrades with increased interference power. The bad geometry leads to higher DOP values as seen in Figure 5. Further investigations will continue to concentrate on the benefits of multi-frequency GNSS receivers. So far, we have assumed only a single-frequency GNSS receiver. With more frequency bands and signals available a multi-frequency GNSS receiver should be more robust against interferences (assuming equal power).



(a) Cumulative distribution function for vertical and horizontal DOP

(b) Cumulative distribution function for visibility

Figure 5: Interference computations for a MULTI-GNSS receiver

References

[1] NovAtel Inc., "GPS-704X Antenna Design and Performance", www.novatel.com.

[2] B. Rembold, "Wellenausbreitung", Institut für Hochfrequenztechnik, RWTH Aachen, 2004.

[3] E.D. Kaplan and C.J. Hegarty, *Understanding GPS Principles and Applications*. 2nd ed. Artech House, INC., 2005.

[4] J. Spilker and B. Parkinson, *Global Positioning System: Theory and Applications - Volume I*. American Institute of Aeronautics and Astronautics, Washington, 1996.

[5] S. Wallner, "Interference Computations Between Several GNSS Systems", Institute of Geodesy and Navigation, University FAF Munich, Germany.

[6] C. Wang, "An Improved Single Antenna Attitude System Based on GPS Signal Strength", Cooperative Research Centre for Satellite Systems, Queensland University of Technology, Australia.

[7] J. Rodriguez and W. Hein and S. Wallner and J. Issler and L. Ries and L. Lestarquit and A. Latour and J. Godet and F. Bastide and T. Pratt and J. Owen, "The MBOC Modulation - A Final Touch for the Galileo Frequency and Signal Plan", InsideGNSS, www.insidegnss.com.

[8] Y. Urlichich, "Russia Approves CDMA Signals for GLONASS - Discussing Common Signal Design", Moscow International Satellite Navigation Forum, www.insidegnss.com.

[9] J.P. Saulay, "STANAG 4665 - Navigation Warfare Operational Planning and Management", Navigation Warfare ad hoc Working Group.

[10] ARINC Research Corporation, *Navstar GPS Space Segment/Navigation User Interfaces*. ICD-GPS-200, Revision C, October 1993.

Position Tracking in Urban Environments using Linear Constraints and Bias Pseudo Measurements

Julia Niewiejska, Felix Govaers, Nils Aschenbruck
University of Bonn - Institute of Computer Science 4
Roemerstr. 164, 53117 Bonn, Germany
{niewiejs, govaers, aschenbruck}@cs.uni-bonn.de
Fraunhofer FKIE, Neuenahrer Str. 20, 53343 Wachtberg, Germany
felix.govaers@fkie.fraunhofer.de

Abstract:
In many GPS-Sensor based tracking applications, the obtained measurements suffer from time a correlated bias. This is due to shadowing and multipath scattering of the wireless GPS signals. In this paper, we applicate a *Schmidt-Kalman Filter* (SKF) in order to improve the tracking process of ground vehicles on roads. We investigate possibilities to integrate bias measurements obtained from road information into the position tracking filter. To this end, we assume a digital map is given which contains road information for the observed region. The estimated position by the sensor data is projected onto the road as a hard constraint. This enables us to detect and to eliminate regular sensor bias by extending the estimate state of the target by an estimate of the sensor bias.

1 Introduction

Nowadays, there are a vast amount of tracking applications based on GPS sensors, as the sensors have become available for low cost and the GPS system is freely available everywhere around the world. However, the accuracy of a position depends on the number of satellites the sensor can receive. Moreover, shadowing and multipath scattering influence the position measurements significantly. In particular, in urban environments or in canyons, the shadowing effect leads to a time correlated *sensor bias* [9, 10].

In this paper, we focus on tracking ground moving target (such as vehicles or pedestrian in an urban area) using road information. The goal is to incorporate additional information in order to improve the track. Assuming perfect detection, linear dynamics and sensor model, and zero mean Gaussian error distributions, the Kalman filter is optimal in the mean squared error sense. In order to obtain a consistent filter, the dynamic model and properties of the sensors must be modeled appropriately. In our approach, we process *pseudo measurements* in order to estimate the current bias of the sensor. This work presents the structure and the parametrization for this virtual sensor.

Assume there is a target moving on roads of which we have digital road maps available. Then, it is possible to obtain a more accurate estimate by extending a Kalman filter as de-

scribed in the the following steps. (1) Filter the position measurement in the Kalman filter. (2) Search the next road segment best matching the current filtering output and project the position onto that segment. (3) Obtain a sensor bias measurement by calculating the line connecting the actual sensor position measurement and the projected position. (4) Let the Kalman filter process this bias measurement in order to estimate the current bias.
By doing so, shadowing effects on the wireless GPS signal are much better compensated by the Kalman filter for the next position data coming in.

2 State Estimation with Linear Constraints

In order to restrict the space where a target is possibly moving to roads, one can use linear constraints. Those are straight lines that represent segments of a road. There are a couple of possibilities to integrate linear constraints with the state estimation: (1) Use the constraints directly in the estimation process. The dynamic model is projected onto the constraints and then used as input for the Kalman filter. (2) Consider the linear constraints as pseudo measurement. (3) *Track-to-road fusion* (T2RF)- the measurement data is filtered without any constraints, then the output is constrained. T2RF is the methodology we use in this paper.

2.1 Road Approximation

Roads are usually not ideally straight lines. They need to be approximated before they can be used as linear constraints. A straight road segment can be directly represented in Hesse Normal Form as a point (r, θ) on the plane. Here, r represents the distance to point of origin, while θ stands for the angle to the x-axis. Roads with curves must first be divided into segments which are then represented as lines as well. Then, a curve consists of a series of straight segments that approximate it.

The road segment closest to the estimated state is relevant for the linear constraint construction. Let n be the normal vector of this segment, then

$$n = [\cos\theta, \sin\theta]^\top \tag{1}$$

Moreover, let $p = [x, y]^\top$ be the position of the target and v its speed on the line. v is perpendicular to n. Thus, the following statement holds according the definition of the dot product: $v^\top n = 0,$ and $p^\top n = r$ due to $p^\top n = [x, y] \cdot \begin{bmatrix}\cos\theta \\ \sin\theta\end{bmatrix} = x \cdot \cos\theta + y \cdot \sin\theta.$ Combining those conditions with the line equation in hough space results in: $D\mathrm{x}_k = d$ where $D = \begin{bmatrix}\cos\theta & \sin\theta & 0 & 0 \\ 0 & 0 & \cos\theta & \sin\theta\end{bmatrix}$, and $d = \begin{bmatrix}r \\ 0\end{bmatrix}$. The state of the object is given by $\mathrm{x}_k = [x, y, \dot{x}, \dot{y}]^\top$. Both variables may be time dependent.

2.2 Projection on a Road

As previously mentioned, the projection of a state on the road is an optimization problem with constraints. The constraints from previous subsection can now be used on the output of the Kalman filter. The road equation is interpreted as a surface $\mathcal{S} = \{\mathrm{x} : D\mathrm{x} = d\}$, containing possible positions of the target, on which the state estimate $\hat{\mathrm{x}}$ will be projected. The proper point can be calculated in different ways depending on the definition of the proximity. Simple metrics such as the euclidean distance may be used. It is also possible to take more information into account, e.g. direction and speed of the target's movement. In the second case a symmetric positive-definite weighting matrix W would be used to define the proximity. The optimization problem is now as follows:

$$\breve{\mathrm{x}} = \mathrm{argmin}_{\mathrm{x}\in\mathcal{S}}(\mathrm{x} - \hat{\mathrm{x}})^\top W(\mathrm{x} - \hat{\mathrm{x}}) \tag{2}$$

The squared euclidean distance $(\mathrm{x} - \hat{\mathrm{x}})(\mathrm{x} - \hat{\mathrm{x}})$ is weighted with matrix W. If $W = I$ only the euclidean distance will be considered. The use of the covariance matrix P of the estimation error for weighting is highly recommended. If $W = P^{-1}$ then the covariance of the constrained result is smaller than the covariance of the unconstrained one.

The optimization equation above is easily solvable with a method called Lagrange multiplier. First, the Lagrange function is constructed:

$$\Lambda(\mathrm{x}, \lambda) = (\mathrm{x} - \hat{\mathrm{x}})^\top W(\mathrm{x} - \hat{\mathrm{x}}) + 2\lambda^\top(D\mathrm{x} - d) \tag{3}$$

Solving gradient equation $\nabla_{\mathrm{x},\lambda}\Lambda(\mathrm{x}, \lambda) = 0$ results in:

$$W(\mathrm{x} - \hat{\mathrm{x}}) + D^\top\lambda = 0 \tag{4}$$

$$D\mathrm{x} - d = 0 \tag{5}$$

After some calculations the resulting state projected on the constraints is as follows:

$$\breve{\mathrm{x}} = \hat{\mathrm{x}} - W^{-1}D^\top(DW^{-1}D^\top)^{-1}(D\hat{\mathrm{x}} - d) \tag{6}$$

3 Filtering with Bias

Filtering with linear constraints provides an improvement of the state estimation if the target moves on roads. The output is a trace constrained by roads that are approximated as lines. However, the accuracy of the road projection depends to a great extent on the output of the measurement filtering in the Kalman filter. This step has a potential of improvement. In many applications, sensors are subject to time correlated bias. To overcome this challenge, one might use the Schmidt Kalman filter [1], which incorporates bias into the Kalman filter.

3.1 Bias Processing

When the filtering of a sensor measurement has been finished, the output can be projected on the road, as described in the previous chapter. The bias can be interpreted as the difference between measurement and the result of the projection. Nevertheless, the bias calculated in such a way suffers the same effects of inaccuracy as the sensor measurement. Therefore, after the bias has been *measured*, it will be filtered in a similar way to the one above. To do this, some changes in the filter have to be made.

Let $\breve{\mathrm{x}}_k^b$ be the output $\hat{\mathrm{x}}_k^b$ of the filter projected onto the approximated road according to section 2.2. Then the bias measurement is defined as follows:

$$\mathrm{z}_k^b = \mathrm{z}_k - \hat{\mathrm{x}}_k^b \tag{7}$$

with z_k being the current sensor position measurement. As only the bias from the target's state is needed, the measurement matrix must be modified:

$$H_k^b = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}. \tag{8}$$

Then, the measurement error covariance is changed to the value γ_b that is appropriate for the bias:

$$R_k^b = \begin{bmatrix} \gamma_b & 0 \\ 0 & \gamma_b \end{bmatrix}. \tag{9}$$

Moreover, with each road projection, the covariance of the filtering estimate changes by the squared distance between the current and the last output. This should be taken into consideration:

$$\breve{P}_k^b = \hat{P}_k^b + \Delta\mathrm{x}\Delta\mathrm{x}^\top, \tag{10}$$

with $\Delta\mathrm{x} = \breve{\mathrm{x}}_k^b - \hat{\mathrm{x}}_k^b$. Now, the bias filtering can be carried out with the modified values. To this end, the bias is interpreted as a measurement. The result is a new bias extended state that is more appropriate, as several additional aspects, e.g. target's speed and direction, is taken into account. After that, the changed values, such as measurement matrix and measurement error covariance, should be set back to ensure proper sensor measurement prediction and filtering in the next time step.

4 Evaluation

In this section, exemplary evaluation results of an implementation of the Kalman filtering with bias and linear constraints are presented. The input is a series of time stamped positions of the target measured with a standard GPS tracker. As background information, the



(a) (b)

Figure 1: (a) Output of the Kalman filter with linear road constraints. The red line is the output of the Kalman filter, the green line shows the result of projecting this output on roads. Where no projection is needed the result of the Kalman filter is taken as is. (b) Result of the Kalman filter with bias and linear road constraints. The green line represents the final output, the red parts are present at places where filtering output with bias does not match the final output.

roads were not given as a set of curves or straight lines. But detailed data on buildings was available. Thus, the roads were approximated from the surrounding buildings' walls. The traced target moves in walking speed and strictly outside of the buildings. Nevertheless, the noise and other factors cause the trace points being placed at least half of the time inside the buildings. It appears as if the target would move through walls.

Our goal was to eliminate this effect. Thus, we applied the Kalman filter with linear constraints on the trace. The result is shown in Fig. 1 (a). The Kalman Filter only smoothes the trace completely ignoring the building data. Then, each trace point is projected onto next road without any feedback for the filter. This leads to straight, unnatural looking trace segments on roads where the projection was used (which was the case most of the time). A total of 334 points had to be projected. It is possible to reduce the frequency of the projection, if the filtering output is more appropriate and if the bias is taken into account. Fig. 1 (b) shows the results of the method described in this paper. The Kalman filter with bias and linear road constraints.

The trace looks now much more natural and there are only a few places (marked red) where the output of the Kalman filter does not match the final projected output of the whole algorithm. These are e.g. places where a filtered trace point lies inside a building. In total there were 46 projections needed which is a significant improvement to the 334 in the case where no bias has been used.

5 Conclusion

In this paper, we presented an improved position tracking scheme for urban environments using road map information. Due to shadowing and scattering, such circumstances often lead to time correlated measurement bias when GPS sensors are utilized. Therefore, it is a great benefit to use a Schmidt-Kalman filter which estimates the sensor bias beneath the target's state. Moreover, background information such as road maps might be included into the filtering process. This yields a reliable estimate for the current sensor bias. Combining both approaches leads to a significant reduction of the amount of states that have to be projected onto the road. This was confirmed by an evaluation using a pedestrian trace of a GPS sensor and a city map picture. In the future, we plan to explore other techniques for the linear constraint filtering and their extension to the bias extended Schmidt-Kalman filter.

References

[1] S. Schmidt, *Applications of State Space Methods to navigation Problems*, Advances in Control Systems (C. T. Leondes), Vol. 3, Academic Press, 1966.

[2] Y. Bar-Shalom and X. R. Li,*Multitarget-Multisensor Tracking: Principles and Techniques*, YBS Publishing, 1995.

[3] Y. Bar-Shalom, X.-R. Li, and T. Kirubarajan, *Estimation with Applications to Tracking and Navigation,* Wiley & Sons, 2001.

[4] S.S. Blackman and R. Popoli, *Design and Analysis of Modern Tracking Systems*, Artech House, 1999.

[5] R. Paffenroth et al., *Mitigation of Biases Using the Schmidt-Kalman filter*, Proceedings of the SPIE Conf. Signal and Data Processing of Small Targets, Vol. 6699, San Diego, CA, Aug. 2007.

[6] Novoselov, R.Y.; Herman, S.M.; Gadaleta, S.M.; Poore, A.B., *Mitigating the effects of residual biases with Schmidt-Kalman Filtering*, Information Fusion, 2005 8th International Conference on , vol.1, no., pp. 8 pp., 25-28 July 2005.

[7] Chun Yang; Blasch, E., *Track Fusion with Road Constraints*, Proceedings of the 10th International Conference on Information Fusion, pp.1-8, 9-12 July 2007.

[8] Koch, W., *Information fusion aspects related to GMTI convoy tracking*, Proceedings of the Fifth International Conference on Information Fusion, vol.2, pp. 1038-1045, 2002.

[9] Youjing Cui; Shuzhi Sam Ge, *Autonomous vehicle positioning with GPS in urban canyon environments*, IEEE Transactions on Robotics and Automation, vol.19, no.1, pp. 15-25, 2003.

[10] Pink, O.; Hummel, B., *A statistical approach to map matching using road network geometry, topology and vehicular motion constraints*, Proceedings of the 11th International IEEE Conference on Intelligent Transportation Systems, pp.862-867, 2008.

Narrowband Passive Sonar Tracking

Kevin Brinkmann and Jörg Hurka
ATLAS ELEKTRONIK GmbH, Sebaldsbrücker Heerstr. 235, 28309 Bremen
Email: {kevin.brinkmann, joerg.hurka}@atlas-elektronik.com

Abstract: Passive sonar sensors are the main source of information for a submerged submarine. Modern sonar systems like ISUS 90 offer a number of acoustic antennas and combined signal processing for broadband and narrowband detection and analysis of target noise. While broadband detection is often used to obtain an overview of the targets surrounding the own ship, narrowband processing enables the detection of target characteristic frequency lines, often produced by vibrations of the propulsion systems on surface ships which are radiated into the water. Narrowband processing is therefore essential for target analysis and classification. Frequency line information can be used to separate targets closely spaced in bearing, e.g. during target crossing situations, which can not be resolved by the broadband passive sonar information alone. In addition, target bearing histories containing frequency line information may be used in the Target Motion Analysis module to infer about target course, speed and range without an own-boat manoeuvre. Therefore, narrowband passive sonar tracking provides an additional valuable source of information and enhances the capabilities of a submarine sonar and combat system. In this paper we present multi-target tracking results for narrowband passive sonar based on a multi-hypothesis tracking approach. The tracking system is designed to automatically track all detectable frequency lines of multiple targets, while at the same time minimizing the number of false tracks.

1 Introduction

Surface ships and other naval platforms radiate broadband and narrowband noise originating from their propulsion system and other machinery into the water. A submerged submarine entirely relies on these acoustic signals for long range detection and surveillance. Because the use of active sonar leads to an increased counter detection probability, submariners usually rely on passive sonar. A typical ATLAS ELEKTRONIK ISUS 90 sonar suite for a modern conventional U 214 class submarine offers three important antenna systems for analyzing the incoming acoustic waves. The Cylindrical Hydrophone Array (CHA), located in the bow of the submarine, the Flank Array (FA) which is located at both sides of the submarine, and the Towed Array (TA), a flexible line array which can be deployed from a winch system in the rear part of the submarine. For the CHA, the FA, and the TA there are in parallel broadband and narrowband signal processing chains. While the narrowband processing is essential for target analysis and classification, broadband detection gives an overview of the targets surrounding the own ship. Apart from the Passive Ranging Sonar (PRS) not considered in the present contribution, passive sonar only provides bearing information. In order to derive target range estimates, bearing tracks of the detected targets of interest are formed. These target bearing histories are input to

the Target Motion Analysis (TMA) module which estimates course, speed and range of the target under consideration. In this way, a tactical picture of the sea area of operation can be compiled. In order to reduce the errors in the TMA solution, high quality bearing tracks entering the TMA algorithm are essential. Information like frequency lines, accessible by analyzing the narrowband processing output, ascribed to a particular bearing track could serve useful additional parameters to enhance the performance of the TMA module, namely it enables the system to derive target range estimates without the necessity of an own boat manoeuvre.

In current operational sonar systems, the initialization, maintenance, and deletion of target bearing tracks as well as tracks formed from the detected frequency lines is typically performed by the operator.

The paper in which we present a novel multi-target tracking algorithm for narrowband passive sonar based on a multi-hypothesis tracking approach is organized as follows. Section 2 reviews some of the narrowband passive sonar specific features which are of interest in the tracking context. In Section 3 we briefly describe our multi-target tracking system and the algorithms used within it. Narrowband passive sonar tracking results for the Flank Array are presented in Section 4. In Section 5, we summarize and draw conclusions.

2 Narrowband Passive Sonar

Narrowband passive sonar aims at the detection of characteristic frequency lines emitted by a target vessel. There are two related but different origins for such frequency lines: i) the propulsion systems on board a naval ship and ii) the physical effect of cavitation. Depending on the origin, it has to be distinguished between tonals, i.e. discrete frequency lines transmitted directly in the water column, and indirect frequency lines. The latter can only be detected by the application of an algorithm which searches for characteristic modulations of the broadband noise emitted by the ship.

The propulsion system of a ship, for example a diesel engine, produces cyclic repetitions of moveable machinery parts due to the operation of the cylinders. For proper functioning, the engine additionally needs a number of ancillary units and several pumps (transport of lubricants or oil). The work of the engine due to its strong coupling to the hull of the ship therefore results in vibrations of various frequencies which are transported into the water. Information on typical frequencies which can be resolved and identified by narrowband passive sonar are the cylinder firing rate, the engine firing rate (cylinder firing rate times number of cylinders) and the crank shaft rate. These frequency lines are typically found in the low frequency range.

Cavitation is the effect of gas bubble formation close to the tips and on the faces of the propeller blades of a ship. The pressure of the moving water in the layer close to the blade is severely lowered, thus allowing vaporization. The bubbles have a characteristic average life time after which they collapse. The bubble implosions can cause severe damages to the propeller blades and additionally are the source of noise generated by the propeller. The generated noise has characteristic modulation frequencies which contain information

about the number of blades and the turning rate of the propeller (propeller shaft rate).

The quantity of interest in narrowband passive sonar is the spectral content of the incoming sound from all directions, integrated over a certain interval of time. Due to the application of a beamforming operation, the incoming sound is easily separable into its directional origins. Two modes of operation can be used to analyze the sound, namely the LOFAR signal processing which is sensitive for direct frequency lines and the DEMON signal processing which has the capability to detect the indirect frequency lines. The acronym LOFAR stands for **LO**w **F**requency **A**nalysis and **R**ecording, DEMON stands for **D**etection of **E**nvelope **MO**dulation on **N**oise.

3 Multiple-Hypothesis Approach

Modern tracking systems are very often designed as multi-target tracking (MTT) systems allowing the simultaneous tracking of a great number of physical objects (a detailed introduction can be found in Ref. [BP99]). An integral part of any such system is a physical sensor, e.g., a radar or a sonar sensor. In the MTT case the sensor typically has a large field of view (FOV) which is scanned in order to process all objects in this region at a time. The sensor signal processing periodically generates lists of detections, e.g., detected radar echoes in the case of active radar, which are passed on to the tracking unit. The task of the tracking unit is the logical combination of recurring sensor detections which emerged from one particular object and the estimation of its kinematical parameters. Due to the fact that the list of detections per scan not only contains measurements from objects of interest but also false alarms this becomes a sophisticated challenge.

Under realistic conditions, the tracking system has to deal with a number of uncertainties like erroneous measurements, limited sensor resolution, imprecise knowledge of the object motion and high false alarm rates. In our proposed approach the tracking unit is composed of three interrelated main elements. One is responsible for state estimation, the other two for data association and management tasks respectively (cf. Ref. [BH09]). The state estimation block in combination with the data association block forms tentative tracks, tentative, because the tracking system will treat them as candidates for real tracks. The data association uses the multi-hypotheses approach, where multiple new detections are assigned to a particular given track and a hypothesis tree is processed in order to consider more than one data interpretation history [BP99, Koc01, BH09].

The management block has to cope with tasks like track extraction, track deletion and track maintenance. In the particular case of narrowband passive sonar as discussed in this contribution, the probability is high that a target emits more than one specific frequency. The management block has the capabilities to extract all of these lines. The tracks formed by the MHT, the multi line tracks (MLT's) hold information about all frequencies emitted by one particular target, thus, for each considered time interval it is characterized by a bearing value and one or more frequency values.

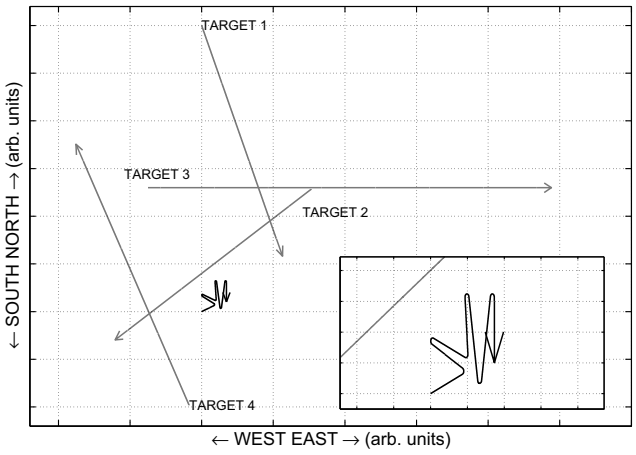


Figure 1: Geometry of the simulated scenario.

4 Simulation Example

In this section we present results of our multi-target tracking algorithm applied on a simulated scenario with a duration of 80 minutes. The scenario contains four different surface vessels traveling with constant course and speed. During the time of the scenario the own submarine performs typical TMA manoeuvres. The target trajectories are shown as red lines in Fig. 1. The inset magnifies the own submarines trajectory. The following frequency lines are given: Target 1: 400 Hz, 450 Hz, 520 Hz, 570 Hz, Target 2: 580 Hz, 620 Hz, 690 Hz, 715 Hz, Target 3: 430 Hz, 525 Hz, 685 Hz, 1000 Hz and Target 4: 867 Hz, 879 Hz, 900 Hz, 917 Hz.

The acoustic information, generated during the simulation, is fed into the sonar signal processing units responsible for extracting direct frequency lines, i.e., the LOFAR processing chain. The detection algorithm extracts possible target frequency line detections.

The list of detections for each time step is subsequently fed into the narrowband passive sonar version of the MHT which performs the estimation and data association steps described in Sec. 3. If a collection of recurring frequency line detections passes the confirmation test, a tentative MLT becomes extracted after a particular amount of time. Deletion of a MLT is initiated if it becomes very unlikely to originate from a true target.

Figure 2 presents the tracking results obtained for the FAS-LOFAR data. In this case the results showing the bearings of the extracted MLT's. These are drawn as blue lines with the respective track number displayed at the beginning of each MLT. The green dots represent bearings of the frequency lines extracted by the detection algorithm. The own course is drawn in black. Each detection occurring in the blind zone of the acoustic sensor is highlighted in magenta. It can be seen that all the MLT's formed by the narrowband version of the MHT originate from one of the four vessels present in the simulated scenario. As

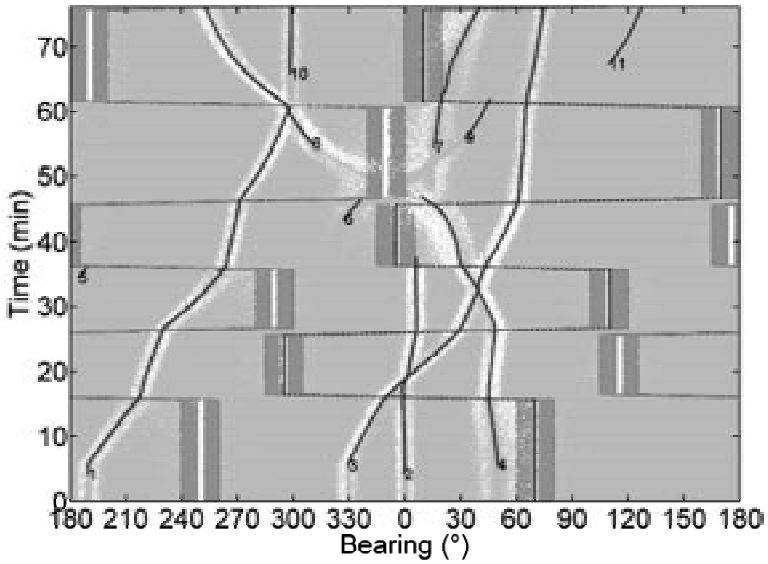


Figure 2: Tracking results showing the bearing of extracted MLT's.

it was the case for the broadband passive sonar results [BH09], the trajectory for one particular target usually consists of several extracted MLT's. However, notice that for the narrowband processing a break in the extracted tracks does not necessarily occur during target crossing situations but mainly due to the imperfections of the sensor. For example, the trajectory of target 4 is composed of MLT's no. 1 and no. 10. The trajectory for target 1 is composed of the MLT's no. 2 and no. 7 and the one for target 2 of MLT's no. 4 and no. 8 respectively. The trajectory of target 3 is build up by just one extracted MLT (no. 3). Being a linear array, the Flank Array in direct frequency line detection mode also leads to the extraction of mirror targets, i.e., the MLT's no. 5, no. 6, no. 9 and no. 11, all of which are mirrors from the target trajectory of vessel 2.

Figure 3 shows the tracking results for the FAS-LOFAR data, especially the frequencies of the extracted MLT's. The frequency lines with the respective numbers of the MLT to which they belong to are shown as blue lines. Because a particular MLT contains more than one frequency line, in this figure specific track numbers occur more than once. The mirror targets (MLT's no. 5, no. 6, no. 9 and no. 11) are clearly observable due to the fact that they perfectly overlay on the frequency lines of MLT no. 4 and no. 8. The profit of having more than one frequency line extracted by the MHT can be seen, for example, by inspection of the frequency lines around 900 Hz belonging to the MLT no. 1 in Fig. 3. At around 30 min the two lines with the higher frequencies are deleted by the MHT. The subsequently confirmed frequency lines (originating from the same vessel as the deleted ones) are correctly add to MLT no. 1, because the lower lines (867 Hz and 879 Hz) are still tracked by the MHT.

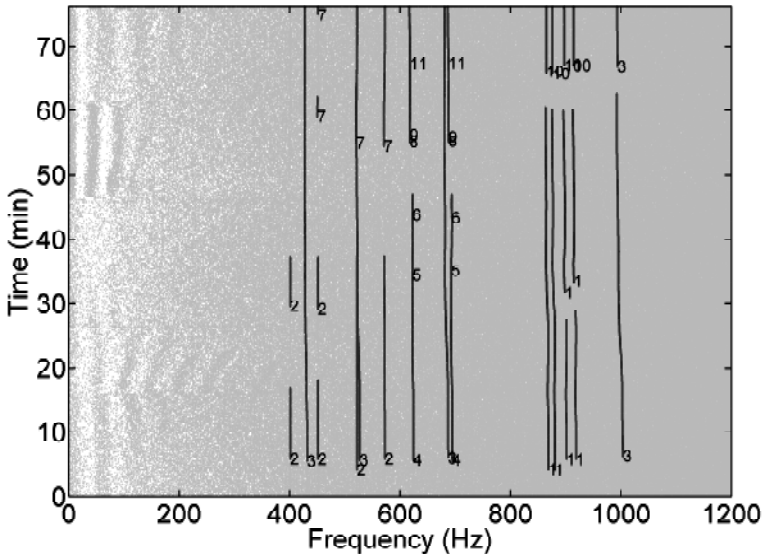


Figure 3: Tracking results showing the frequencies of extracted MLT's.

5 Summary and Conclusion

This paper introduced a multi-target tracking approach for narrowband passive sonar targets. After a review of some sonar specific features and a brief algorithm description, an application of this algorithm on the results of a simulated scenario was presented.

It was shown that this algorithm has the capability to automatically track the relevant frequency lines of all targets without a manual input by the user. It therefore has a great potential for a future tracking system.

References

[BH09] K. Brinkmann and J. Hurka. Broadband Passive Sonar Tracking. In S. Fischer, E. Maehle, and R. Reischuk, editors, *Informatik 2009 - Im Focus das Leben*, volume P-151 of *GI-Edition - Lecture Notes in Informatics (LNI)*, 2009.

[BP99] S. Blackman and R. Popoli. *Design and Analysis of Modern Tracking Systems*. Artech House, 1999.

[Koc01] W. Koch. Target Tracking. In Stergios Stergiopoulos, editor, *Advanced Signal Processing Handbook*, volume 10, chapter 8, pages 1–31. CRC Press LLC, 2001.

Maneuver-Adaptive Multi-Hypothesis Tracking for Active Sonar Systems

Kathrin Seget[1,2], Arne Schulz[1] and Ulrich Heute[2]

[1]Bundeswehr Technical Centre for Ships and Naval Weapons,
Naval Technology and Research (WTD 71)
Research Department for Underwater Acoustics and Marine Geophysics (FWG),
Klausdorfer Weg 2-24, 24148 Kiel

[2]Institute for Circuit and System Theory,
Faculty of Engineering, University of Kiel,
Kaiserstrasse 2, 24143 Kiel

Abstract: In undersea surveillance, active sonar systems are commonly used to detect submarines. These sonar systems allow high detection ranges, but the interaction of sound with the sea bottom may lead to a high number of false alarms as well, especially in shallow-water environments. Therefore, automatic detection and tracking procedures are needed to provide helpful assistance to sonar operators. The Multi-Hypothesis Tracking approach presented in this paper is one of these procedures. It is based on nonlinear Kalman Filtering.
In Kalman Filtering the assumption on underlying target dynamics is essential and has considerable impact on the overall tracking performance. As targets usually maneuver, their dynamics are varying and hidden. To include variable target dynamics, a Multi-Hypothesis tracking algorithm is adapted to consider target maneuvers by estimating and adjusting the process-noise level in the Kalman Filter equations. The level of process noise is determined for every track hypothesis individually based on the estimated velocities of the target. The impact on the tracking result is shown by applying the presented approach to different multistatic sonar datasets and comparing it to results gained by tracking with one global level of process noise. Tracking results are quantified by several tracking-performance metrics.

1 Introduction

Active sonar systems are used in Anti-Submarine Warfare (ASW) to detect submarines. Sonar contacts may include a high number of false alarms mainly due to the interaction of sound with the sea bottom and the sea surface, especially in shallow-water environments. To provide helpful assistance to sonar operators, automatic detection and tracking techniques are applied. Data association in the presented tracking approach is realised by Multi-Hypothesis Tracking (MHT) where a single track is represented by a set of weighted track hypotheses with the weight denoting the probability that the respective hypothesis is the true target track. The estimation of a target's state within the single hypotheses is

realised by applying the Unscented Kalman Filter (UKF) [JU04]. In Kalman Filtering the assumptions on the underlying target dynamics are essential. In many tracking approaches the target's dynamics are modelled as a target travelling with a constant velocity, known as the Nearly Constant Velocity (NCV) model. Deviations of the target's behaviour from this assumption are modelled as process noise, usually set a priori and fixed. But since targets usually change their dynamic behaviour regularly, a fixed process noise does not lead to optimal tracking results. In this paper a method is proposed to estimate and adapt the level of process noise. The determination of the level of process noise is based on the estimated velocities of the track-state hypotheses. To show the impact of an adaptively chosen level of process noise the algorithm is applied to different multistatic datasets with maneuvering as well as non-maneuvering targets. If a sonar system operates in a multistatic geometry, received data can be fused appropriately to improve tracking performance. In this case a centralised fusion strategy is applied [SSH10].

2 MHT Algorithm

Tracking using the MHT scheme as described in [KKU06] is done in the Cartesian plane. Thus, state vectors $\mathbf{x}$ are defined as Cartesian vectors with information on position and velocity. Assuming a linear dependency of the subsequent states, the underlying system can be described in matrix notation:

$$\mathbf{x}_{k+1} = \mathbf{A} \cdot \mathbf{x}_k + \mathbf{w}_k \tag{1}$$

with the system matrix $\mathbf{A}$, $\mathbf{w}$ is a Gaussian distributed random variable modelling the process noise (compare section 3). Sonar contacts $\mathbf{z}$ which are processed by the MHT contain information on range r, angle φ between contact and receiver and, if a Doppler can be extracted, the range rate $\dot{r}$. The vectors $\mathbf{z}$ are nonlinearly dependent on the state vectors $\mathbf{x}$ according to the measurement function h and distorted by additive White Gaussian noise $\mathbf{v}$. To process the nonlinear measurements for updating existing hypotheses, the UKF is used applying an Unscented Transform [JU04] in the filtering step where hypotheses states are transformed to allow for an appropriate update of the hypotheses states. To limit the number of hypotheses, gating, pruning and merging [BP99] are applied to the MHT algorithm. Furthermore, sequential track extraction [vK98] is included in the track management of the algorithm to confirm and delete tracks.

3 Dynamic Model and Process Noise

In Kalman filtering, sequential estimation of the conditional probability density function $p(\mathbf{x}_k|\mathbf{Z}^k)$ of a state $\mathbf{x}_k$ at the discrete point of time k is performed with $\mathbf{Z}^k = \{\mathbf{z}_1, \mathbf{z}_2, ..., \mathbf{z}_k\}$ denoting all measurements processed until time k. The prediction step utilises the assumptions on target dynamics specified by the matrix $\mathbf{A}$ in (1) for linear dynamics. The random variable $\mathbf{w}_k$ models the uncertainties of the assumed target dynamics as a Gaussian dis-

tributed random variable with zero-mean and covariance matrix

$$\mathbf{Q}_k = E[\mathbf{w}_k \cdot \mathbf{w}_k^T] \tag{2}$$

with E denoting the expectation operator. The predicted state $(\hat{\mathbf{x}}_k^-, \hat{\mathbf{P}}_k^-)$ is determined by the previous state $(\hat{\mathbf{x}}_{k-1}, \hat{\mathbf{P}}_{k-1})$ and the matrices $\mathbf{A}$ and $\mathbf{Q}$, specified by the system model. Due to relatively low measurement rates used in active sonar (e.g. sampling period $T = 60$s) an appropriate estimate of the acceleration is impossible. Thus, a dynamic model that does not include any acceleration in the target state should be chosen [BSB00] and an appropriate choice for the motion model in submarine tracking is a NCV model. The covariance matrix $\mathbf{Q}$ of the discretised process noise modelling accelerations is

$$\mathbf{Q} = \begin{bmatrix} T^3/3 & 0 & T^2/2 & 0 \\ 0 & T^3/3 & 0 & T^2/2 \\ T^2/2 & 0 & T & 0 \\ 0 & T^2/2 & 0 & T \end{bmatrix} \cdot q, \tag{3}$$

with q denoting the continuous-time process noise intensity, which is derived by the auto-correlation function [BSRLK01]

$$E[\tilde{v}(t)\tilde{v}(t-\tau)] = q \cdot \delta_0(\tau). \tag{4}$$

As a guideline for the choice of the intensity of the process noise, q should be set such that changes of the target's velocity during one sampling period T are of the order

$$\Delta v_T \simeq \sqrt{qT} \Leftrightarrow q \simeq \frac{\Delta v_T^2}{T}. \tag{5}$$

These guidelines are now used for a maneuver-adaptive tracking approach. The target states $\mathbf{x}$ include velocities estimated during Kalman Filter processing. Since velocities are estimated in $x-$ and $y-$direction of the Cartesian coordinate plane separately, a level of process noise for each of these directions can be determined individually. The levels of process noise $q_x(k, N)$ and $q_y(k, N)$ are determined applying (5) as an average using the last N estimated velocities with their corresponding time stamps t based on [Sch09]

$$q_x(k, N) \quad = \quad \sum_{n=1}^{N} \frac{(\dot{x}_{k-(n-1)} - \dot{x}_{k-n})^2}{t_{k-(n-1)} - t_{k-n}} \cdot \frac{1}{N}, \tag{6}$$

adequately for $q_y(k, N)$.
In MHT, every track consists of several track hypotheses including the current Cartesian velocities. Thus, (6) is applied to every track hypothesis to estimate an individual level of process noise. Only tracks that have already been confirmed are subject to an adaptive estimation of process noise. Due to small association gates, a low level of process noise might lead to a missed association of a contact to a true track, especially when target maneuvers appear after a period of non-maneuvering. The MHT prevents gates from becoming such small that missed association due to starting maneuvers occur. If no contact can be associated to an already extracted track, its level of process noise is significantly increased.

4 Tracking Results

To show the influence of the presented method to adjust the level of process noise adaptively, it has been included in the MHT-tracking algorithm and applied to different datasets. Results are expressed by several tracking-performance metrics [CdT06] and compared to metrics obtained by processing the data with a fixed level of process noise. In both datasets the sonar sensors used are buoy systems with a fixed source and two separate receivers operating bistatically leading to a multistatic geometry. The multistatic processing of the data requires an appropriate data fusion. For this paper a centralised fusion strategy as presented in [SSH10] is performed.

4.1 ARL:UT

In the ARL:UT sonar dataset two simulated targets were injected into real experimental sonar data [CCL06], [LC07]. The targets feature two different, but constant velocities. Target 1 is a slowly moving target (approximately 2 knots) and target 2 is a faster moving target (approximately 10 knots). Table 1 summarises the tracking results separately for target 1 and target 2. The value for a fixed process noise level for further analysis is set to $q = q_x = q_y = 0.01\mathrm{m}^2/\mathrm{s}^{-3}$. If no contact can be associated to an already extracted track when applying an adaptive level of process noise, q_x and q_y are increased to $q_x = q_y = 0.1\,\mathrm{m}^2/\mathrm{s}^{-3}$. The value N, determining the number of preceding track hypothesis states used to calculate the q_x and q_y according to (6) is set to $N = 1$ in a first step. Both approaches yield higher quality tracks for the slowly moving target. The track probability of detection (TPD), the ratio of the time the target is tracked to the time the target is present, is higher in case of slowly moving targets. The same holds true for the track localisation error (TLE) and the latency (LAT), the number of pings needed to extract the track . Moreover, it is obvious, that for the ARL:UT data, an adaptively estimated level of process noise does not influence the tracking performance significantly. The track false alarm rate (TFAR) only increases slightly. The TLE is hardly changed for both targets. This is due to the fact that the dynamics of the target trajectories follow the assumption of targets travelling with a constant velocity precisely. In this case, filtered target states are close to optimum and cannot be improved by an adaptation of the process noise.

Table 1: Tracking performance metrics for a fixed and an adaptive level of process noise applied to the ARL:UT data.

	Process Noise q	
	Fixed	Adaptive
TPD [slow/fast]	0.95/0.77	0.95/0.77
TFAR	0.38	0.43
TLE[slow/fast]	35.10/105.54	35.13/104.79
LAT [slow/fast]	2/13	2/13

4.2 SEABAR07 Run A01

From the SEABAR07 trial, conducted by the NATO Undersea Research Center in 2007, run A01 supplies the data for this paper. During this run the target is performing maneuvers.
The positive influence of an adaptive determination of process noise on the localisation accuracy can be read from table 2, which lists certain tracking-performance metrics for comparison. The TPD is equal for both approaches. TFAR is slightly increased for the adaptive approach. Applying an adaptively estimated level of process noise influences the track localisation accuracy positively. The TLE decreases considerably.

Table 2: Tracking performance metrics for a fixed and an adaptive level of process noise applied to the SEABAR'07 data.

	Process Noise q	
	Fixed	Adaptive
TPD	0.93	0.93
TFAR	0.29	0.30
TLE [m]	155.49	110.44
LAT [pings]	5	5

5 Conclusions and Outlook

The presented approach of including an adaptively estimated level of process noise within a Nearly Constant Velocity model has potential to increase the performance of a Multi-Hypothesis Tracking algorithm. The level of process noise is determined individually for every track hypothesis by using past and current filtered target states from which a deviation in the estimated velocities is calculated. Based on the deviation, the level of process noise is derived.
Tracking results obtained for two different datasets show that results for targets which are travelling with a constant velocity are hardly influenced because there targets already follow the Nearly Constant Velocity model precisely. But besides, maneuvering targets are tracked with a higher localisation accuracy. Thus, the presented approach effects the algorithm in such a way that localisation accuracy can be improved. A higher localisation accuracy is achieved nearly without influencing the tracking quality considering further performance metrics.
In this paper the level of process noise has been determined using changes in the estimated velocities directly. A further development could be to model the level of process noise as a target state and include it in the Kalman Filtering algorithm. Thus, estimation errors would be considered and past information would be included in the estimation process. To apply the presented approach to further datasets, including moving transmitters and receivers (e.g. Low Frequency Active Sonar (LFAS) systems), and analyse the influence of certain parameters within the tracking algorithm in more detail is subject to future work.

Acknowledgements

This work was made possible through collaboration between NURC, a NATO Research Centre, and WTD71-FWG (DEU) and by the Deployable Multistatic Sonar Joint Research Project (JRP), a collaboration between NURC, MOD (U.K.), and the Office of Naval Research (U.S.). The authors would like to thank B. La Cour for making the ARL:UT dataset available.

References

[BP99] S. Blackman and R. Popoli. *Design and Analysis of Modern Tracking Systems*. Artech House Radar Library, 1999.

[BSB00] Y. Bar-Shalom and W. D. Blair, editors. *Multitarget-Multisensor Tracking: Applications and Advances*, volume 3. Artech House, MA, 2000.

[BSRLK01] Y. Bar-Shalom, X. Rong Li, and T. Kirubarajan. *Estimation with Applications to Track and Navigation*. John Wiley & Sons, 2001.

[CCL06] B. R. La Cour, C. W. Collins, and J. L. Landry. Multi-everything Sonar Simulator (MESS). In *Proceedings of the 9th International Conference on Information Fusion*, Florence, Italy, July 2006.

[CdT06] S. Coraluppi and P. de Theije. Benchmark Evaluation of Multistatic Trackers. In *Proceedings of the 9th International Conference on Information Fusion*, Florence, Italy, July 2006.

[JU04] S. Julier and J. K. Uhlmann. Unscented Filtering and Nonlinear Estimation. *Proceedings of the IEEE*, 92(3):401–422, March 2004.

[KKU06] W. Koch, J. Koller, and M. Ulmke. Ground target tracking and road map extraction. *ISPRS Journal of Photogrammetry & Remote Sensing*, 61:197–208, 2006.

[LC07] B. La Cour. Bayesian Multistatic Tracking with Doppler-Sensitive Waveforms. In *Proceedings of OCEANS 2007*, Aberdeen, Scotland, 2007.

[Sch09] V. Scheltow. Analysis of the Influence of the Motion Model on the Tracking Result. Diploma thesis, Institute for Circuit and Systems Theory, Christian-Albrechts University Kiel, 2009. (in German).

[SSH10] K. Seget, A. Schulz, and U. Heute. Multi-Hypothesis Tracking and Fusion Techniques for Multistatic Active Sonar Systems. In *Proceedings of the 13th International Conference on Information Fusion*, Edinburgh, Scotland, July 2010.

[vK98] G. van Keuk. Sequential Track Extraction. *IEEE Transactions on Aerospace and Electronic Systems*, 34(4):1135–1148, October 1998.

A Multi-agent Concept for Multistatic Sonar

Frank Ehlers

NATO Undersea Research Centre (NURC)
19126 La Spezia
Italy
ehlers@nurc.nato.int

Abstract. In this paper we present a novel and relevant application for multi-agent systems which is to implement autonomy for Multistatic Sonar. Multistatic Sonar is used in Antisubmarine Warfare and exploits multiple aspects on the target generated by multiple acoustic sources and receivers. We propose to use a team of unmanned underwater vehicles starting with search operations as covert receivers and then switching to pursuit operations as actively communicating multistatic receivers. This results in a safe, persistent and efficient surveillance system. With the help of a simulation setup, we explain what are the specific challenges related to the Multistatic Sonar setup. These challenges are mainly related to the highly nonlinear measurement model and to communication constraints. We then briefly outline a potential "baseline" solution for the overall system architecture capable to face these challenges.

Keywords: data fusion, multistatic sonar, multi-agent control

1 Introduction

As an example for a multi-agent surveillance application, the problem tackled in this paper is that of decentralized sonar detection and tracking with a team of autonomous agents (typically AUVs, Autonomous Underwater Vehicles) [1]. The problem is further made more complex by communication constraints: Underwater acoustic communication is notoriously affected by bandwidth and range limitations; in addition, the tracking team members wish to remain covert, and hence they do not want to communicate to each other, unless to hand over to the other team members not only the information but also the tracking task. They have to do so for two reasons: (i) because only the diversity generated by multiple aspects on the target allows a sufficiently high probability of detection [2] and (ii) because the target can be much faster than a single AUV.
However, adding 'autonomy' to multistatic sonar has to face especially the challenge that 'teamwork' is required between team members including the coordination of work schedules on potential targets (including a lot of clutter targets) while communication between team members can only happen on a very limited basis. In other words, the autonomous multistatic sonar will fail if the AUVs show pathological behavior, e.g. they all stick to the same target (which might not even be the real

target, but a false alarm) and it will fail if they all communicate (indicating where they are and allowing the target to make a plan to evade).
The contribution of this paper is a complete concept for the implementation of autonomy in multistatic sensor networks. Although the specific algorithms for this implementation are not given in detail, the high-level description of the concept is useful to explain special demands on sensor data fusion algorithms for autonomous applications.

2 Description of a Multistatic Sonar System

In this section, we introduce the multistatic measurement system. The description summarizes state-of-the-art technology for manned systems. Furthermore, it serves as the basis for the plans (which are then further outlined in section 3) on how to add autonomy to this surveillance system.

2.1 Teamwork in Multistatic Sonar

Multistatic sonar is by default a coordinated effort aimed at gaining a more accurate view of the target and discriminating it from the background. This is true already for manned systems. The plan for the autonomous system is to use the degrees of freedom added by the autonomy of its assets in order to improve the quality of the measurement. Autonomy provides additional degrees of freedom because the assets are designed for the specific surveillance task only and do not have to be as much under constraints as manned platforms are which have normally multiple objectives and parallel missions.

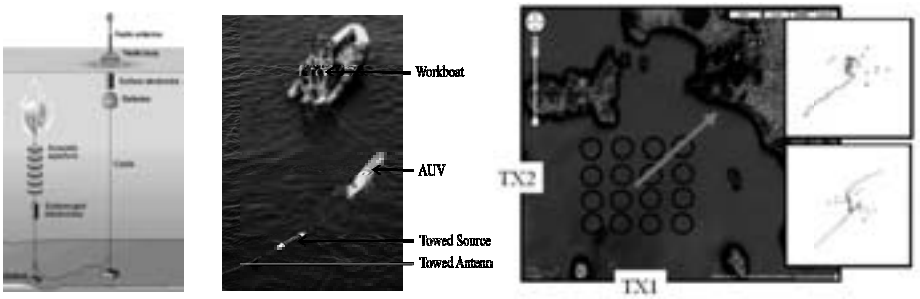


Fig. 1. (Left) Stationary acoustic source, (Middle) Deployment of the OEX Explorer AUV and its towed source, together with a slim towed array. (The workboat is only necessary for deployment and recovery. During the operation the AUV is operating at about 80m depth.) (Right) Test bed: Geometry in an example environment (right/main picture). Simulation results of teamed AUVs trapping a target: For a slow target (right/top right) and a faster target (right/bottom right).

2.2 Advantages of Multistatic Sonar

Multistatic active sonar has the obvious benefit of combining data from multiple pairs of sonar transmitters and receivers. When using multiple sources, a single receiver can already exploit these "data fusion benefits" of different bistatic aspects, without communicating, i.e. it can operate completely covert. This covertness does not allow the target acting cleverly, thus its stealth capability diminishes. The cost of losing covertness can be quantified by comparing the detection performance for two different Target Strength values: assigning to the clever target the lowest value and for the innocent target a substantially higher value.
However, to finally discriminate the target from background noise or clutter targets, the cross-fixing of tracks from multiple receivers is necessary. This has been demonstrated by several sea trials. I.e. in the final stage of target tracking (or target classification) the receivers have to give up their covertness for the benefit of a low false track rate (or 'low false alarm rate' when speaking about the entire autonomous surveillance system).

2.3 Physical agents carrying sonar receivers

Fixed and moving assets can be mixed. At NURC, a transmitter buoy is available and as receivers AUVs (Fig. 1) are used towing a receiving array and a quite powerful transmitter (given limited battery capacity) that is proposed in our context to be used for reliable long-range communication purposes. The source buoys can also be used for underwater communication (given an adequate receiving unit), and they can receive/exchange commands & data via radio (allowing the utilization of communication gateway buoys). The AUVs have low speed compared to the speed of the target.

2.4 Multistatic data fusion architecture

Data fusion merges the data taken from the different receivers in the multistatic system into a coherent picture. This picture reveals a series of contact points, which create a track showing the target's progress through the water [4].

2.5 Example Concept of Use

As an example, the idea is to implement the scenario of a multistatic survey and pursuit in clutter as a test bed (or simulation/experimentation environment). The geometry of the test bed is depicted in Fig. 1 (right). (Note: The region depicted has just been chosen as an example.) Two stand-off sources ensonify ("illuminate") a shallow water area. Since sound is reflected from a complex sea bottom structure, false alarms are generated by these clutter returns. The best way to overcome this is to send AUVs, equipped with towed arrays into the area, using their good angular resolution and, hence, their ability to discriminate target echoes from background

echoes when they are close enough. As a target, a submarine is hiding in this area. Its evasive action is to move to North East. The task of N (in Fig. 1 (right): N=16) AUVs is to organize themselves in a way that they keep the most accurate track of the target under the constraints of both: a minimal usage of communication and a minimal number of AUVs involved in the tracking task. Minimal communication is important, because covertness is a big advantage. Involving a minimal number of AUVs is important to prevent being exhausted by false alarms/false tracks (which cannot be avoided in difficult measurement environments).

3 System concept

In this section we describe the system concept for the autonomous multistatic surveillance system. We divide the entire problem into subtopics by making appropriate approximations and simplifications. This leads to a chain (or sequence) of tasks for which a solution is much easier to find than for the original entire problem.

3.1 Decision making necessary for a computationally feasible solution

For the heterogeneous multi-agent concept described in section 2, we can formally write the solution of the corresponding control problem for the system following the description as a stochastic nonlinear minimax dynamic game with noisy measurements [6]. However, we will not be able to solve it with standard methods (even numerically) because of its complexity. Thinking about a particle stream like simulation, it is shown in [7] that taking out particles associated with low probability greatly minimizes the demand on processing power in the numerical simulation. We interpret “taking out particles” in our context as “making decisions”, and by introducing a reasonable structure for the decision making process, we generate a computationally feasible solution of the formally written control problem.

3.2 Evasive target behavior

As far as target behavior is concerned, the general formulation in 3.1 covers the observer design for a maximally ‘clever’ target. However, while implementing the concept 3.1, simplifications are made and this optimal behavior of the surveillance team is not guaranteed anymore. Therefore, we put the developments described next under an ‘outer training loop’, meaning that once these developments are implemented, a simulation based learning phase starts where iteratively clever target behaviors and adaptive surveillance team parameters are found. This iteration should converge to a stable surveillance system because of its in general ‘worst case design’ implemented by the minimax formulation in subsection 3.1.

3.3 Multi-agent decision making

The decision making process envisioned for the multistatic surveillance task should be data driven, because we know from previous experiments (as described in section 2), that the quality of the achievable data fusion depends on the quality of the contributing sensors which then depend on the hypothetical target track itself. In order to keep the overall complexity of the decision making process low, we separate the individual planning (data driven because it depends on the specific target track) from the team behavior. This separation idea is further explained in subsection 3.5. Before that, we have to introduce a grid-based processing and control scheme in subsection 3.4.

3.4 Grid-based processing and control scheme

In this subsection we introduce the term 'virtual Sonobuoy' (virtSB). A virtSB is meant to be a sensor located in a grid cell, whereby a grid is covering the complete surveillance area. Each grid cell contains several virtSBs, belonging to each multi-bistatic receiving unit and to each multistatic combination of receiving units. Each virtSB is used to describe a partially observable Markov decision process (POMDP).
We assume that the virtSB has (virtually) the capability to (i) pursue a target occupying the same (or neighboring cells), (ii) ignore the contact data from this cell (or neighboring cells), or (iii) wait for more data before making a decision between (i) and (ii). How the virtSBs will be used to lead to AUV actions will be shown in subsections 3.7 and 3.8. Before that, we especially deal with the multistatic virtSBs in subsections 3.5 and 3.6.

3.5 Multistatic virtSBs

The virtSBs for multistatic actions would require communication between their mother-AUVs (from which they are stemming) to perform the joint pursuit action. But this communication should occur as rarely as possible. Furthermore, the overall concept has to prevent also a pathological behavior where all AUVs pursue the same target track.
We propose to overcome this challenge by introducing a learning phase especially for teamwork. In team sports, team members are exploring the value of collaboration during training sessions. For example in basketball, they exercise on the so-called 'no-look pass' between them. We make the analogy between 'no-look' and no-communication and propose a training (or learning) phase for the specific autonomous teamwork features. We explain in the next subsection how this analogy influences the overall design architecture of the multistatic surveillance team.

3.6 Efficient team behavior

We assume that a sufficient number of receivers for tracking a target is only the minority of all available receivers. A solution of this (now simplified) problem can be

found if we recall the ‘minority problem’, well studied in game theory [5]. The ‘minority problem’ can be solved by learning. There exists also an analogy between the proposed learning scheme and the Potts model known in statistical physics.

3.7 virtSBs with weights

The virtSBs stemming from different mother-AUVs have different sensor quality, as bistatic (or multi-bistatic) sonar geometries are range and aspect dependent. The multistatic virtSBs found in the multistatic activation map have a potentially better quality than the single receiver virtSBs. The quality of each virtSB can be calculated in the same manner as for the sensor model in a multistatic tracker [4]. In [4], measurement uncertainties and the detection performance are appropriately modeled.

3.8 virtSBs generate plans

The virtSBs inherit from their mother-AUVs or mother-‘team of AUVs’ constraints on their motion. Whereas previously we were assuming a virtSB could pursue a target with no constraints on its own motion, we now have to introduce that own motions in-line with the motion of the mother-AUV/team are more likely.
Given a sensor model in subsection 3.7 and a motion model here, the virtSBs are now able to perform path planning with the goal to follow their targets.

3.9 Movement equations AUVs

Onboard each AUV, it is then possible to collect all plans generated from the virtSBs so far. Each plan comes with a length in time and a quality. Clustering algorithms are able to shrink the collection of plans to a few options. The AUV is then choosing the option that helps most to reach the operation goal defined by the system’s operator.

3.10 Waypoint for AUVs and Decisions on Communication

Given a decision on a plan inside the AUV in subsection 3.9, this plan also includes waypoints that the physical AUV has to reach next and in the near future. Also, in the case that the plan includes active communication to exchange target track information, this communication has to be executed. Speaking in terms of a behavioral design of the AUV control, this simply means that input to waypoint and communication behaviors has to be provided.

3.11 Real-time measurement/simulation

Simulations play a crucial role in the concept. They are used in subsection 3.6 and 3.2 in order to learn the multistatic behavior (or activation map) and to learn adaptability to clever target behavior. The usage of simulations, like here in the sense of a reinforcement learning structure, is a common procedure and part of a numerical

solution of the problem described in 3.1. The demonstration of the concept at sea needs real physical assets and real-time measurement facilities, which are developed and tested now in the framework of NURC's collaborative ASW project [3].

4 Summary

Multistatic operations with autonomous underwater vehicles enable new anti-submarine warfare missions. Under the umbrella of game and control theory with a probabilistic multi-sensor setup for data fusion and tracking, a complete concept for the implementation of autonomy in multistatic sensor networks has been developed. The concept identifies special demands for sensor data fusion algorithms and the interfaces to the overarching control, decision and communication processes. A realization of this concept is aiming to produce teamed (or swarming) multiple AUVs which autonomously perform data based (and target adaptive) multistatic surveillance.

Acknowledgments. The author thanks the members of the GLINT 09 and SUBNET 09 teams, and K. McCoy, T. Curtin, J. Pitton, S. Kemna and R. Streit for initial discussions and support.

References

[1] F. Ehlers, "Multi-Robot Teamwork in Multistatic Sonar," IEEE ICRA2010 Workshop: Search and Pursuit/Evasion in the Physical World: Efficiency, Scalability and Guarantees, Alaska, 3 May, 2010.

[2] F. Ehlers,"Particle filter for information management in an autonomously operating activated distributed sensor network," Proceedings of OCEANS'07, Aberdeen, June 2007.

[3] D. T. Hughes, R. Been, and A. Vermeij, "Heterogenous underwater networks for ASW: Technology and techniques," in Maritime Systems and Technology, Genova, Italy, 2008.

[4] M. Daun, F. Ehlers,"Tracking Algorithms for Multistatic Sonar Systems", EURASIP Journal on Advances in Signal Processing, 2010.

[5] H.G. Schuster, Complex Adaptive Systems, Scator Verlag, 2001.

[6] C. D. Charalambous, "Stochastic Nonlinear Minimax Dynamic Games With Noisy Measurements," IEEE Trans. Automatic Control, vol. 48, no. 2, Feb. 2003.

[7] H. J. Kappen,"Path integrals and symmetry breaking for optimal control theory," J. Stat. Mech. (2005) P11011.

Microphone Networks for Shooter Localization*

David Lindgren[1†] **Fredrik Gustafsson**[2] **Hans Habberstad**[1]

[1]Swedish Defence Research Agency, firstname.lastname@foi.se
[2]Linköping University, fredrik@isy.liu.se

Abstract: Shooter localization in a wireless network of geographically distributed microphones is studied. The arrival time of both the acoustic muzzle blast (MB) from the gunfire and the ballistic shock wave (SW) from the bullet can be registered by the microphone nodes and considered as measurements. To relax the requirements on accurate node clocks, the approach to base the localization on the time difference of MB and SW at each sensor is analyzed. Cramér-Rao lower bound analysis indicates how a lower bound of the root mean square error depends on the synchronization error for different approaches to use the measurements. The localization problem is formulated in a separable nonlinear least squares framework. A simplified ballistic model with only one, assumed unknown, parameter is used. Results from field trials with different types of ammunition demonstrate excellent results using the MB–SW difference for both the position and the aiming angle of the shooter.

1 Introduction

Today, there are more than a few acoustic shooter localization systems commercially available. One or more microphone arrays are commonly used, each synchronously sampling acoustic phenomena associated with gunfire, see [MB06]. Some of these systems are mobile, some have even been proposed to be based on microphone arrays that soldiers carry on their helmets.

In this work we consider shooter localization based on singleton microphones geographically distributed in a wireless sensor network, which indeed is a less common approach. We focus on a typical issue in wireless networks; the effects of limited sensor synchronization. Even high-end node clocks tend to drift over time and thus need periodical adjustments. For localization algorithms that rely on accurate timing like the ones based on *time difference of arrival* (TDOA), this is of course bad news, since regardless if the synchronization is supported by the GPS or by communication/network techniques, the synchronization procedures are associated with costs in battery life and/or communication resources that usually must be kept at a minimum.

We have earlier derived fundamental estimation bounds for shooter localization models

*This work is funded by the VINNOVA supported Centre for Advanced Sensors, Multisensors and Sensor Networks, FOCUS, at the Swedish Defence Research Agency, FOI.

†Corresponding author, Tel +46 8 555 030 00, Fax +46 8 555 031 00

with respect to the synchronization errors, see [LWGH10]. An accurate method independent of the synchronization errors is in focus of our investigations, the *MB–SW difference model*. A computationally efficient bullet deceleration model is also an important part of the studies. Algorithms are tested on data from field trials. In this work we summarize the methods described in [LWGH10], and extend the numerical evaluation from the simplified plane implementation earlier used to a more realistic 3-dimensional setting. Here, we also consider the ballistic deceleration parameter as an unknown, which means that a priori knowledge of the ammunition is no longer needed.

Closely related results have previously been presented in [WKD09], where fundamental sniper localization estimation bounds are derived with respect to both sensor position errors and synchronization errors. Furthermore, the authors in [SMAL+04] study the synchronization error impact on the sniper localization of an urban network by using Monte Carlo simulations. One of the results is that the inaccuracy increased significantly (> 2 m) for synchronization errors exceeding approximately 4 ms. 56 small wireless sensor nodes were modeled. Also [Dan05] should be mentioned, where combinations of directional and omnidirectional acoustic sensors for sniper localization are evaluated by perturbation analysis.

2 Localization Principle

Two acoustical phenomena associated with gunfire will be exploited to determine the shooter's position: the *muzzle blast* and the *shock wave*. The principle is to detect and time stamp the phenomena as they reach microphones distributed over an area, and let the shooter's position be estimated by, in a sense, the most likely point, considering the microphone locations and detection times.

The muzzle blast (MB) is the sound that probably most of us associate with a gun shot; the "bang". The MB is generated by the pressure depletion in effect of the bullet leaving the gun barrel. The sound of the MB travels at the speed of sound in all directions from the shooter. Provided that a sufficient number of microphones detect the MB, the shooters position can be more or less accurately determined.

The shock wave (SW) is formed by supersonic bullets. The SW has (approximately) the shape of an expanding cone, with the bullet trajectory as axis, and reaches only microphones that happen to be located inside the cone. The SW propagates at the speed of sound in direction away from the bullet trajectory, but since it is generated by a supersonic bullet, it always reaches the microphone before the MB, if it reaches the microphone at all. A number of SW detections may alone primarily reveal the *shooting direction*. Of course, the most informative approach is to use both MB and SW in combination.

3 Estimation Framework

It is assumed that the coordinates of the microphones are known with negligible error, that the arrival times of the MB and SW at each microphone are measured with significant synchronization error, and that the shooter position and aim direction are the sought parameters. The signal models (for the MB, SW, etc.) are generally of the form

$$y = h(x, \theta; p) + e, \tag{1}$$

where y is a vector with the measured detection times, h is a nonlinear function, and where θ represents the unknown parameters apart from the sought shooter position $x \in \mathbb{R}^3$. The error e is assumed to be stochastic. Given the M sensor locations in $p \in \mathbb{R}^{M\times 3}$, nonlinear optimization can be performed to estimate x using the nonlinear least squares (NLS) criterion

$$\hat{x} = \text{argmin}_x \min_\theta V(x, \theta; p), \tag{2a}$$

$$V(x, \theta; p) = ||y - h(x, \theta; p)||^2_R. \tag{2b}$$

The loss function norm R is chosen by consideration of the expected error characteristics. Numerical optimization, for instance the Gauss-Newton method, can here be applied to get the NLS estimate.

For computational efficiency, we exploit sub-linear structures in the signal models and apply the weighted least squares method to the parameters entering linearly, *the separable least squares method* (SLS). With that motivation, the signal model (1) is rewritten as

$$y = h_N(x, \theta_N; p) + h_L(x, \theta_N; p)\theta_L + e. \tag{3}$$

Note that θ_L enters linearly here. The NLS problem can then be formulated as

$$\hat{x} = \text{argmin}_x \min_{\theta_L, \theta_N} V(x, \theta_N, \theta_L; p),$$
$$V(x, \theta_N, \theta_L; p) = ||y - h_N(x, \theta_N; p) - h_L(x, \theta_N; p)\theta_L||^2_R.$$

Since θ_L enters linearly it can be solved for by linear least squares (the arguments of $h_L(x, \theta_N; p)$ and $h_N(x, \theta_N; p)$ are suppressed for clarity):

$$\hat{\theta}_L = \text{argmin}_{\theta_L} V(x, \theta_N, \theta_L; p) = (h_L^T R^{-1} h_L)^{-1} h_L^T R^{-1}(y - h_N), \tag{4a}$$

$$P_L = (h_L^T R^{-1} h_L)^{-1}. \tag{4b}$$

Here, $\hat{\theta}_L$ is the weighted least squares estimate and P_L is the covariance matrix of the estimation error. This simplifies the nonlinear minimization to

$$\hat{x} = \text{argmin}_x \min_{\theta_N} ||y - h_N + h_L(h_L^T R^{-1} h_L)^{-1} h_L^T R^{-1}(y - h_N)||^2_{R'}, \tag{5a}$$

$$R' = R + h_L P_L h_L^T. \tag{5b}$$

This general SLS approach is applied to 3 different combinations of signal models for the MB and SW detection times.

4 Signal Models

Thus, we outline 3 shooter localization models that may be derived in the SLS framework:

- **MB**. This model only utilizes the MB, and ignores the SW. Given that the acoustic environment enables reliable detection of the muzzle blast, the MB model promises the most robust estimation algorithms. It also allows global minimization with low-dimensional exhaustive search algorithms; two parameters enter the model linearly (speed of sound and shooting time), and, besides the position in $\mathbb{R}^3$, none nonlinearly; $\dim \theta_N = 0$, $\dim \theta_L = 2$. This model is thus suitable for initialization of algorithms based on the subsequent models.
- **MB;SW**. For this model, every MB and SW measurement is stacked into a complete equation system. This total MB;SW model keeps all information from the observations, and should thus provide the most accurate and general estimation performance. However, the complexity of the estimation problem is large; $\dim \theta_N = 5$ (direction, speed of sound, initial bullet speed, ballistic parameter) and $\dim \theta_L = 1$ (shooting time).
- **MB–SW**. Algorithms based on the models above require that the synchronization error in each microphone is either negligible or can be described with a statistical distribution. The MB–SW model relaxes such assumptions by eliminating the synchronization error by taking time differences of the two pulses at each microphone. This also eliminates the shooting time as a parameter, $\dim \theta_N = 5$, $\theta_L = 0$.

The SW depends on the bullet's somewhat involved ballistic trajectory. However, in the tradeoff between accuracy and computational efficiency, we claim that a reasonable compromise is to assume that the bullet follows a straight line at the decelerating velocity

$$v = v_0 - rd, \tag{6}$$

where v_0 is the initial velocity, d the traveled distance, and r an unknown ballistic parameter. The parameter r is estimated along with the other unknowns (v_0 etc.), and may besides the shooter location provide extra information to support ammunition classification etc.

5 Cramér-Rao Lower Bound

Cramér-Rao analysis provides a lower bound for the estimation error, particularly the localization root mean square error (RMSE). The bound may be evaluated for a specific shooter location, parameter setting, microphone positioning, and so on.

We have analyzed a scenario with 14 distributed microphones deployed for camp protection. The resulting localization bounds for the MB, MB;SW, and MB–SW model are depicted in Figure 1. Here, σ_e is the detection error level, and σ_b is the clock synchronization error. The error bounds are confined to the horizontal plane, since the height component

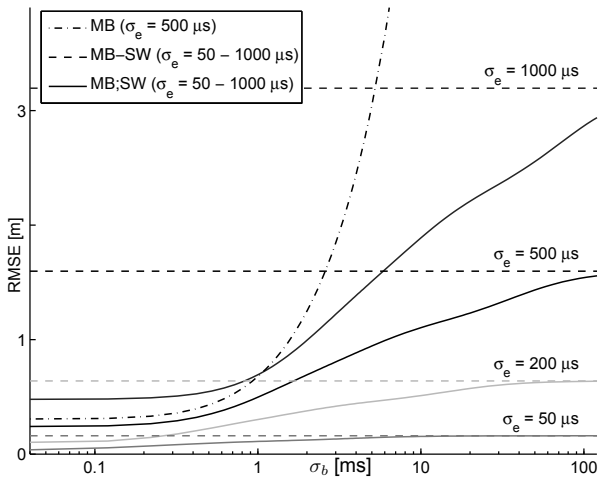


Abbildung 1: Cramér-Rao RMSE bound for the MB, the MB–SW, and the MB;SW models, respectively, as a function of the synchronization error (STD) σ_b, and for different levels of detection error σ_e.

is both very uncertain and assumed less interesting in the scenario. Based on this example, the following recommendations could be used as a rule-of-thumb: If $\sigma_b \gg \sigma_e$, then the MB–SW model should be used. If σ_b is moderate, then the MB;SW model should be used.

6 Experimental Data

A field trial to collect acoustic data on non-military small arms fire is conducted. 10 microphones are placed on an approximate 50 m radius around a fictitious camp. A total of 42 rounds are fired from 3 positions up to 400 meters away from the camp center. The shooter aims at a cardboard target at the opposite side of the camp. The resulting localization root mean square errors (RMSE) for the 3 shooter positions, together with the theoretical Cramér-Rao bounds are given in Table 1.

7 Conclusions

We have summarized a framework for estimation of shooter location and aiming angle from wireless networks where each node has a single microphone. Measurements on the acoustic muzzle blast (MB) and the ballistic shock wave (SW) may together reveal both the shooter's position and aim direction. A separable least squares (SLS) framework was

Tabelle 1: Horizontal plane localization RMSE and theoretical bound for the 3 different shooter positions using the MB–SW and MB;SW model, respectively, beside the aim RMSE for the MB;SW model. The aim RMSE is with respect to the aim at $\hat{x}$ against the target.

Shooter/camp Distance	400 m	200 m	0 m (Inside Camp)
RMSE($\hat{x}^{\text{MB;SW}}$)	22 m	9 m	2.4 m
MB;SW Bound	1.2 m	0.3 m	0.05 m
RMSE ($\hat{x}^{\text{MB–SW}}$)	28 m	7 m	N/A
MB–SW Bound	3.6 m	0.9 m	N/A
RMSE($\hat{\alpha}'$)	$0.12°$	$0.38°$	$0.61°$

proposed to limit the parametric search space and to enable the use of global grid-based optimization algorithms (for the MB model), eliminating potential problems with local minima. A simple ballistic model with one unknown parameter was used. The models have been analyzed on a scenario example and validated on data from a field trial. Compared to our earlier work, the analysis and evaluation have been conducted in 3 dimensions, and by assuming that the ballistic parameter is unknown, no prior knowledge about the ammunition is required anymore.

When the network node synchronization is perfect, both MB and SW measurements should be stacked into one large signal model for which SLS is applied. On the other hand, when the error becomes comparable to the detection error for MB and SW, the performance quickly deteriorates. In that case, the time difference of MB and SW at each microphone is preferably used, which eliminates the effect of any synchronization error. The effective number of measurements decreases in this approach, but as Cramér-Rao analysis can show, the root mean square position error is comparable to that of the ideal stacked model, at the same time as the synchronization error distribution may be completely disregarded.

Both the bullet speed and bullet retardation occur as nuisance parameters in the proposed signal models. Future work will investigate how the information in these parameters, together with the MB/SW signal forms, should be used to identify weapon and ammunition.

Literatur

[Dan05] E. Danicki. Acoustic sniper localization. *Archives of acoustics*, 30(2):233–245, 2005.

[LWGH10] David Lindgren, Olof Wilson, Fredrik Gustafsson, and Hans Habberstad. Shooter Localization in Wireless Microphone Networks. *To appear in Journal of Advances in Signal Processing*, 2010.

[MB06] J. Millet and B. Balingand. Latest Achievements in Gunfire Detection Systems. In *Proc. of the RTO-MP-SET-107 Battlefield Acoustic Sensing for ISR Applications*, Neuilly-sur-Seine, France, 2006.

[SMAL+04] Gyula Simon, Miklós Maróti, Ákos Lédeczi, György Balogh, Branislav Kusy, András Nádas, Gábor Pap, János Sallai, and Ken Frampton. Sensor network-based countersniper system. In *Proc. of SenSys*, Baltimore, MD, USA, 2004.

[WKD09] Gene T. Whipps, Lance M. Kaplan, and Raju Damarla. Analysis of sniper localization for mobile asynchronous sensors. In *Proc. of the SPIE Signal Processing, Sensor Fusion, and Target Recognition XVIII*, volume 7336, 2009.

Performance analysis of passive emitter tracking using TDOA, AOA and FDOA measurements

Regina Kaune
Fraunhofer FKIE, Dept. Sensor Data and Information Fusion
Neuenahrer Str. 20, 53343 Wachtberg, Germany
regina.kaune@fkie.fraunhofer.de

Abstract: This paper investigates passive emitter tracking using a combination of Time Difference of Arrival (TDOA) measurements with further different types of measurements. The measurements are gained by exploiting the signal impinging from an unknown moving emitter. First, a combined set of TDOA and Angle of Arrival (AOA) measurements is processed using the Maximum Likelihood Estimator (MLE). Then, a Gaussian Mixture (GM) filter is used to solve the tracking problem based on TDOA and Frequency Difference of Arrival (FDOA) measurements. In Monte Carlo simulations, the superior performance of the combined methods in contrast to the single TDOA approach is shown and compared with the Cramér-Rao Lower Bound (CRLB).

1 Introduction

Many applications require fast and accurate localization and tracking of non-cooperative emitters. In many cases, it is advantageous not to conceal the observation process by using active sensors, but to work covertly with passive sensors. The estimation of the emitter state is based on various types of passive measurements obtained by exploiting signals emitted by targets, [KMK10, Bec05, and references therein].
Measurements of Time Difference of Arrival (TDOA) and Frequency Difference of Arrival (FDOA) are obtained from a network of several spatially dislocated sensors. Here, a minimum of two sensors is often needed.
In the absence of noise and interference, a single TDOA measurement localizes the emitter on a hyperboloid with the two sensors as foci. By taking additional independent TDOA measurements from at least four sensors, the three-dimensional emitter location is estimated from the intersections of three or more hyperboloids. If sensors and emitter lie in the same plane, one TDOA measurement defines a hyperbola of possible emitter locations. That is why, the localization using TDOA measurements is called hyperbolae positioning. The sign of the measurement defines the branch of the hyperbola on which the emitter is located. The two-dimensional emitter location is found at the intersection of two or more hyperbolas from at least three sensors. This intersection point can be calculated by an analytical solution, see e.g. [SHH08, HY08].
Alternatively, a pair of two sensors moving along arbitrary but known trajectories can be used for localizing an emitter using TDOA measurements. In this case, the emitter location can be estimated by filtering and tracking methods based on further measurements over

time. The localization of an unknown, non-cooperative stationary emitter using TDOA measurements from a sensor pair has already been investigated in [Kau09].
In this paper, the gain in performance by combining TDOA measurements from one sensor pair with further different kinds of passive measurements is analyzed . The focus is on localization and tracking a non-cooperative moving emitter.
Firstly, TDOA measurements are combined using additional Angle of Arrival (AOA) measurements from one of the two sensors. The measurement set based on the TDOA and the combination of TDOA and AOA measurements are processed using the Maximum Likelihood Estimator (MLE). Secondly, measurement information is increased by FDOA measurements which are gained by differentiating the Frequencies of Arrival (FOA) of the sensor pair. In this case, relative motion between sensors and emitter is necessary. The Gaussian Mixture (GM) filter described in [Kau09] is used to obtain comparable results of the single TDOA and the combined TDOA/ FDOA method. In Monte Carlo simulations, the performance of the different methods is analyzed and compared with the Cramér-Rao Lower Bound (CRLB).

2 Problem description

A two-dimensional emitter-sensors scenario is considered. Let $\mathbf{e}_k = (\mathbf{x}_k^T, \dot{\mathbf{x}}_k^T)^T$ be the emitter state at time t_k, where $\mathbf{x}_k = (x_k,\ y_k)^T \in \mathbb{R}^2$ denotes the position and $\dot{\mathbf{x}}_k = (\dot{x}_k,\ \dot{y}_k)^T \in \mathbb{R}^2$ the velocity. Two sensors with the state vectors

$$\mathbf{s}_k^{(i)} = \left(\mathbf{x}_k^{(i)^T}, \dot{\mathbf{x}}_k^{(i)^T}\right)^T, \qquad i = 1, 2, \tag{1}$$

observe the emitter and receive the emitted signal. They move along arbitrary but known trajectories.
The emitter is assumed to move with constant velocity. Therefore, the emitter state can be modeled from the previous time step t_{k-1} by adding white Gaussian noise:

$$\mathbf{e}_k = \mathbf{F}_k\, \mathbf{e}_{k-1} + \mathbf{v}_k, \quad \mathbf{v}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}), \tag{2}$$

with

$$\mathbf{F}_k = \begin{pmatrix} 1 & 0 & t_k - t_{k-1} & 0 \\ 0 & 1 & 0 & t_k - t_{k-1} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \tag{3}$$

where $\mathbf{v}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{Q})$ means that $\mathbf{v}_k$ is zero-mean normally distributed with covariance $\mathbf{Q}$.
The TDOA measurement in the range domain is given by:

$$h^{\mathrm{r}}(\mathbf{e}_k) = ||\mathbf{x}_k - \mathbf{x}_k^{(1)}|| - ||\mathbf{x}_k - \mathbf{x}_k^{(2)}||, \tag{4}$$

where $||\cdot||$ denotes the vector norm and $\mathbf{r}_k^{(i)} = \mathbf{x}_k - \mathbf{x}_k^{(i)}$ denotes the relative position vector between the emitter and sensor i. The measurement process is modeled by adding

white Gaussian noise to the measurement function:

$$z_k^{\mathrm{r}} = h^{\mathrm{r}}(\mathbf{e}_k) + u_k^{\mathrm{r}}, \quad u_k^{\mathrm{r}} \sim \mathcal{N}(0, \sigma_{\mathrm{r}}^2) \tag{5}$$

where σ_{r} denotes the standard deviation of the measurement error in the range domain. The measurement noise u_k^{r} is i.i.d., the measurement error is independent from time to time, i.e. mutually independent, and identically distributed.

3 Combination of TDOA and AOA measurements

Let $\mathbf{s}^{(1)}$ be the location of the sensor, which takes the bearing measurements. The additional azimuth measurement function at time t_k is:

$$h^{\alpha}(\mathbf{e}_k) = \arctan \frac{x_k - x_k^{(1)}}{y_k - y_k^{(1)}} \tag{6}$$

Addition of white noise yields:

$$z_k^{\alpha} = h^{\alpha}(\mathbf{e}_k) + u_k^{\alpha}, \quad u_k^{\alpha} \sim \mathcal{N}(0, \sigma_{\alpha}^2), \quad u_k^{\alpha} \text{ i.i.d.}, \tag{7}$$

where σ_{α} is the standard deviation of the AOA measurement. AOA and TDOA measurement noise are assumed to be uncorrelated from each other.
At each time step, one azimuth and one TDOA measurement are taken. The azimuth measurement defines a line of possible emitter locations and the TDOA measurement localizes the emitter on a hyperbola. This pair of nonlinear measurements must be processed with nonlinear estimation algorithms.

CRLB

It is important to know the optimal estimation accuracy that can be achieved with the available measurements for the problem under consideration. The CRLB, the inverse of the Fisher information $\mathbf{J}$, provides a lower bound on the estimation accuracy for an unbiased estimator. The CRLB of the combined measurement set is calculated over the fused Fisher information. The Fisher information matrix (FIM) at time t_k is the sum of the FIMs based on the TDOA and the AOA measurements:

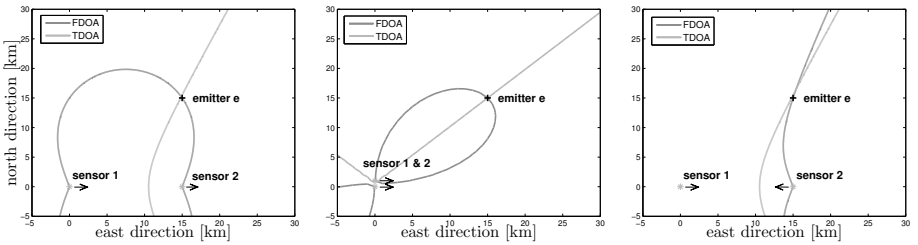
$$\mathbf{J}_k = \frac{1}{\sigma_{\mathrm{r}}^2} \sum_{i=1}^{k} \left(\frac{\partial h^{\mathrm{r}}(\mathbf{e}_i)}{\partial \mathbf{e}_k} \right)^T \frac{\partial h^{\mathrm{r}}(\mathbf{e}_i)}{\partial \mathbf{e}_k} + \frac{1}{\sigma_{\alpha}^2} \sum_{i=1}^{k} \left(\frac{\partial h^{\alpha}(\mathbf{e}_i)}{\partial \mathbf{e}_k} \right)^T \frac{\partial h^{\alpha}(\mathbf{e}_i)}{\partial \mathbf{e}_k}, \tag{8}$$

where

$$\frac{\partial h^{\mathrm{r}}(\mathbf{e}_i)}{\partial \mathbf{e}_k} = \frac{\partial h^{\mathrm{r}}(\mathbf{e}_i)}{\partial \mathbf{e}_i} \frac{\partial \mathbf{e}_i}{\partial \mathbf{e}_k}. \tag{9}$$

Therefore, the localization accuracy depends on the sensor-emitter geometry and the standard deviation of the TDOA and the azimuth measurements.

4 Combination of TDOA and FDOA measurements



(a) tail flight (b) parallel flight (c) flight head on

Figure 1: Combination of TDOA and FDOA measurements in three different scenarios

The FDOA measurement function depends not only on the emitter position but also on its speed and course: $h^{\mathrm{ff}}(\mathbf{e}_k) = \frac{f_0}{c} h^{\mathrm{f}}(\mathbf{e}_k)$, where f_0 is carrier frequency of the signal. Multiplication with $\frac{c}{f_0}$ yields the measurement equation in the velocity domain:

$$h^{\mathrm{f}}(\mathbf{e}_k) = (\dot{\mathbf{x}}_k^{(1)} - \dot{\mathbf{x}}_k)^T \frac{\mathbf{r}_k^{(1)}}{||\mathbf{r}_k^{(1)}||} - (\dot{\mathbf{x}}_k^{(2)} - \dot{\mathbf{x}}_k)^T \frac{\mathbf{r}_k^{(2)}}{||\mathbf{r}_k^{(1)}||}. \tag{10}$$

Under the assumption of uncorrelated measurement noise from time step to time step and from the TDOA measurements, we obtain the FDOA measurement equation in the velocity domain:

$$z_k^{\mathrm{f}} = h^{\mathrm{f}}(\mathbf{e}_k) + u_k^{\mathrm{f}}, \quad u_k^{\mathrm{f}} \sim \mathcal{N}(0, \sigma_{\mathrm{f}}^2), \tag{11}$$

where σ_{f} is the standard deviation of the FDOA measurement. The associated TDOA/ FDOA measurement pairs may be obtained by using the Complex Ambiguity Function ([St81]).
Fig. 1 shows the situation for different sensor headings after taking one pair of TDOA and FDOA measurements. The green curve, i.e. the branch of hyperbola, indicates the ambiguity after the TDOA measurement. The ambiguity after the FDOA measurement is plotted in magenta. The intersection of both curves presents a gain in information for the emitter location. This gain is very high if sensors perform a tail flight, see Fig. 1 (a).

CRLB

The Fisher information at time t_k is the sum of the Fisher information based on the TDOA and the FDOA measurements:

$$\mathbf{J}_k = \frac{1}{\sigma_{\mathrm{r}}^2} \sum_{i=1}^{k} \left(\frac{\partial h^{\mathrm{r}}(\mathbf{e}_i)}{\partial \mathbf{e}_k} \right)^T \frac{\partial h^{\mathrm{r}}(\mathbf{e}_i)}{\partial \mathbf{e}_k} + \frac{1}{\sigma_{\mathrm{f}}^2} \sum_{i=1}^{k} \left(\frac{\partial h^{\mathrm{f}}(\mathbf{e}_i)}{\partial \mathbf{e}_k} \right)^T \frac{\partial h^{\mathrm{f}}(\mathbf{e}_i)}{\partial \mathbf{e}_k}. \tag{12}$$

The FIM of the FDOA measurements not only depends on the geometry and the standard deviation of the measurements. But also they strongly depend on the relative speed vectors between emitter and sensors.

5 Simulation Results

A moving emitter is considered to compare the performance of the single TDOA approach and the combined methods. Fig. 2 (a) shows the measurement situation. Sensors fly one after another at a constant speed of 50 m/s and perform one maneuver to ensure observability. The emitter moves at a constant velocity of 40 m/s in east south direction. TDOA and FDOA measurements are gained from the network of the two sensors. Sensor $\mathbf{s}^{(1)}$ takes the azimuth measurements. TDOA measurement standard deviation in the range domain is assumed to be 100 m (0.33 μs), the standard deviation of the azimuth measurements is assumed to be 3 degree and FDOA measurement standard deviation is assumed to be 1 m/s (assuming a carrier frequency of 3 GHz this corresponds to 10 Hz in the frequency domain, the carrier frequency is assumed to be known).
The results shown here are the product of 1000 Monte Carlo runs with a sampling interval of two seconds. A total of 120 s is regarded.
Firstly, the combined measurement set of TDOA and AOA measurements is processed using the MLE which is implemented with the simplex method due to Nelder and Mead as numerical iterative search algorithm. For each Monte Carlo run, the emitter state is computed once processing the complete measurement set. In comparison, a MLE only based on TDOA measurements is performed. This MLE is initialized with a point generated randomly from a Gaussian distribution centered at the true emitter location with standard deviation of 500 m in the x and y direction. The velocity entries are set to 10. In contrast, the intersection point of the first TDOA/AOA measurement pair provides the initialization point of the combined method. In Fig. 2 (b), the results of the combination of TDOA and AOA are labeled with the acronym TAOA MLE, the results of the TDOA approach with the acronym TDOA MLE.

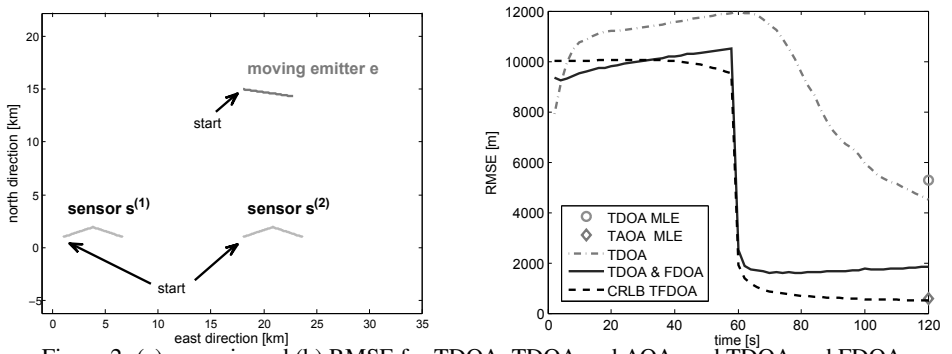


Figure 2: (a) scenario and (b) RMSE for TDOA, TDOA and AOA, and TDOA and FDOA

Secondly, the combination of TDOA and FDOA is exploited using the static GM filter described in [Kau09]. Initially, the first TDOA measurement is approximated by a GM in the Cartesian state space incorporating prior information about the area in which the emitter must lie. Additionally, a maximal emitter speed of 60 m/s in x and y direction is assumed. The updates are done using EKFs for the incoming TDOA and FDOA measurements. The performance of the combined filter (acronym TDOA & FDOA) is compared to the GM

filter only based on TDOA measurements (acronym TDOA).
In Fig. 2 (b), the Root Mean Square Errors (RMSE) of the different algorithms are plotted against the time in seconds. It shows the superior performance of the combined methods. Due to the initialization, the GM filter are in the initial phase better than the CRLB. CRLB for the TDOA/FDOA approach are shown with additional assumptions (CRLB TFDOA). In this scenario, the best results are obtained with the combination of TDOA and AOA.

6 Conclusions

For passive emitter tracking in sensor networks different measurement types can be obtained by exploiting the signal impinging from the target. Some of them can be taken by single sensors; e. g. bearing measurements. Others can only be collected by a network of sensors. A minimum of two sensors is needed here. TDOA and FDOA measurements belong to this group. FDOA measurements are highly dependent on the relative motion between emitter and sensors.
The combination of different measurements leads to a significant gain in estimation accuracy. In the investigated scenario, this gain is very high in the case of combining TDOA and AOA measurement information.

References

[BSLK01] Y. Bar-Shalom, X. Rong Li, and T. Kirubarajan. *Estimation with Applications to Tracking and Navigation*. Wiley-Interscience, 2001.

[Bec05] Klaus Becker. Three-dimensional Target Motion Analysis using Angle and Frequency Measurements. *IEEE Trans. on Aerospace and Electronic Systems*, 41(1): 284–301, 2005.

[HY08] K. C. Ho and Le Yang. On the Use of a Calibration Emitter for Source Localization in the Presence of Sensor Position Uncertainty. *IEEE Trans. on Signal Processing*, 56(12): 5758–5772, 2008.

[Kau09] R. Kaune. Gaussian Mixture (GM) Passive Localization using Time Difference of Arrival (TDOA). *Informatik 2009-Workshop Sensor Data Fusion: Trends, Solutions, Applications*: 2375–2381, 2009.

[KMK10] R. Kaune, D. Mušicki, W. Koch. On passive emitter tracking. chapter in *Sensor Fusion* sciyo.com: submitted for publication.

[MKK10] D. Mušicki, R. Kaune, W. Koch. Mobile Emitter Geolocation and Tracking Using TDOA and FDOA Measurements. *IEEE Trans. on Signal Processing*, 58(3), part 2: 1863–1874, 2010.

[MK08] D. Mušicki and W. Koch. Geolocation using TDOA and FDOA Measurements. *11th International Conference on Information Fusion*, Cologne, Germany, June 30 2008-July 3 2008.

[SHH08] So, Chan, Chan. Closed-Form Formulae for Time-Difference-of-Arrival Estimation. *IEEE Trans. on Signal Processing*, 56(6): 2614–2620, 2008.

[St81] S. Stein. Algorithms for Ambiguity Function Processing. *IEEE Trans. Acoustic, Speech and Signal Processing*, 29(3): 588–599, 1981.

Localization of Piecewise Curvilinearly Moving Targets Using Azimuth and Azimuth Rates

Julian Hörst and Marc Oispuu
Fraunhofer FKIE, Dept. Sensor Data and Information Fusion
Neuenahrer Str. 20, 53343 Wachtberg, Germany
{julian.hoerst, marc.oispuu}@fkie.fraunhofer.de

Abstract: Detailed knowledge of the estimation performance for passive tracking of maneuvering targets is of fundamental interest. Therefore, the maximum achievable localization accuracy for maneuvering targets from azimuth measurements as well as from combined azimuth and azimuth rate measurements are calculated in this paper. These targets perform a piecewise curvilinear motion with an unknown number of maneuvers at unknown times. It is proven that in addition to the azimuth angles also azimuth rates contain valuable information about the kinematics of maneuvering targets that can be exploited with advantage for state estimation.

1 Introduction

State estimation of an emitting source from passive bearing measurements collected by a single moving observer is a widely investigated problem. This problem is commonly referred to as Target Motion Analysis (TMA) [Bec01] and is encountered in various fields like wireless communications, as well as airborne radar and underwater sonar applications. Aspects of the TMA problem examined in the literature include bearings-only estimation algorithms, estimation accuracy, and target observability [Bec01, Bec96].

In many cases, the targets are not moving inertially (i.e. non-accelerated), but are partly strongly maneuvering. Commonly maneuvering targets can be characterized by the so-called *curvilinear motion* model described in [BN97]. This model assumes constant and simultaneously active tangential (i.e. along-track) and normal (i.e. cross-track) accelerations a_{t} and a_{n}. An approximate solution of the curvilinear motion equation has also been presented in [BN97] for the case that the relative change of velocity is much lower than 1. The evaluation of the Cramér-Rao bound (CRB) has been realized in [RA03, RAG04] with the limiting condition that the maneuver change-over times and the maneuvers are exactly known.

In [OH10], we considered maneuvering targets performing a curvilinear motion in each maneuver segment (see Fig. 1) known as the *piecewise curvilinear motion* model established by Becker [Bec05]. It is important to emphasize that the maneuver change-over times are unknown, i.e. these parameters have to be estimated.

In contrast to [OH10], where we only investigated azimuth measurements, we consider additional azimuth rate measurements in this work. The azimuth rate can be obtained,

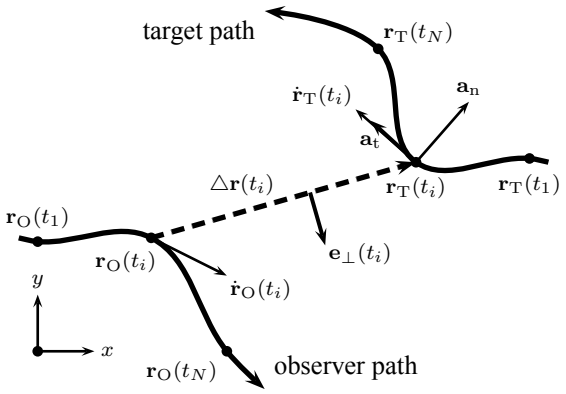


Figure 1: Scenario with an arbitrary moving sensor and a piecewise curvilinearly moving target.

e.g., by joint processing of sensor array signals. Since acquiring these quantities from raw signals is an estimation problem, the CRB of the azimuth angles and azimuth rates can be derived [WE95]. Recently, several estimation approaches have been proposed, e.g., in [MBS05] for a planar array to obtain azimuth angles, elevation angles and the corresponding rates.

2 Estimation problem

We consider the scenario depicted in Fig. 1. A maneuvering target moves along a trajectory $\mathbf{r}_\mathrm{T}(t) = (x(t), y(t))^T$ with velocity $\dot{\mathbf{r}}_\mathrm{T}(t) \in \mathbb{R}^{2\times 1}$ and constant tangential acceleration $a_\mathrm{t} = |\mathbf{a}_\mathrm{t}|$ and normal acceleration $a_\mathrm{n} = |\mathbf{a}_\mathrm{n}|$. Furthermore, a single observer moves along an arbitrary trajectory $\mathbf{r}_\mathrm{O}(t) \in \mathbb{R}^{2\times 1}$. The target observer geometry is given by the relative vector $\triangle\mathbf{r}(t) = \mathbf{r}_\mathrm{T}(t) - \mathbf{r}_\mathrm{O}(t)$ and its velocity $\triangle\dot{\mathbf{r}}(t) = \dot{\mathbf{r}}_\mathrm{T}(t) - \dot{\mathbf{r}}_\mathrm{O}(t)$. The observer's objective is to estimate the target state from passively measured line-of-sight azimuth angles α and azimuth rates $\dot{\alpha}$.

2.1 Motion model

The state of a target moving on a plane with constant tangential and normal acceleration can be completely described by the position components of $\mathbf{r}_\mathrm{T}(t)$, two velocity components of $\dot{\mathbf{r}}_\mathrm{T}(t)$ given by the velocity $v(t)$ and the course $\varphi(t)$, and the acceleration components a_t and a_n. The special cases of inertial motion ($a_\mathrm{t} = a_\mathrm{n} = 0$), straight-line acceleration ($a_\mathrm{t} \neq 0$, $a_\mathrm{n} = 0$), and circular motion ($a_\mathrm{t} = 0$, $a_\mathrm{n} \neq 0$) are included in this model. All target parameters are comprised in the parameter vector $\mathbf{x}(t) = (x(t), y(t), v(t), \varphi(t), a_\mathrm{t}, a_\mathrm{n})^T \in \mathbb{R}^{6\times 1}$. From the results of our previous work, we know that the exact solution of the motion equation has the form

$$\mathbf{x}(t) = \mathbf{f}[\mathbf{x}(t_0); t, t_0]\,, \tag{1}$$

where (1) describes the temporally evolution of the target state from t_0 to t. The derivation of the components in (1) can be found in [OH10]. Mentionable, the initial time t_0 can be replaced by any reference time $t_\mathrm{r} > t_0$. The state at t_r can be written shortly as $\mathbf{x}_\mathrm{r} = \mathbf{x}(t_\mathrm{r})$.

In the case of piecewise curvilinear motion, the dimension of the target state $\mathbf{x}(t)$ increases by three components with each maneuver change-over. That means, two acceleration components and the maneuver change-over time are added to the target state. With M maneuvers, the augmented target state is specified by

$$\mathbf{x}(t) = \left(x(t), y(t), v(t), \varphi(t), \mathbf{a}^T, \tilde{\mathbf{t}}^T\right)^T \tag{2}$$

with $\mathbf{a} = (a_{\mathrm{t},0}, a_{\mathrm{n},0}, ..., a_{\mathrm{t},M}, a_{\mathrm{n},M})^T \in \mathbb{R}^{2+2M\times 1}$ and $\tilde{\mathbf{t}} = (\tilde{t}_1, ..., \tilde{t}_M)^T \in \mathbb{R}^{M\times 1}$. Here, $\tilde{t}_m$ is the change-over time of the m-th maneuver and $a_{\mathrm{t},m}$ and $a_{\mathrm{n},m}$ denote the tangential and normal acceleration in the time interval $[\tilde{t}_{m-1}, \tilde{t}_m]$, $m = 1, ..., M$.

Similar to (1), the target state can be parameterized by the state at another time, e.g. by $\mathbf{x}_m = \mathbf{x}(\tilde{t}_m)$. Since the reference state is commonly the current state, the state for M maneuvers in the time interval $[t_0, t_\mathrm{r}]$ is given by

$$\mathbf{x}(t) = \begin{cases} \mathbf{f}[\mathbf{x}_1; t, \tilde{t}_1] & \text{for } t_0 \le t \le \tilde{t}_1 \\ \vdots & \vdots \\ \mathbf{f}[\mathbf{x}_m; t, \tilde{t}_m] & \text{for } \tilde{t}_{m-1} < t \le \tilde{t}_m \\ \vdots & \vdots \\ \mathbf{f}[\mathbf{x}_M; t, \tilde{t}_M] & \text{for } \tilde{t}_{M-1} < t \le \tilde{t}_M \\ \mathbf{f}[\mathbf{x}_\mathrm{r}; t, t_\mathrm{r}] & \text{for } \tilde{t}_M < t \le t_\mathrm{r} \end{cases}, \tag{3}$$

where the state at some arbitrary time t is related to the reference state by

$$\mathbf{f}[\mathbf{x}_m; t, \tilde{t}_m] = \mathbf{f}[\mathbf{f}[\cdots \mathbf{f}[\mathbf{x}_\mathrm{r}; t, t_\mathrm{r}]; \cdots]; t, \tilde{t}_m]\,. \tag{4}$$

2.2 Measurement model

For the sake of simplicity, we assume that the detection probability is equal to 1 and the false alarm rate is equal to 0. The measured azimuth α_m and azimuth rate $\dot{\alpha}_\mathrm{m}$ at time t_i, $i = 1, ..., N$, are given by

$$\begin{aligned} \alpha_\mathrm{m}(t_i) &= \alpha(t_i) + w_\alpha(t_i)\,, \\ \dot{\alpha}_\mathrm{m}(t_i) &= \dot{\alpha}(t_i) + w_{\dot{\alpha}}(t_i)\,, \end{aligned} \tag{5}$$

where $w_\alpha(t_i)$ and $w_{\dot{\alpha}}(t_i)$ denote the measurement error, and

$$\begin{aligned} \alpha(t) &= \arctan \frac{\triangle x(t)}{\triangle y(t)}\,, \\ \dot{\alpha}(t) &= \frac{\triangle \dot{x}(t)\triangle y(t) - \triangle x(t)\triangle \dot{y}(t)}{\triangle x^2(t) + \triangle y^2(t)} = \frac{\triangle \dot{\mathbf{r}}^T(t)\,\mathbf{e}_\perp(t)}{\triangle r(t)} \end{aligned} \tag{6}$$

indicate the true azimuth angle and the true azimuth rate. In (6), $\triangle\dot{\mathbf{r}}(t) = (\triangle\dot{x}(t), \triangle\dot{y}(t))^T$, $\mathbf{e}_\perp(t)$ is a unit vector orthogonal to the relative vector $\triangle\mathbf{r}(t) = (\triangle x(t), \triangle y(t))^T$, and $\triangle r(t) = |\triangle\mathbf{r}(t)|$ denotes the distance between observer and target (Fig. 1). The observer state (position $\mathbf{r}_\mathrm{O}(t)$ and velocity $\dot{\mathbf{r}}_\mathrm{O}(t)$) is assumed to be exactly known. With this, the equations in (6) only depend on the target state.

We assume that the azimuth and azimuth rate measurements are independent of each other, that the measurement noise vectors are zero-mean Gaussian, and that the measurement covariances read $\mathbf{W}_\alpha = \sigma_\alpha^2\,\mathbf{I}_N$ and $\mathbf{W}_{\dot\alpha} = \sigma_{\dot\alpha}^2\,\mathbf{I}_N$. Here, σ_α^2 and $\sigma_{\dot\alpha}^2$ denote the constant noise variances and $\mathbf{I}_N$ denotes the $N \times N$-dimensional identity matrix. We note that in practice, the variances may change from time to time and have to be estimated.

With the previous considerations, the problem can be stated as follows: Estimate the target state $\mathbf{x}_\mathrm{r}$ at some reference time t_r from all measurements. We consider two measurement sets, only azimuth measurements and both azimuth and azimuth rate measurements.

3 Cramér-Rao bound (CRB)

The CRB provides a lower bound on the estimation accuracy of any unbiased estimator and its parameter dependencies reveal characteristic features of the estimation problem. Let $\mathbf{x}_\mathrm{r}$ denote an unknown parameter vector and let $\hat{\mathbf{x}}_\mathrm{r}(\boldsymbol{\psi}_\mathrm{m})$ be some unbiased estimate of $\mathbf{x}_\mathrm{r}$ based on the measurement set $\boldsymbol{\psi}_\mathrm{m}$. The CRB is given by the inverse Fisher Information Matrix (FIM)

$$\mathbf{J}_\psi(\mathbf{x}_\mathrm{r}) = \mathrm{E}\left\{\left(\frac{\partial\mathcal{L}(\boldsymbol{\psi}_\mathrm{m};\mathbf{x}_\mathrm{r})}{\partial\mathbf{x}_\mathrm{r}}\right)^T\left(\frac{\partial\mathcal{L}(\boldsymbol{\psi}_\mathrm{m};\mathbf{x}_\mathrm{r})}{\partial\mathbf{x}_\mathrm{r}}\right)\right\}, \tag{7}$$

where $\mathrm{E}\{\cdot\}$ denotes the expectation operation and

$$\mathcal{L}(\boldsymbol{\psi}_\mathrm{m};\mathbf{x}_\mathrm{r}) = -\frac{1}{2}\ln\left(\det(2\pi\mathbf{W}_\psi)\right) - \frac{1}{2}(\boldsymbol{\psi}_\mathrm{m} - \boldsymbol{\psi}(\mathbf{x}_\mathrm{r}))^T\,\mathbf{W}_\psi^{-1}(\boldsymbol{\psi}_\mathrm{m} - \boldsymbol{\psi}(\mathbf{x}_\mathrm{r})) \tag{8}$$

is the log-likelihood function. Inserting (8) into (7), performing the expectation operation, and using the noise covariances in Section 2.2, we obtain the FIMs

$$\begin{aligned}\mathbf{J}_\alpha(\mathbf{x}_\mathrm{r}) &= \frac{1}{\sigma_\alpha^2}\sum_{i=1}^{N}\left(\frac{\partial\alpha(t_i)}{\partial\mathbf{x}_\mathrm{r}}\right)^T\frac{\partial\alpha(t_i)}{\partial\mathbf{x}_\mathrm{r}},\\ \mathbf{J}_{\dot\alpha}(\mathbf{x}_\mathrm{r}) &= \frac{1}{\sigma_{\dot\alpha}^2}\sum_{i=1}^{N}\left(\frac{\partial\dot\alpha(t_i)}{\partial\mathbf{x}_\mathrm{r}}\right)^T\frac{\partial\dot\alpha(t_i)}{\partial\mathbf{x}_\mathrm{r}}\end{aligned} \tag{9}$$

for the measurements of the azimuth ($\boldsymbol{\psi}_\mathrm{m} = \boldsymbol{\alpha}_\mathrm{m}$) and the azimuth rate ($\boldsymbol{\psi}_\mathrm{m} = \dot{\boldsymbol{\alpha}}_\mathrm{m}$). It is important to emphasize that the above given FIMs denote the maximum achievable information at an arbitrary reference time t_r based on N measurements. Finally, the CRBs read $\mathbf{J}_\alpha^{-1}(\mathbf{x}_\mathrm{r})$ and $(\mathbf{J}_\alpha(\mathbf{x}_\mathrm{r}) + \mathbf{J}_{\dot\alpha}(\mathbf{x}_\mathrm{r}))^{-1}$ for the azimuth-only case and the case using both azimuth and azimuth rate.

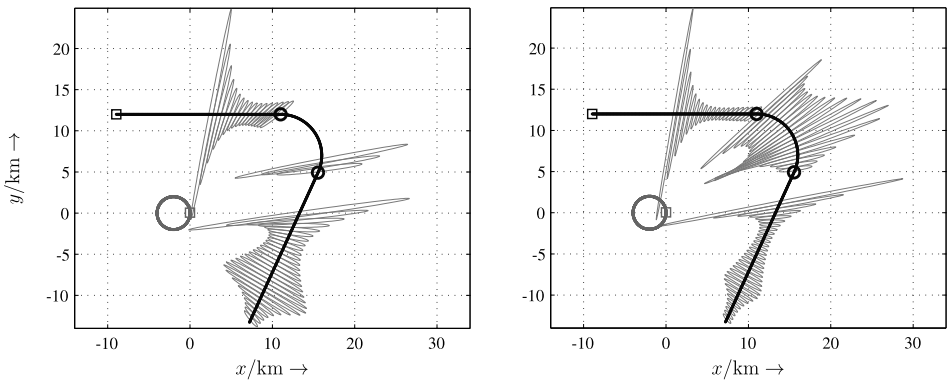


Figure 2: Observer path (green) with starting point (green square); target trajectory (blue) with target starting point (blue square) and target maneuver change-over points (blue circles). *Left*: CRB for the azimuth-only case (red ellipses). *Right*: CRB for combined azimuth and azimuth rate measurements (red ellipses).

4 Simulation results

We consider the 2D scenario in Fig. 2. The target starts with the initial state $(-9\,\mathrm{km}, 12\,\mathrm{km}, 50\,\mathrm{m/s}, 90^\circ)^T$ and performs the following maneuvers:

$$
\begin{aligned}
(t_0, a_{\mathrm{t},0}, a_{\mathrm{n},0}) &= (0\,\mathrm{s}, 0\,\mathrm{m/s^2}, 0\,\mathrm{m/s^2})\,, \\
(\tilde{t}_1, a_{\mathrm{t},1}, a_{\mathrm{n},1}) &= (400\,\mathrm{s}, 0\,\mathrm{m/s^2}, 0.5\,\mathrm{m/s^2})\,, \\
(\tilde{t}_2, a_{\mathrm{t},2}, a_{\mathrm{n},2}) &= (600\,\mathrm{s}, 0\,\mathrm{m/s^2}, 0\,\mathrm{m/s^2})\,.
\end{aligned}
$$

The observer moves counterclockwise along a circular path with constant velocity. This is parameterized by $\mathbf{r}_\mathrm{O}(t_0) = (0\,\mathrm{km}, 0\,\mathrm{km})^T$, $|\dot{\mathbf{r}}_\mathrm{O}(t_0)| = 50\,\mathrm{m/s}$, $\varphi_\mathrm{O}(t_0) = 0^\circ$, and $a_{\mathrm{O,n}} = -1.25\,\mathrm{m/s^2}$. The sensor collects one measurement per second, where the measurement noise is zero-mean Gaussian distributed with covariances given in Section 2.2. The corresponding standard deviations are $\sigma_\alpha = 2^\circ$ and $\sigma_{\dot{\alpha}} = 1\,\mathrm{mrad/s}$ which are similar to the assumptions in [W+08].

In Fig. 2, the results of the Cramér-Rao analysis are given. The bounds are illustrated by means of $90\,\%$ confidence ellipses. In the left part of the figure the CRB for the azimuth-only case are depicted, whereas in the right part of the figure the CRB for the combination of azimuth and azimuth rate measurements are shown. Note that for visualization purposes, ellipses have been drawn every $10\,\mathrm{s}$ and ellipses with an extent of over $30\,\mathrm{km}$ have been omitted.

It can be recognized that for all three legs of the target motion the estimation accuracy for the combined azimuth and azimuth rate case (right) is higher and converges faster than for azimuth-only measurements (left). Especially for the turn motion in the middle leg, reasonable estimation accuracies can be given for the combined azimuth/azimuth rate case, whereas azimuth-only measurements provide only inferior accuracies. Additionally, in Fig. 2, a decreasing estimation accuracy can be seen after a maneuver has taken place which is due to the dimension change of the target state at this time.

5 Conclusions

The considered target motion model subsumes the models described in the literature. We have presented an exact solution of the corresponding motion equation and have derived the CRB for the case that the maneuver accelerations and change-over times are unknown. In a Cramér-Rao analysis, we have found that the extension of the target state by acceleration components and change-over times leads to a declined estimation accuracy. Nevertheless, the achievable estimation accuracy can be significantly improved by using additional azimuth rate measurements.

References

[Bec96] K. Becker. A General Approach to TMA Observability from Angle and Frequency Measurements. *IEEE Trans. Aerosp. Electron. Syst.*, 32:487–494, January 1996.

[Bec01] K. Becker. Target Motion Analysis (TMA). In S. Stergioulos, editor, *Advanced Signal Processing Handbook*, chapter 9, pages 284–301. CRC Press, New York, NY, 2001.

[Bec05] K. Becker. Passive Aufklärung manövrierender Ziele aus Winkel- und Frequenzmessungen. FKIE Report 95, Research Establishment of Applied Science (FGAN), Wachtberg, April 2005. In German language.

[BN97] R. A. Best and J. P. Norton. A New Model and Efficient Tracker for a Target with Curvilinear Motion. *IEEE Trans. Aerosp. Electron. Syst.*, 33:1030–1037, July 1997.

[MBS05] W. Montlouis, O. Bayat, and B. Shafai. DOA and Angular velocity estimation using planar array with subspace based initialization. In *Proc. IEEE Military Communications Conference (MILCOM)*, pages 2797 – 2801, Atlantic City, NJ, October 2005.

[OH10] M. Oispuu and J. Hörst. Azimuth-only Localization and Accuracy Study for Piecewise Curvilinearly Moving Targets. In *Proc. ISIF 13th International Conference on Information Fusion*, Edinburgh, UK, July 2010. To be published.

[RA03] B. Ristic and M. S. Arulampalam. Tracking a manoeuvring target using angle-only measurements: algortihms and performance. *IEEE Trans. Signal Processing*, 83:1223–1238, 2003.

[RAG04] B. Ristic, M. S. Arulampalam, and N. Gordon. *Beyond the Kalman Filter: Particle Filters for Tracking Applications*. Artech House, Boston, MA, 2004.

[W+08] J. Wang et al. Passive Location to Moving Emitter Using DOA and Its Rate of Change. *Congress on Image and Signal Processing 2008 (CISP'08)*, 4, 2008.

[WE95] T. Wigren and A. Eriksson. Accuracy Aspects of DOA and Angular Velocity Estimation in Sensor Array Processing. *IEEE Signal Processing Lett.*, 2:60–62, April 1995.

Robust Likelihood Function for Blind Mobile Terminal Localization

Vadim Algeier, Reiner Thomä, Wolfgang Koch
vadim.algeier@tu-ilmenau.de
reiner.thomae@tu-ilmenau.de
Wolfgang.Koch@fkie.fraunhofer.de

Abstract: The focus of this contribution is put on the blind localization of mobile terminals in urban scenarios. The proposed localization method exploits multipath propagation, which is typical of urban terrain. It is assumed that each multipath component is characterized by its direction of arrival (DoA) and relative time of arrival (RToA) parameter. In essence the developed method compares the multipath parameters predicted by the ray tracing algorithm with those measured by the single observing station. In this context inevitably occur a multipath order problem, which is tackled in detail within a proposed likelihood function definition. The latter follows an approach known from the robust statistics. Its intention is to provide a robust and reliable position estimate having at hand minimum a priori information about system parameters.

1 Introduction

There is a rapid growth of wireless applications that require the knowledge of the mobile terminal's location [KH06]. In most cases the cooperative position estimation methods [GG05] can be used. In this contribution, however, we concentrate on the *blind mobile localization* (BML), which presumes no cooperation of the mobile terminal referred to as *mobile station* (MS) with the *observing station* (OS). This problem is typical of non-subscribed user localization, e.g. in emergency, security, and safety applications [KST06].

In the preceding paper [ADKT08] we formulated the boundary conditions and proposed a possible solution of the BML task. Its goal is to develop a method for geo locating of a MS in a non-cooperative mode, i.e. solely from the signals radiated by the MS, implying that neither MS nor the cellular infrastructure are involved in the positioning process. The idea of the proposed BML method implies a correlation of measured direction of arrival (DoA) and relative time of arrival (RToA) parameters with data pre-calculated by the ray tracing (RT) analysis [Mau05]. Formulation of an appropriate correlation criterion is a key task of the underlying likelihood function and turns into a challenge in the noisy case. Since the measurement process induces missing detections of the true propagation paths or conversely produces the false multipaths. That is, the number of true multipaths as well as the corresponding multipath parameter are distorted, which makes the unique mapping impossible. In particular, the *multipath model order* problem, i.e. a need to compare hypothetical positions with different number of predicted multipaths, requires careful consideration. In [ADKT08] the probabilistically motivated likelihood function was proposed. It relies on

the known error statistics, i.e. false alarm rate and detection probability. The likelihood definition presented in this work does not need a priori knowledge about clutter parameter, which is a more realistic assumption when working with real-life data. Furthermore, it is able to cope with the multipath model order problem.

The remainder of the paper is organized as follows. In section 2 we present the measurement error model and in section 3 the robust version of the likelihood function for the BML is introduced.

2 Measurement model

We start with introduction of frequently used symbols. Let $\boldsymbol{\xi}_{\mathrm{O}}$ denote the known OS position. The MS position $\boldsymbol{\xi}_{\mathrm{M}}$ is unknown and must be estimated. For this purpose we introduce some hypothetical MS position $\boldsymbol{\xi}$, which can be arbitrarily chosen in the particular scenario. All positions are specified in a 2D Cartesian space as follows:

$$\boldsymbol{\xi}_{\mathrm{O}} = \begin{bmatrix} x_{\mathrm{O}} \\ y_{\mathrm{O}} \end{bmatrix}, \quad \boldsymbol{\xi} = \begin{bmatrix} x \\ y \end{bmatrix}, \quad \boldsymbol{\xi}_{\mathrm{M}} = \begin{bmatrix} x_{\mathrm{M}} \\ y_{\mathrm{M}} \end{bmatrix}. \tag{1}$$

Let $\mathbf{h}\left(\boldsymbol{\xi}_{\mathrm{O}}, \boldsymbol{\xi}\right)$ be the function, which represents the result of the RT analysis. For particular transmitter position $\boldsymbol{\xi}$ and receiver poisition $\boldsymbol{\xi}_{\mathrm{O}}$ function $\mathbf{h}$ gives, i.e. RT analysis predicts, a set of multipath parameters consisting of $M_{\boldsymbol{\xi}}$ multipath components:

$$\mathbf{h}\left(\boldsymbol{\xi}_{\mathrm{O}}, \boldsymbol{\xi}\right) = \left\{\mathbf{h}^{m}\left(\boldsymbol{\xi}_{\mathrm{O}}, \boldsymbol{\xi}\right)\right\}_{m=1}^{M_{\boldsymbol{\xi}}}. \tag{2}$$

$\mathbf{h}^{m}\left(\boldsymbol{\xi}_{\mathrm{O}}, \boldsymbol{\xi}\right)$ contains parameters describing the m-th multipath. Subscript $\boldsymbol{\xi}$ in $M_{\boldsymbol{\xi}}$ indicates that the number of predicted multipaths can be different for different hypothetical MS positions. For the sake of notation simplicity we use M instead of $M_{\boldsymbol{\xi}}$ whenever there is no danger of ambiguity. For the same reason an explicit specification of OS location can be ignored, i.e. in the sequel we use $\mathbf{h}_{\boldsymbol{\xi}}^{m}$ instead of $\mathbf{h}^{m}\left(\boldsymbol{\xi}_{\mathrm{O}}, \boldsymbol{\xi}\right)$. $\mathbf{h}_{\boldsymbol{\xi}}^{m}$ has the following structure:

$$\mathbf{h}_{\boldsymbol{\xi}}^{m} = \begin{bmatrix} l_m & \varphi_m \end{bmatrix}^{\mathrm{T}}, \tag{3}$$

i.e. each multipath component is specified by its relative path length and azimuth DoA. Azimuth DoA is in general bounded by:

$$\varphi_m \in \left[-\pi, \pi\right]. \tag{4}$$

l_m is a meter valued parameter, which relates to (the actually observable) *relative* or *excess delay* by:

$$l_m = \tau_m \cdot c_{\mathrm{Light}}. \tag{5}$$

c_{Light} denotes the speed of light. We use l_m instead of τ_m since it is more illustrative and convenient to use a meter valued parameter, while characterizing position estimate. The relative path length l_m is obtained from the complete path length l'_m, by:

$$l_m = l'_m - \min\left(l'_1, l'_2, \ldots, l'_M\right). \tag{6}$$

In this way l_m's are calculated from l'_m's also within RT analysis, since latter provides full multipath lengths. With definitions made above l_m is bounded by:

$$l_m \in [0, l_{\max}], \tag{7}$$

where $l_{\max} = \max(l_1, l_2, \ldots, l_M)$ relates to *total excess delay* $\tau_{\max}$. Let us now define the measured multipath set $\mathbf{z}$ consisting of K multipath components:

$$\mathbf{z} = \left\{\mathbf{z}^k\right\}_{k=1}^{K}. \tag{8}$$

Each measured multipath component $\mathbf{z}^k$ is characterized by its DoA and RToA parameter in the same manner as in the case of predicted multipath from (3). The measurement process incorporates different types of errors, which are introduced in detail in [Alg10] and will be briefly recapitulated here. On the one hand, there are false alarms and missing detections of true multipaths. That is, even if we would know the true MS position $\boldsymbol{\xi}_{\mathrm{M}}$ the number of measured multipaths K and predicted multipaths $M_{\boldsymbol{\xi}_{\mathrm{M}}}$ will be in general different. On the other hand, there is additive Gaussian distortion, which can lead to misassociations especially when the multipaths are closely spaced. All those errors can significantly deteriorate the localization accuracy or may result in completely wrong coordinates. Therefore, a careful choice of an underlying likelihood function is essential.

3 M-type likelihood function

As already mentioned in the introduction we wish to develop a likelihood criterion, which, firstly, requires minor statistical information about the clutter and, secondly, is able to cope with the multipath order problem. Let us demonstrate the problems arising, when the classical definition of the likelihood function is straightforwardly applied to the single observation:

$$\mathcal{L}(\boldsymbol{\xi}) = \prod_{k=1}^{K} \mathcal{N}\left(\mathbf{z}^k; \mathbf{h}_{\boldsymbol{\xi}}^k, \mathbf{C}^k\right). \tag{9}$$

Hereby $\mathbf{C}^k$ denotes the measurement covariance matrix of the k-th multipath. With noise variances $\sigma_{l_k}^2, \sigma_{\varphi_k}^2$ corresponding to relative path length and DoA respectively, $\mathbf{C}^k$ is defined as follows:

$$\mathbf{C}^k = \mathrm{diag}\left[\sigma_{l_k}^2, \sigma_{\varphi_k}^2\right]. \tag{10}$$

The values of the noise variances are known and depend on the array configuration, system bandwidth, SNR. They are typically different for every measured multipath. However, for the sake of simplicity, we assume equal variances for all multipaths.

According to the *maximum likelihood estimation* principle the location maximizing $\mathcal{L}(\boldsymbol{\xi})$ is the most likely to be the correct MS position estimation, i.e. the corresponding *maximum*

likelihood estimator (MLE) is expressed as follows:

$$\begin{aligned}\widehat{\boldsymbol{\xi}}_{\mathrm{M}}^{MLE} &= \underset{\boldsymbol{\xi}}{\operatorname{argmax}} \left(\prod_{k=1}^{K} \mathcal{N}\left(\mathbf{z}^k; \mathbf{h}_{\boldsymbol{\xi}}^k, \mathbf{C}^k\right) \right) \\ &\propto \underset{\boldsymbol{\xi}}{\operatorname{argmin}} \left(\sum_{k=1}^{K} \frac{1}{2} \left(\mathbf{h}_{\boldsymbol{\xi}}^k - \mathbf{z}^k\right)^{\mathrm{T}} \left(\mathbf{C}^k\right)^{-1} \left(\mathbf{h}_{\boldsymbol{\xi}}^k - \mathbf{z}^k\right) \right) \\ &= \underset{\boldsymbol{\xi}}{\operatorname{argmin}} \left(\sum_{k=1}^{K} \sum_{i=1}^{R} \frac{1}{2} \frac{\left(\mathbf{h}_{\boldsymbol{\xi}}^k(i) - \mathbf{z}^k(i)\right)^2}{\mathbf{C}^k(i,i)} \right) = \underset{\boldsymbol{\xi}}{\operatorname{argmin}} \left(\sum_{k=1}^{K} \sum_{i=1}^{R} \frac{\left(u_{\boldsymbol{\xi}}^{k,i}\right)^2}{2} \right), \quad (11)\end{aligned}$$

where R denotes the number of parameters describing a single multipath. In our case $R = 2$, however, it changes if further parameters e.g. elevation DoA or Doppler shift are considered. $\mathbf{C}^k(i,i)$ addresses the i-th diagonal element, i.e. the variance of the i-th multipath parameter characterizing the k-th multipath component, while $\mathbf{h}_{\boldsymbol{\xi}}^k(i)$ addresses the i-th entry of the vector $\mathbf{h}_{\boldsymbol{\xi}}^k$. $u_{\boldsymbol{\xi}}^{k,i} = \frac{\left(\mathbf{h}_{\boldsymbol{\xi}}^k(i) - \mathbf{z}^k(i)\right)}{\sqrt{\mathbf{C}^k(i,i)}}$ is the i-th normalized residual of the k-th multipath component for the hypothesized MS position $\boldsymbol{\xi}$.

The aim is to find position $\widehat{\boldsymbol{\xi}}_{\mathrm{M}}$ so that the corresponding multipath parameters $\mathbf{h}_{\widehat{\boldsymbol{\xi}}_{\mathrm{M}}}$ generated via RT yield the best match with the measured parameters $\mathbf{z}$. However, the likelihood function from (9), i.e. the estimator from (11), shows some deficiency in practical use. Firstly, the predicted multipath parameters must be associated with the measured ones, since the relation between them is unknown. Secondly, (9) is inapplicable if the number of predicted multipaths $M_{\boldsymbol{\xi}}$ at the particular position $\boldsymbol{\xi}$ is not equal to the number of measured multipaths K. Whereas in the case with $K < M_{\boldsymbol{\xi}}$ simply the subset of predicted multipaths with the highest weight can be chosen, the opposite case with $K > M_{\boldsymbol{\xi}}$ is more crucial. Typically, the joint weight $\mathcal{L}(\boldsymbol{\xi})$ diminishes with the number of factors considered, since the individual weights are mostly smaller than 1. Such behavior inevitably causes a higher joint weight for positions with smaller $M_{\boldsymbol{\xi}}$ penalizing those with higher $M_{\boldsymbol{\xi}}$ values. However, this fact counteracts the idea that the position with the highest number of associations is most likely to be the correct one. Thirdly, the product structure of $\mathcal{L}(\boldsymbol{\xi})$ makes it very sensitive to small values of the individual weights. Hence a single outlier with the vanishing individual weight decreases drastically the joint weight even if the remaining individual weights are significant.

The first point addressed as *data association* problem can be seen as an independent task and is not discussed in this contribution. The interested reader is referred to [Alg10]. Instead we concentrate in the following on the remaining two points and define the modified likelihood criterion inspired by the concepts of robust statistics.

Classical statistical methods rely heavily on model assumptions which are often not met in practice. E.g. if there are outliers in the data or if the assumption of the noise signal distribution (typically Gaussian) is violated, classical methods often have very poor performance. Robust statistics seeks to provide concepts which describe the behavior of statistical procedures not only under strict parametric models, but also in neighborhoods of such models, see [Hub81]. *M-estimator* ("M" stands for "maximum likelihood-type") is a

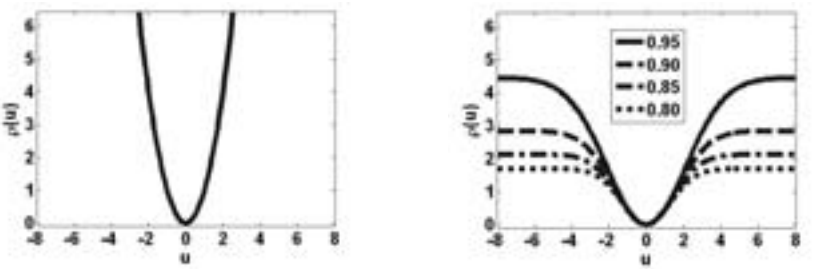
Table 1: Two possible ρ-functions for M-estimators.

Name	$\rho(u)$
Least squares	$u^2/2$
Welsh	$c_\alpha^2/2\left(1-\exp\left(-\left(u/c_\alpha\right)^2\right)\right)$

popular robust technique, which generalizes maximum likelihood estimation from (11) to:

$$\widehat{\boldsymbol{\xi}}_{\mathrm{M}}^{ME}=\underset{\boldsymbol{\xi}}{\operatorname{argmin}}\left(\sum_{k=1}^{K}\sum_{i=1}^{R}\rho\left(u_{\boldsymbol{\xi}}^{k,i}\right)\right). \tag{12}$$

where ρ is a symmetric, positive-definite function, i.e. $\rho(u)>0$ for all u accept 0, with a unique minimum at zero, i.e. $\rho(0)=0$. The idea is to choose ρ in such a way to provide the estimator desirable properties in terms of bias and efficiency. There is a wide choice of underlying functions, see [Rey83]. Figure 1 and table 1 present two possible alternatives for one dimensional case. The first candidate in table 1 is the least squares estimator, which corresponds to the maximum likelihood case from (11). Note, that the ρ-function of a least squares estimator is not bounded and residuals cause its quadratic increase, which explains the strong effect of outliers, see figure 1(a).



(a) Least squares. (b) Welsh; with varying asymptotic efficiency.

Figure 1: Shapes of ρ-functions.

The second candidate is the Welsh function, which is the Gaussian-like curve turned upside down, thus ensuring absolute suppression of outliers lying far from zero, see figure 1(b). This attractive property represents a decisive factor by the choice of an appropriate ρ-function in the context of BML problem. It allows, furthermore, to compare MS locations with different number of detected multipaths.

Using the Welsh function and applying standard statistical calculation the *M-type* likelihood function can be expressed as follows:

$$\mathcal{L}\left(\boldsymbol{\xi}\right)_{W}^{ME}=\prod_{k=1}^{K}\prod_{i=1}^{R}\frac{\exp\left(-\rho\left(u_{\boldsymbol{\xi}}^{k,i}\right)\right)}{\sqrt{2\pi\mathbf{C}^{k}(i,i)}}\propto\prod_{k=1}^{\min(K,M_{\boldsymbol{\xi}})}\exp\left(\frac{c_\alpha^2}{2}\sum_{i=1}^{R}\exp\left(-\left(\frac{u_{\boldsymbol{\xi}}^{k,i}}{c_\alpha}\right)^2\right)\right). \tag{13}$$

The tuning constant c_α used in the definition of Welsh function allows to adapt the underlying *M-estimator*, i.e. make it either more robust or more efficient. Further details and the derivation of (13) are presented in the full paper.

4 Conclusions

We presented the robust likelihood function for the *blind mobile localization* task. The likelihood function provides a non-linear mapping between the observation and the state space exploiting the knowledge of specific sensor properties combined with RT prediction, thus giving an answer to the fundamental question, how likely a particular hypothetical MS location is to produce an observed set of multipath components. In essence, the underlying likelihood function evaluates the proximity of the measured and predicted multipath components. A special feature of the developed likelihood criterion is that it is able to tackle the *multipath order problem* inevitably arising when positions with different number of multipath components are compared. Furthermore, it provides a robust and reliable position estimate having at hand minimum a priori information about the clutter parameters.

References

[ADKT08] V. Algeier, B. Demissie, W. Koch, and R. Thomä. State Space Initiation for Blind Mobile Terminal Position Tracking. *EURASIP Journal on Advances in Signal Processing*, 2008:14, 2008. Article ID 394219, 14 pages.

[Alg10] Vadim Algeier. *Blind localization of mobile terminals in urban scenarios*. PhD thesis, Ilmenau University of Technology, 2010.

[GG05] F. Gustafsson and F. Gunnarsson. Mobile positioning using wireless networks: possibilities and fundamental limitations based on available wireless network measurements. *IEEE Signal Processing Magazine*, 22(4):41–53, July 2005.

[Hub81] Peter J. Huber. *Robust Statistics*. New York: John Wiley & Sons, 1981.

[KH06] K. W. Kolodziej and J. Hjelm. *Local Positioning Systems: LBS Applications and Services*. Taylor & Francis Group, New York, NY, USA, 2006.

[KST06] Shohei Kikuchi, Akira Sano, and Hiroyuki Tsuji. Blind Mobile Positioning in Urban Environment Based on Ray-Tracing Analysis. *EURASIP Journal on Applied Signal Processing*, 2006:12, 2006. Article ID 38989, 12 pages.

[Mau05] Jürgen Maurer. *Strahlenoptisches Kanalmodell für die Fahrzeug-Fahrzeug-Funkkommunikation*. PhD thesis, Universität Karlsruhe, 2005.

[Rey83] W. J. J. Rey. *Introduction to Robust and Quasi-Robust Statistical Methods*. Berlin Heidelberg: Springer-Verlag, 1983.

Localization of Radioactive Source Carriers in Person Streams

Monika Wieneke
Fraunhofer FKIE, Department Sensor Data and Information Fusion
Neuenahrer Str. 20, 53343 Wachtberg, Germany
monika.wieneke@fkie.fraunhofer.de

Abstract: The localization and tracking of radioactive sources in public facilities like airports or stations is a problem of highest security relevance. The accumulation and the severity of terrorist attacks during the past decade give reason to the assumption that future attacks could also involve radioactive material packaged with conventional explosives. The only way to avoid such kind of attacks is to localize and arrest the person carrying the material to its destination. But since radiation is not perceivable by human beings, the security guards are largely dependent on technical decision support to perform this task. We consider a security assistance system comprising three gamma scintillation detectors that are distributed along a corridor wall to check passing people for radioactive material. Furthermore, the system consists of a set of tracking sensors simultaneously providing the positions of all persons during their walk through the corridor. In this paper we propose and evaluate techniques to estimate the assignment of radioactive detections to person tracks. These techniques provide a measure for each person reflecting the probability that the person is a radioactive source carrier.

1 Introduction

In the context of intelligent surveillance of public places, the observation and analysis of persons by distributed sensor systems increasingly gains in importance. The detection of hazardous material in busy areas as well as its assignment to a person is a challenging task that cannot be performed without technical decision support. However, the application of conventional technologies and the corresponding courses of action lead to long waiting times and pressure of work for the security personnel. This situation can be extremely relieved by security assistance systems that continuously observe an area by distributed sensor systems, that call the security guards only in case of a detection and that finally give a hint to those persons who can be assumed to carry the detected source.

In this work we concern ourselves with the localization of radioactive source carriers in person streams. The discussions about potential substances used for terrorist attacks are not only coined by the already applied improvised explosive devices (IED) but also by the fear of improvised nuclear devices, better known as dirty bomb or radioactive dispersion device (RDD). An RDD consists of a conventional explosive wrapped up with radioactive material. The conventional explosive conduces to disperse the radioactive material in the environment. Although this type of threat has not been put into practice so far, of growing

concern are numerous accidents involving a loss or theft of radioactive sources that could possibly be used for a dirty bomb. Hence, there is an increasing need for security assistance systems that are able to localize such material either on the way to the creation place of the bomb or already packaged on the way to its detonation place.

In this work we consider the transportation of radioactive material by a person walking through a public facility. In such a scenario a security assistance system for source carrier localization is ideally equipped with multiple sensors of complementary type. We propose a combination of scintillation detectors for radiation detection with tracking sensors for determining the positions of the persons. While the strength of radiation detectors lies in their detection capability, their substantial weakness is given by a limited spatio-temporal resolution capability. Hence, a single detector is not able to reliably localize the source and to assign it to a person, whereas tracking sensors enable a precise localization of all persons and thus reduce the search space to a countable set of potential source positions. A security assistance system combining sensors for hazardous substances with tracking data has been first proposed by Wieneke and Koch [WK09]. Within such a kind of sensor system localization means the calculation of an assignment probability between a series of radiation detections and each person track. The decision whether a person is a source carrier or not can thus be interpreted as a classification task.

We will propose three techniques to estimate the assignment between a detection series and a person track and evaluate their capability of finding the source carrier. For this first analysis we completely exclude potential position uncertainties in the tracking data and assume the positions of the persons inside the surveillance area to be exactly known. Persons are considered as point objects.

2 Measurement Process and Sensor Model

The radiation strength of a radioactive source is called activity [VS07]. The activity α of a source is defined as the number of radioactive decays per time unit. The SI unit[1] of activity is Becquerel (Bq). One Bq corresponds to one decay per second. From a statistical point of view the activity is the expected value of the number of decays per time unit. The actual number of decays in a certain time interval randomly deviates from the expected value. The frequency of the numbers follows a Poisson distribution. Let a_k be the number of decays during the current time interval k. The Poisson distribution is a discrete probability distribution that assigns probabilities $\mathcal{P}_\alpha(X = a_k) = \exp(-\alpha)\alpha^{a_k}/a_k!$ to numbers $a_k \in \mathbb{N}_0$.

There are three main types of radiation: alpha, beta and gamma radiation, listed by their increasing capability of penetrating matter. Gamma rays can cause serious damage when absorbed by living tissue, and are thus a health hazard. A gamma scintillation detector counts the number of emitted gamma particles that hit the detector surface. The number of emitted gamma particles per decay is given by the decay scheme of the radiator. For the sake of simplicity, we assume a material with one gamma emission per decay.

[1] SI – International System of Units

The strength of radioactive radiation at a detector r is inversely proportional to the square of the distance d_r from the detector to the source (inverse square law). In other words, the emitted particles are equally distributed on the surface of a sphere with radius d_r and the area A_r of the absorbing part of the detector marks a section of this surface. Besides the source particles, the detector counts particles of the background radiation with rate α_B.

For a stationary source the relation between the measured counts c_r at detector r and the source activity is hence given by Eq. (1), where $\theta = [x, y, \alpha]^\mathsf{T}$ is the source parameter vector with position $[x, y]^\mathsf{T}$ and activity α.

$$\mathcal{P}_{\lambda(\theta)}(X = c^r) \qquad \text{with} \qquad \lambda(\theta) = \alpha_B + \alpha \cdot \frac{A_r}{4\pi d_r^2} \tag{1}$$

Recall that the actual particle emission a_k and the actual background radiation a_B are both random variables. The decision whether a measured countrate is greater than the background a_B, i.e. whether a real source is present, is a problem of statistical testing. A decision threshold with type I error 0.05 and type II error 0.05 is given by Eq. (2) [VS07]

$$\tau = 1.65\sqrt{a_B/\Delta}\,, \tag{2}$$

where Δ is the detector interval. The background rate a_B is determined in advance by a long-term measurement.

3 Accumulation of Counts

A first simple approach to the problem of source carrier localization is the accumulation of counts. Let C_r^m be the accumulation variable of person m with respect to detector r that is gradually increased as follows: 1.) $C_r^m := 0\ \forall r$. The procedure starts when person m enters the detection area of detector r. 2.) As long as the person is inside the detection area all measured counts c_k^r are compared with the decision threshold τ. If $c_k^r - a_B > \tau$ we accept the hypothesis that a source is present and add the measured counts to the accumulation variable: $C_r^m = C_r^m + (c_k^r - a_B)$. If $c_k^r - a_B \leq \tau$ we continue. 3.) When the person leaves the detection area the accumulated counts are divided by the retention time of the person: $C_r^m = C_r^m / T_r^m$. 4.) Steps 1.) up to 3.) are processed for all passed detectors. Hence, each person m collects a personal countrate C_r^m for each detector r. When person m leaves the surveillance area the variables $C_r^m, r = 1, \ldots, R$ of all R detectors are summed up to C^m. The greater C^m the more suspicious is the person. Thus, in this first approach the detector counts are either fully included into the accumulation process (if person is inside the detection area) or completely ignored (if person is outside).

A second approach for source carrier localization can be derived by introducing count weights. With these weights w_k^{rm} the counts during the detection phase of a person are no longer fully included but partially corresponding to the relation of the person's path segment to the detector position $\mathbf{p}^r$, as shown in Eq. (3), where $\mathcal{N}()$ denotes a gaussian.

$$w_k^{rm} = \frac{\int_{(k-1)\Delta}^{k\Delta} \mathcal{N}([x_t^m, y_t^m]^\mathsf{T}; \mathbf{p}^r, \mathbf{\Sigma}^2)\ dt}{\sum_{m=1}^{M} \int_{(k-1)\Delta}^{k\Delta} \mathcal{N}([x_t^m, y_t^m]^\mathsf{T}; \mathbf{p}^r, \mathbf{\Sigma}^2)\ dt} \tag{3}$$

4 Activity as Additional State Estimate

In this section we present a novel approach to source carrier localization based on the treatment of activity as an additional state estimate. Let $\mathcal{Z}^k$ the series of all measurements up to scan k including the position measurements of the tracking sensors. We want to calculate the joint density $p(\mathbf{x}_k^m, \alpha^m|\mathcal{Z}^k) = p(\alpha^m|\mathbf{x}_k^m, \mathcal{Z}^k)p(\mathbf{x}_k^m|\mathcal{Z}^k)$ for person m. The second factor is calculated by the tracker. We assume that person m is inside detection area r. Because the conjugate prior of the Poisson distribution is the Gamma distribution, the first factor can be transformed to Eq. (6) (m is omitted in the following)

$$p(\alpha|\mathbf{x}_k, \mathcal{Z}^k) = p(\alpha|\mathbf{x}_k, c_k^r, \mathcal{Z}^{k-1}) \propto \underbrace{\mathcal{P}(c_k^r|\mathbf{x}_k, \alpha)}_{\text{Poisson}} \times \underbrace{\mathcal{G}(\alpha; \mu_{k|k-1}, \nu_{k|k-1})}_{\text{Gamma}} \tag{4}$$

$$\propto \frac{\alpha^{a_k}\exp(-\alpha)}{a_k!} \times \frac{\nu_{k|k-1}^{\mu_{k|k-1}}}{\Gamma(\mu_{k|k-1})}\alpha^{\mu_{k|k-1}-1}\exp(-\nu_{k|k-1}\alpha) \tag{5}$$

$$\propto \alpha^{\mu_{k|k-1}-1+a_k}\exp(-(\nu_{k|k-1}+1)\alpha)\,, \tag{6}$$

where a_k is the measured count at the person's position calculated by solving Eq. (9) for $\alpha =: a_k$. For a source moving parallel to the x-axis, the relation between the measured count c_k^r and the activity α is given by Eq. (9) (applicable also for arbitrary path segments).

$$c_k^r = \alpha_B\Delta + \int_{(k-1)\Delta}^{k\Delta} \frac{\alpha A_r}{4\pi d_{rt}^2}\,dt = \alpha_B\Delta + \int_{(k-1)\Delta}^{k\Delta} \frac{\alpha A_r}{4\pi(h_r^2 + (x_t - p_{\mathrm{x}}^r)^2)}\,dt \tag{7}$$

$$= \alpha_B\Delta + \int_0^{\Delta} \frac{\alpha A_r}{4\pi(h_r^2 + (x_{(k-1)\Delta} + v_{(k-1)\Delta}t - p_{\mathrm{x}}^r)^2)} \tag{8}$$

$$= \alpha_B\Delta + \frac{\alpha A_r}{4\pi v_{k\Delta}h_r}\left[\tan^{-1}\frac{x_{k\Delta} - p_{\mathrm{x}}^r}{h_r}\right] - \frac{\alpha A_r}{4\pi v_{(k-1)\Delta}h_r}\left[\tan^{-1}\frac{x_{(k-1)\Delta} - p_{\mathrm{x}}^r}{h_r}\right], \tag{9}$$

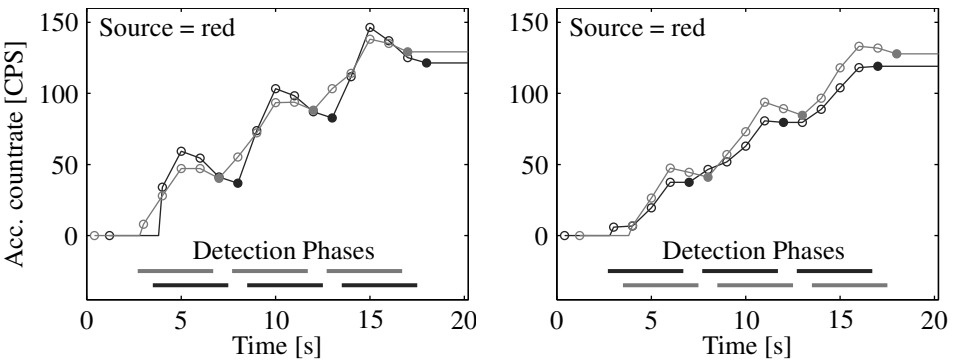
where p_{x}^r is the x-position of detector r, $v_{k\Delta}$ is the person's velocity at time $k\Delta$, h_r is the shortest distance between person and detector. Eq. (6) is a Gamma density with parameters $\mu_{k|k-1}+a_k$ and $\nu_{k|k-1}+1$. The expected value of α is $(\mu_{k|k-1}+a_k)/(\nu_{k|k-1}+1)$ leading to the update formulae $\mu_{k|k} = \mu_{k|k-1} + a_k$ and $\nu_{k|k} = \nu_{k|k-1} + 1$. The average deviation of the expected value is inversely proportional to the probability of being the source carrier. The greater the deviation the less suspicious is the person.

5 Experimental Results

The surveillance area consists of a corridor with three gamma scintillation detectors that are equally distributed at distances of 5 m along a single side of the corridor. Two persons traverse the corridor from the left to the right, walking on after the other at a constant distance. Their velocity is 1 m/s. The source activity is 250 kBq. Hence, from the decision threshold in Eq. (2) a detection radius of 2 m can be derived for the scintillation detectors

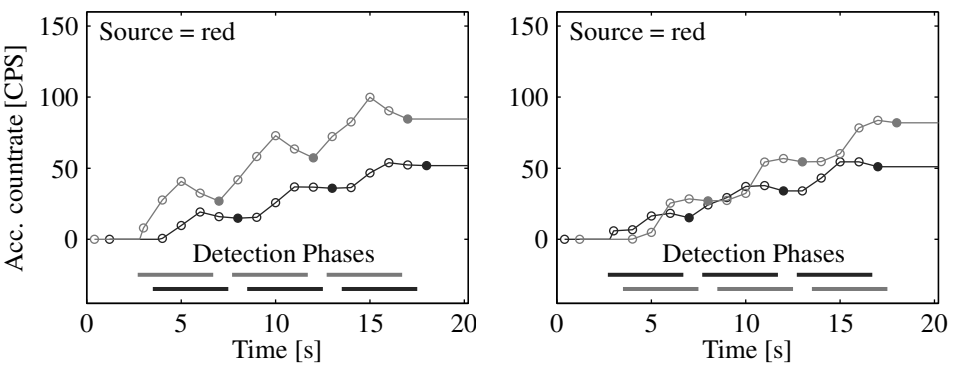
$r = 1, 2, 3$. The detectors synchronously work at a detection interval Δ of 1 s. The background radiation a_B under the photo peak is 10 counts per second [CPS].

In the following we will evaluate all three approaches for source carrier localization: the accumulation of countrates (ACR), of weighted countrates (AwCR) and the deviation of the estimated activity (αDev). Fig. 1 – Fig. 3 show the approaches for an exemplary data set. The distance between the persons is 0.8 m. $h_r = 0.8\,\mathrm{m}\ \forall r$. The simple ACR approach in Fig. 1 obviously has the worst discrimination capability. The weights in AwCR lead to an improvement. The best approach is αDev. For ACR and AwCR the carrier probabilities



(a) Source carrier in the front (b) Source carrier in the back

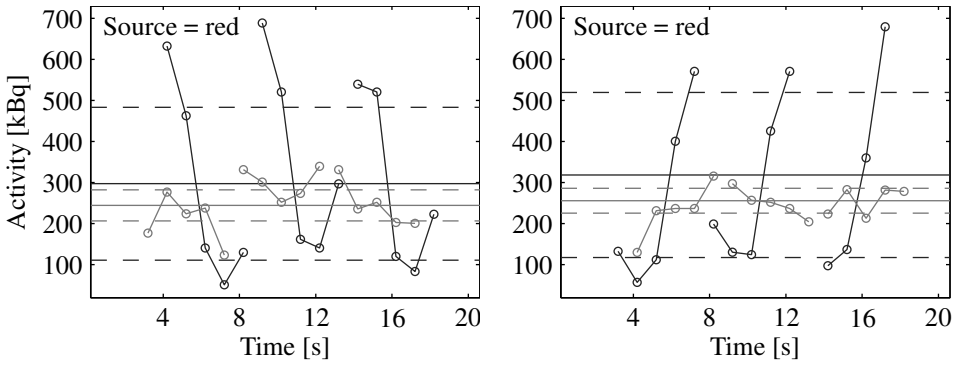
Figure 1: Accumulated counts as carrier criterion (example). Red lines correspond to the true carrier. The filled circles show the accumulated countrate after the person has left a detection area. Carrier probabilities: $P_C = 0.515$ in Fig. 1(a) and $P_C = 0.518$ in Fig. 1(b).



(a) Source carrier in the front (b) Source carrier in the back

Figure 2: Accumulated weighted counts as carrier criterion (example). For further description see Fig. 1. Carrier probabilities: $P_C = 0.620$ in Fig. 2(a) and $P_C = 0.616$ in Fig. 2(b).

are calculated as $P_C(m) = C^m/(C^1 + C^2)$. In the αDev approach the carrier probability is evaluated as $P_C(m) = 1 - \mathrm{dev}(\alpha)^m/(\mathrm{dev}(\alpha)^1 + \mathrm{dev}(\alpha)^2)$. The results in Tab. 1 confirm the evaluation based on Fig. 1 – Fig. 3. Besides the worse discrimination capability, ACR also makes wrong decisions when the persons are close to each other.



(a) Source in the front

(b) Source in the back

Figure 3: Deviations from estimated activity as carrier criterion (example). The red lines correspond to the source carrier. The solid lines show the estimated activity. The dashed lines represent the average deviation. Carrier probabilities: $P_C = 0.832$ in Fig. 3(a) and $P_C = 0.869$ in Fig. 3(b).

Distance Carrier	1.0 m front	0.8 m front	0.6 m front	0.4 m front	1.0 m back	0.8 m back	0.6 m back	0.4 m back
ACR	0.523 –	0.516 –	0.509 1% F	0.501 30% F	0.526 –	0.515 –	0.508 –	0.506 –
AwCR	0.655	0.617	0.574	0.536	0.654	0.613	0.573	0.540
αDev	0.864	0.827	0.760	0.683	0.845	0.810	0.764	0.712

Table 1: Average over the carrier probabilities P_C of the true carrier. 100 simulations per walking distance and carrier (front/back). ACR involves false decisions ($P_C < 0.5$).

6 Summary and Future Work

In this work we presented three approaches to the problem of source carrier localization in person streams. The novel αDev approach based on activity deviations could be proven to be the best in terms of discrimination. The activity estimation in αDev is realized by a recursive Bayesian filter using Gamma densities. Future evaluation will also include asynchronous detectors, position uncertainty, real data and various source strengths. The methodical future activities will consider persons as extended objects and shielding effects.

References

[VS07] Hans-Gerrit Vogt and Heinrich Schultz. *Grundzüge des Praktischen Strahlenschutzes (German)*. Hanser Fachbuchverlag, 3 edition, 2007.

[WK09] Monika Wieneke and Wolfgang Koch. Combined Person Tracking and Classification in a Network of Chemical Sensors. *Elsevier Journal of Critical Infrastructure Protection*, 2(1–2), 2009.

Information Fusion:
Protection of Military Forces and Homeland Security

Kaeye Dästner, Bastian Köhler, Felix Opitz

Air Dominance and Data Fusion (ISEG 42)
EADS Deutschland GmbH
Wörthstraße 85
89077 Ulm, Germany
kaeye.daestner@eads.com
bastian.koehler@eads.com
felix.opitz@eads.com

Abstract: Information fusion is a key factor to enable Homeland security and military protection systems. Applications have to detect, track, classify and identify objects in real time. Sensor resources must be allocated to these objects of interest and the whole situation has to be analyzed concerning threats. Whenever necessary counter measures have to be initialized.

1 Introduction

1.1 Homeland Security

Homeland security refers to civil application domains, like the surveillance of borders, infrastructures or critical facilities or events.

E.g. border security handles the surveillance of several thousand kilometres of borderline within different terrains and climate conditions. The task is to prohibit illegal immigration, illegal fishing, smuggling, piracy or terroristic infiltration.

1.2 Protection

Protection is related to the military domain, and addresses the protection of own forces in peacekeeping or peace enforcement out of area missions. Typical applications address the protection of military camps, convoys, naval platforms in critical situations, or military airports and sites.

Also interoperability issues for combined and joined forces have to be ensured. Therefore, external C2-components as well as surveillance systems, like site surveillance, air defence systems or civil air traffic management have to be integrated. Finally the different data links (Link11, Link 16, Link22) are used to distribute the recognized picture and additional real-time or non real-time information.

1.3 Information Fusion Topics

The main issue of information fusion is to keep sea or land areas, and the lower air space under surveillance. Therefore, objects within these areas have to be detected, tracked, classified and whenever possible identified.

To get the necessary information items and to ensure a good coverage of the surveillance area the information fusion has to deal with a distributed sensor suite. The sensors are mounted on top of surveillance towers, land vehicles and naval platforms and are linked together via fibre optical and radio networks.

For spectral diversity different types of sensors are commonly used. Ground surveillance radars are used because of their good detection and tracking capabilities for large surveillance areas und different climate conditions. Electro optical systems contribute to classification and identification. These systems consist of long range day light cameras, infra red systems and laser rangers which are mounted on top of a pan tilt head. This pan tilt head allows directing the electro optical systems onto the different objects.

Video processing software involves the detection of moving objects within image sequences (change detection), the steering of a camera to follow a designated object (template matching), or to recognize targets by comparison with predefined patterns (automatic target detection), e.g. by neuronal networks. Hence, also the video processing software addresses information fusion items mentioned above.

2 Radar Processing

2.1 Radar: Sensors

For ground surveillance tasks pulse and FMCW radars are used. Pulse radars are normally realized as multi mode radars to support the ground surveillance task in an optimal manner. The modes can be switched by the user and therefore allow an active interaction. Often they possess modes for

- Acquisition,
- Single target tracking,
- (Single or multi) sector surveillance (TWS),
- Panorama surveillance (TWS) and
- Jammer detection

The acquisition mode is used to direct the radar beam manually into a small sector.

Within the single target tracking mode the radar catches a designated target and controls the antenna in such a way that the target is kept. STT results in a very high update rate for the track which also helps to improve course and speed estimations.

The most important modes are the track while scan modes of ground surveillance radars. These allow the scanning of one or multiple predefined sectors. In these scanning modes the update rate is lower as within single target tracking. This is a real challenge for the filter and association algorithms because of the spectrum of ground target types (e.g. pedestrians, vehicles) and the corresponding bandwidth in dynamic variations. Mechanically scanning radars allow also to scan several sectors in a sequential manner. Electronically scanning antennas offers higher performances because, they allow quasi simultaneous scanning of several sectors and offer a constant update rate for the objects. Further they allow a parallel tracking of dedicated objects with a higher update rate or the intensive observation of special point of interests.

Finally, a 360° (panorama) surveillance is possible with ground surveillance radars.

2.2 Radar: Detection – Tracking - Classification

The detection of targets is most times based on Doppler processing and so called CFAR algorithms. Special attention must be given to the plot extraction with respect to resolution phenomena. Tracking uses non linear filter algorithms to take the measured Doppler value into account. Further the filters must be adapted to the different radar modes. For plot track association a gating algorithm is used, which depends on the false alarm density. Complex multi dimensional data association techniques [BP99] are used to overcome the uncertainty in state prediction and the possible long update intervals.

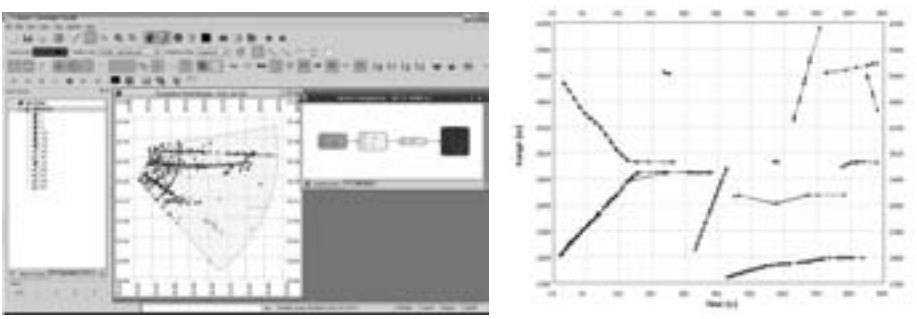


Figure 1: Sector Scan: Fusion++ Radar Tracks vs. Time

Because of the object spectrum track initialisation is done very carefully. The object spectrum contains pedestrians as well as vehicles, which depends on very different dynamic models. Therefore a score based initialisation mechanism was selected. This offers fast initialisation for fast vehicles and a very stable initialisation for slow moving objects simultaneously [BP99].

The tracking performance can be improved by taking into account digital terrain elevation data (DTED) or road map information (VMAP).

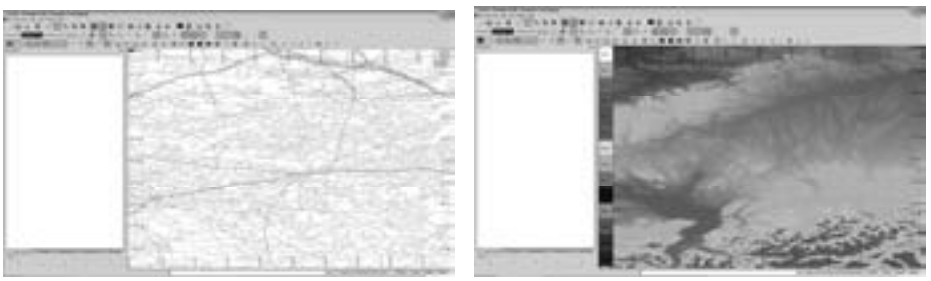


Figure 2: Fusion++ with VMAP and DTED

Performance tests were executed within the European and Arabian environment. Especially in the challenging environment of heavy sandstorms a good performance was important.

3 Video Processing

3.1Video: Sensors

The sensors consists of IR and long range day light TV cameras which are mounted on a pan tilt head. Often a laser ranger is also available. The sensor delivers a video for each camera. Further a data stream is provided, which contains the commands to steer the pan tilt head and the cameras. Status messages offer the actual direction and camera settings.

3.2 Video: Detection – Tracking - Classification

Four main software components are used to create a georeferenced view of the camera image.:

- Alignment
- Video processing, change detection and correlation
- Georeferencing with DTED
- Sensor tracking
- Visualisation with a display area.

A real challenge for georeferenced tracking with camera systems is the alignment of the camera system. The Alignment of the camera has to be very accurate since small uncertainties of the elevation can cause great uncertainties in the range estimation [DKO09].

Different aspects have to be taken into account when a camera system has to be aligned. On one hand the pan-tilt-head has to be adjusted in north orientation and horizontal direction. Otherwise the calculated target direction does not direct to the correct DTED position.

Further the optical axis within the camera system can be misaligned with respect to the pan/tilt system or misalignment effects occur through zooming operations with respect to the calculated field of view (bore sight alignment).

On the other hand two cameras (IR and TV) mounted on one pan-tilt-head may be misaligned to each other. Than the tracked targets are misaligned in range and bearing and fusion of these tracks becomes difficult.

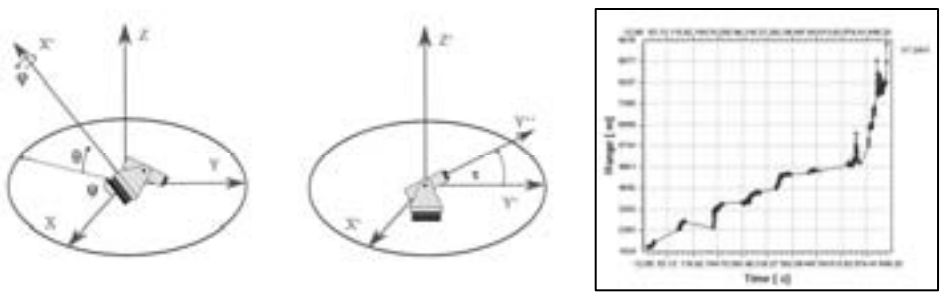


Figure 3: Alignment and coordinate systems: Earth vs. pan-tilt system (left), pan-tilt vs. camera (optical axis, centre); Video tracking: Stop and Go scenario (right)

After the alignment a video based detection and tracking software is used to identify moving objects in a scene, e.g. a landscape with moving vehicles and pedestrians. The software detects these objects automatically and delivers their pixel coordinates in the video picture.

If an object is detected within one video frame it is described in image coordinates. With respect to the current platform parameter, given by azimuth and elevation the direction of the target can be calculated. [DKK09]

Digital Terrain Elevation Data allows the range estimation of an object that is moving on the ground. By calculating the intersection between the direction and a DTED layer the range can be found [Ba09].

Once the extracted image is georeferenced the object can be tracked with a special Filter to handle different motion models of targets.

Additionally an automatic target recognition (ATR) is available, which distinguishes between different object classes.

4 Conclusion

Potential applications of information fusion for homeland security and protection are shown and various techniques of information fusion which are involved in the sensor integration and processing are described. These handle the detection, tracking and classification process for the radar and video based processing. These are key factors for effective and user friendly surveillance systems.

References

[AR05] Acharya, T.; Ray, A. K.: Image Processing Principles and Applications, Wiley, 2005.

[Ba09] Bartelme, N.: Geoinformatik: Modelle, Strukturen, Funktionen. Springer, Berlin, 2009.

[BP99] Blackman, S.; Popoli, R.: Design and Analysis of Modern Tracking Systems. Artech, Norwood, MA, 1999.

[Br01] Bremaud, P.: Markov Chains - Gibbs Fields, Monte Carlo Simulation, and Queues, Texts in Applied Mathematics , Vol. 31, Springer, 2001.

[DKK09] Dästner, K.; Köhler, B.; Kouemou, G.; Opitz, F.: Sensor Integration in the Security Domain. Informatik 2009, Lübeck.

[DKO07] Dästner, K.; Kausch, T.; Opitz, F.: An Object Oriented Approach for Data Fusion. GI Jahrestagung (2) 2007.

[DKO09] Dästner, K.; Köhler, B.; Opitz, F.: Fusion of IR / CCD video streams and digital terrain models for multi target tracking. Informatik 2009, Lübeck.

[KO07] Kuouemou, G.; Opitz, F.: Hidden Markov Models in Radar Target Classification, Radar 2007.

[OFD07] Opitz, F.; Filusch, J.; Dästner, K.; Kausch, T.: Information Fusion in Anti Asymmetric Warfare and Low Intensity Conflicts. Fusion 2007, Québec, Canada, 9-12 July 2007.

[Op08] Opitz, F.: Sensor Suite and Data Fusion in Security Applications. SET-125 Symposium on “Sensors and Technology for Defence Against Terrorism (DAT)” Bundesakademie für Wehrverwaltung und Wehrtechnik, Mannheim, Germany, 22-25 April 2008.

[Wa04] Watkinson, J.: The MPEG Handbook. MPEG-1, MPEG-2, MPEG-4 (MPEG-4 Part 10/H.264/AVC included), Butterworth Heinemann, 2004.

[WOZ01] Wang, Y.; Ostermann, J.; Zhang, Y.-Q.: Video Processing and Communications, Prentice-Hall Signal Processing Series, Prentice Hall, 2001.

Knowledge based Information Evaluation for Integrated Ship Bridge Systems

Reinhard Müller [1], Michaela Demuth[2]

[1] Hochschule Wismar, Bereich Seefahrt, [2] Schiffahrtsinstitut Warnemuende
Richard-Wagner-Str. 31
D – 18119 Rostock
reinhard.mueller@hs-wismar.de
michaela.demuth@hs-wismar.de

Abstract: The following paper describes a data fusion oriented approach for a novel information handling on integrated ship's bridge systems using agents. A Joint Data Laboratories based data fusion model was used for the ship navigational data handling. First experiments supported by data from a Voyage Data Recorder, extern knowledge data bases and various defined information layers will be discussed. The goal of the investigation was the information relevance evaluation in the navigational process current circumstances. An applied example will be discussed.

1 Introduction

The most ship's bridge equipment devices follow their own philosophy of displaying information and announcing warnings or alarms when determining safety- or operation-critical states. The technical navigation systems are used on board vessels in a different manner. Very often these systems have an integrative mode of operation. However, they have actually a life of its own depending of the equipment manufacturers. The sense and handling of the announced warnings or alarms in complex situations is not acceptable in ship handling procedures, because of the device-centric decision-making process. The nautical staff have not the ability to distinguish the acoustically and visually cognition of alerts due to the lack of classified or graduated critical information appearances. The currently used alert systems do not work in accordance to the state of vessel's vitality neither with the technical periphery nor with the navigational situation or task.

A cross-system data management for process information harmonizing and weighting is necessary. Such a management system would be aimed to bundle sensor-information, warnings and alarms efficiently. Determined process-dependent information sets and hierarchies have to be carried out for handling this challenge. The goal of a new data and alert management should be the discharge of the nautical staff during the operational task by confining the permanent information overflow. The integrative load of the operator during sorting and prioritizing of process information and process handling has to be minimized by a vital display of significant and qualified process data. This dynamic needs a holistic approach which can be reached by the theory and methods of

Data Fusion (DF). The sensor data of vessel's technical system like engine and bridge systems may be correlated with its current circumstances as the current maneuvering behavior of own ship, navigational conditions (spaces, targets and weather) and rules of collision avoidance and grounding. It is essential to associate the measured local states of solely sensors with the global overall situation and to post-process the data in respect to this focus.

2 Data Fusion in Ship's Bridge Information Systems

The probably most complex system using methods of DF is the human being by itself. It is able to recognize and to evaluate different situations in shortest time frames because of its sense organs and its capability of cross-linking. At all times a decision tree will be running through in the background. The created result may be different every time due to the external influences or due to the growing number of the decision tree branches. The technical DF aims to reverse engineering of the human model. At this the measuring sensors will be connected to a net-working system which is able to recognize an event by additional aspects [BI98]. The pyramid tapering of a voluminous dataflow into process-depending knowledge is the central task of a data cross-linking on ship's bridge.

2.1 The Architecture of the used Model

The theory of DF is summarized in the generic Approach of a DF System [Ha01], [Bi03]. The data of different resources will be pre-processed, linked and refined and will get a new and higher quality after each level of processing by using the generic architecture of a multi sensor fusion system like in [SGD99]. The performance of a DF System is also based in the prognostic of relevant data. The system may detect unsteadiness of the behavior rapidly. That means for the ship's system, that the functionalities and the interrelationship of sensors can be treated by the methods of DF in a very efficient manner [Da07]. The goal is a cross-linked description of an operating ship of her movement, her load, her mission related to the interaction with the surroundings like the navigational space, the targets, the wind and water. The more exact and complete the accordant situation may be qualified the more accurate the data of technical system may be interpreted according to the given context.

2.2 The Knowledge based Information

The ship's sensor system affords the technical measurement recognition of the environment in more or less complete manner. So, any situation may be interpreted as a structured extract from the real world and may be formed as information, which can be combined, abstracted and related to each other information. Putting expert knowledge into the investigation the result of a multi-sensor-system is a correlated and integrated outcome. The information η itself may be defined as $\eta = \{r, x, h, p, k\}$, with r as a relation between units, x as location, t as time, p as probability, and k as a class membership [St06].

The data of the Ship-Bridge Information Management has to be collected and to be sort according to a hierarchical structure. Therefore a multi-layer model [Ha06] based on a performed graph architecture is used. The first layer called Data Layer provides the origin sensor data to the next higher layer called Information Layer which handles the data by given processing functions. The resulted outcomes build up the Knowledge Layer. Each layer has included a restricted knowledge base containing the kind of data and operational tools. The integration and controlling of ship's data is carried out in all three layers as a dynamical graph. Its building up depends on the operational result in every existing knot. That means, only the relevant connections of the knots inside the tree will be treated. The knots are implemented as special structures and are called as agents. They contain operational intelligence with respect to their controlling domain.

3 Implementation

The automatically estimation of the kinematical state of the own ship is a central intermediate process of the implemented DF methods for getting the reference to the context. Due to the use of a context information carried out by DF may be weighted in relation to their relevance. That means, not all displayed information is necessary in every situation of the navigational process. The opportunity is given for splitting off the information in temporary important or unimportant, necessary or needless and other groups by taking a context into account. Therefore f is the elemental function for mapping the parameter **I** used for describing the current kinematical state of the own ship into a real number **R**. By means of a measurable set **I** of the parameter the integral of f is estimated by:

$$s = \int f d\mu = \sum_{i=1}^{n} \alpha_i \mu(I_i), \quad \sum_{i=1}^{n} \alpha_i = 1$$

with $\alpha_i \in$ R and $\mu(I_i)$ is the measure. The identification of the weights α_i was done in an empirical way. It based on interviewing of experts and analytic evaluation of data records of a real ship during voyage.
Beside the dynamical parameter of ship's moving behavior like speed, rate of turn, drift and its derivatives the actuating elements like revolutions, rudder angles, thrusters and engine mode also belong to the parameter set for estimating the kinematical ship state s_{cin}. with

$$f : R \mapsto N_0, f(s) = s_{cin}, s_{cin} \in S = \{0,1,2,3\} \quad . \quad (2)$$

The different characteristics of the well-defined parameters are clustered into four introduced classes of the ship-state set **S**: 3 – *Underway*, 2 – *Maneuvering*, 1 – *Drifting* and 0 – *Moored*. The cluster *Underway* consists of all the ship-states, which means the classical voyage of a vessel. An anchored vessel (harbor, anchorage), which are moving with very low or none velocities belongs to the cluster of *Moored*. If the moving behavior of the ship is solely resulted by external influences like wind and current the ship-state class is *Drifting*. All others ship-states like coastal or pilotage trade, maneuvering, positioning and tugging are clustered by *Maneuvering*.

4 The experimental Use

During the sea trail of a motor vessel before ship's delivery the NMEA data were recorded by the Voyage Data Recorder (VDR) for about 36 hours. In the following figure the recorded sea trail NMEA – data are shown. The sea trail consists of the full required maneuvering program before ship's delivery. The record includes engine, rudder and thrusters maneuvers as well anchor, tug and blackout maneuvers. An extract of the focused parameters are shown in the diagram as normalized values.

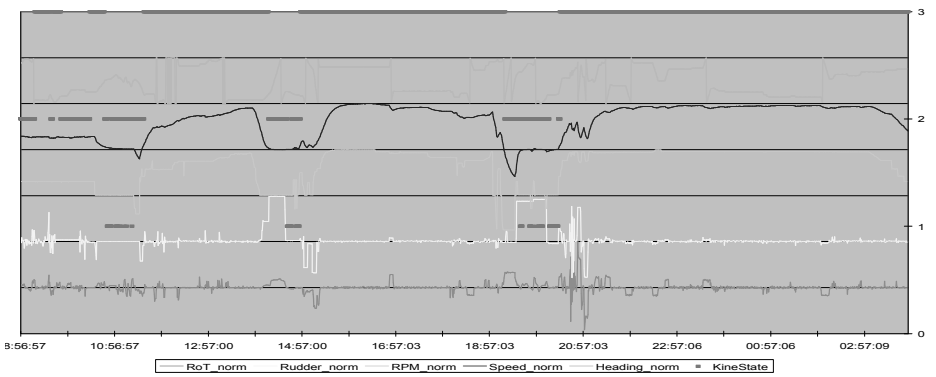


Figure 1: Assignment of the cinematic ship-state based on dynamical parameter of moving behavior (3 – class *Underway*, 2 – class *Maneuvering*, 1 – class *Drifting*, 0 – class *Moored*)

After the use of the implemented DF methods on the recorded data a mapping of the ship-states is carried out by the agents and marked by the red dots in the diagram for each time step. For example, in the time frame of about 40 minutes at 11 o'clock of the time scale the estimated kinematical state of the own ship alternates between the classes *Drifting* and *Maneuvering*. The mapping alternation here depends on the measured Rate of Turn, which is influenced by the switching use of the ship's thrusters during the testing and adjusting period. The estimated kinematical ship-state influences the operational capability of the implemented agents and the priority of the processed data. The information will be evaluated and controlled in accordance to the current context in terms of their relevance, their availability and their quality. Thus, the temporary consideration of the parameter Heading in the cluster *Underway* has to be evaluated differently as in the cluster *Maneuvering*.

5 Results

The following figure shows the results of the applied DF approach for Ship-Bridge Information Management. The input data are real NMEA data. A 30 minutes time sequence of one navigational parameter POSITION is abstracted displayed. The diagram range is a color gradient from green to red, which reflects from a very good to a very low availability and quality of the data according the context shown by the color of the value marker. Each color describes one kinematical state of the own ship estimated by the algorithm. The origin data of position recorded by the Voyage Data Recorder are marked by the white triangle. For this data the kinematical state estimation and an agent controlling do not exist. Three times the position signal is lost each for 4.5 minutes. No position value is available, the quality is zero. A concert of acoustic alerts produced by each technical devise handling with position data like Track pilot, Conning Display, ECDIS, Weather Charts and Radar would be the consequence at the ship's bridge.

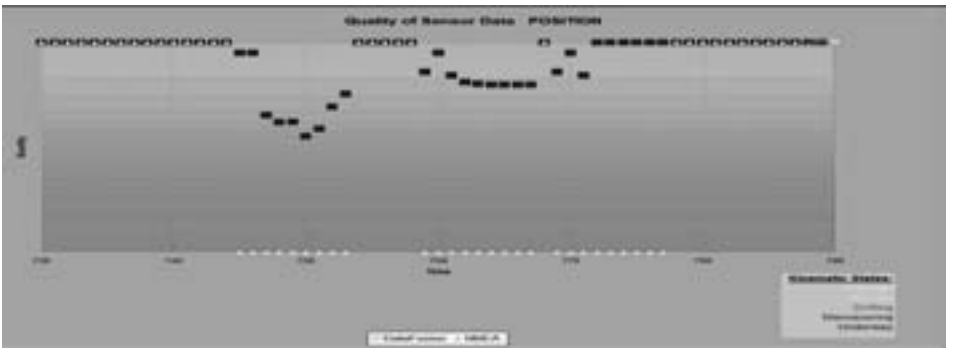


Figure 2: Comparing the quality and availability of origin and fused data, (One time step = 30 sec on the time scale)

Looking at the second group of displayed values marked by colored quads, it is evident, that the missing origin signal is substituted by alternative data representing positioning data. This data may be delivered from a competitive sensor or estimated on the basis of a co-operative sensor. The quality of the substituted value is evaluated by the algorithm with respect to the context. The kinematical state of the own ship is estimated as a maneuvering behavior (brown marked quads) during the first and second period of signal missing. In this state the positioning data are appraised as necessary and the threshold of a satisfactory quality is higher than in the state *Underway*. So, the quality of the substituted data reaches lower in opposite to the third period of missing signal. There the quality of the substituted values is regarded as satisfactory good. Here, the consequences at the ship's bridge could be an alert handling reduction by a centralized technical device. The exemplary diagram shows the different processing operations of the implemented algorithm in an effectual way. The lost information handling may be analyzed and controlled in a context depending method by the use of a DF approach for a ship's bridge information system. This approach gives the possibility for a different and automatically adapted conditioning of bridge alerts.

6 Summary

Currently the data flow of an INS on ship's bridges is characterized by a partly exchange of navigational data of special integrated bridge sub-systems. The technical components are directly connected with the necessary sensors. There is no cross-link or cross reference between the devices. A data pre-processing or –checking is made only by the respective component of the technical system. In case of an operational fault of any sensor each technical sub-system depending on this sensor data will announce its own acoustical alert. The consequence is an alert distress of the nautical staff and acceptance failures to the technical systems. Taking the knowledge of the sensor architecture into account the alert frequency may be reduced by substitution of a sensor fault by competing or corresponding sensors. A second step in reaching a silenced bridge is a context depending evaluation of a sensor fault. Alerts without a relevance to the defined context may be quieted. During a four years lasting research project ("DGON-bridge" funded by the German Federal Ministry of Economics and Technology) it was shown that the goal of a differenced ship alert handling for reducing the overload of the nautical bridge personnel is reachable by the theory of DF using Agent Model. It was proofed that the alert disturbances can be measurable reduced since the sensor data of the technical system ship are cross-linked and focused under identified circumstances. The implemented method is applicable in real-time and well-proofed by the good outcomes.

Literature

[Im09] IMO NAV CG on IBS: Draft Performance Standards for Bridge Alert Management (in progress)

[BI98] R.R. Brooks, S.S. Iyengar: Multi-Sensor Fusion. Prentice Hall PTR Jersey, 1998.

[Ha01] Hall, D. L.: Handbook of multi sensor data fusion. CRC Press, 2001.

[Da07] Darms, M.: Eine Basis-Systemarchitektur zur Sensordatenfusion von Umfeldsensoren für Fahrerassistenzsysteme. VDI Verlag, Reihe 12, Nr. 653, 2007.

[Bi03] J. Biermann: A Knowledge – based Approach to Information Fusion for the Support of Military Intelligence. NATO RTO Symposium on Military Data and Information Fusion IST – 040 / RSY -012, Prague, Czech Republic, 20.-22.10.2003.

[SGD99] Sick, B.; Grass, W.; Donner, K.: Multisensorfusion bei der Verarbeitung mikrosystem¬basierter Signale. Informationstechnik und Technische Informatik, 41, 1999.

[CY90] J.J. Clark, A.L. Yuille: Data Fusion for Sensory Information Processing Systems. Kluwer Academic Publishers, Boston, 1990.

[Ha06] Hall, D.: Increasing Operator Capability through Advances in Visualization Systems. SensorFusion Conference, Washington D.C., 2006.

[Ko00] Koch, W.: Multisensordatenfusion als Kernfunktion in Überwachungssystemen. Information Technology Nr. 47, Oldenburg, S. 79-86, 2005.Abel, K.; Bibel, U: Formatierungsrichtlinien für Tagungsbände. Format-Verlag, Bonn, 2000.

[St06] Steinberg, A.: Threat Assessment Issues and Methods. SensorFusion Conference, Washington D.C., 2006.

Probabilistic Classification of Disease Symptoms caused by *Salmonella* on *Arabidopsis* Plants

Marek Schikora[1,3], Adam Schikora[2], Karl-Heinz Kogel[2],
Wolfgang Koch[1] and Daniel Cremers[3]

[1] Fraunhofer FKIE, Dept. Sensor Data and Information Fusion, Wachtberg, Germany
{marek.schikora, wolfgang.koch}@fkie.fraunhofer.de

[2] IPAZ Phytopathology, Justus Liebig University Giessen, Germany
{adam.schikora, karl-heinz.kogel}@agrar.uni-giessen.de

[3] Department of Computer Science, Technical University of Munich, Germany
daniel.cremers@in.tum.de

Abstract: Several reports have linked food poisoning with the consumption of raw vegetables and fruits contaminated by *Salmonella*. Most studies suggested an extracellular lifestyle of *Salmonella* on plants. However, more recent studies show that *Salmonella* are also able to colonize the intracellular compartment of various plant tissues causing chlorosis and eventually death of infected organs. The aim of this work is to present a probabilistic classification algorithm for disease symptoms on *Arabidopsis thaliana* plant in order to improve the current biological research. The algorithm itself uses images of *Arabidopsis thaliana* leaves as input and consists of two steps. The first step is the detection of pixels belonging to a leaf. This is done with a globally optimal color segmentation method. The second step is realized with a probabilistic framework to classify each pixel. Finally a morbidity rate is computed based on the classification result.

1 Introduction

In recent time, several reports have linked food poisoning to the consumption of *Salmonella*-contaminated raw vegetables and fruits. Most studies suggested an extracellular lifestyle of *Salmonella* on plants. However, recent results have shown that *Salmonella* bacteria are also able to colonize the intracellular compartment of various plant tissues, causing chlorosis and eventually death of infected organs [SCCH08]. Moreover, similar to other plant pathogens this bacterium triggers complex host defense responses in *Arabidopsis thaliana*. Among other reactions to pathogenic bacteria, plants induce also so-called hypersensitive response (HR). Core of this reaction is the programmed cell death (PCD). PCD is a very tightly controlled process, in which infected areas or organs are sacrificed in order to stop the invaders. On leaves, PCD can be easily visualized since it causes yellowing of plant tissues. This process depends on numerous factors and can



Figure 1: Input images for the proposed classification algorithm. Left image: healthy plant. Right image: sick plant

be prevented by successful pathogens. To investigate how *Arabidopsis thaliana* defends itself and how bacteria interfere with plant immunity, we want to analyze the impact of different bacterial mutants on plant tissues. To solve this task, plants are infected with *Salmonella* and images of infected leaves are taken at different time points after infection. Typical input images can be seen in Figure 1. The task here is to establish an objective measurement for the disease rate in these leafs. This is done in two steps. First, for each pixel in an image the decision has to be drawn if it belongs to the leaf or not. This is done using a convex energy functional whose minimum is the desired segmentation. This topic is presented in the second section. Second, each pixel belonging to a leaf has to be assigned to a class (healthy vs. sick). This classification procedure is described in Section 3.

The workflow of the proposed algorithm is visualized in Figure 2.



Figure 2: Workflow: First the input image is segmented into foreground (black) and background (white). Then for each foreground pixel a classification is performed. Unhealthy classified pixels are marked cyan.

2 Color Segmentation

The problem of extracting relevant objects from images can be seen as the segmentation of an image into two regions, foreground and background. All pixels labeled as foreground count as part of an object and are interesting candidates for further analysis. Image segmentation is a common task in computer vision, and many solutions have been proposed for this problem. Currently, the best solutions are provided by variational approaches. Three main classes of variational approaches exist for image segmentation, the first one being level sets [OS88, CV01]. The main advantage is that the energy functional being minimized is formulated continuously, so there is no need for discretization. On the other hand, the local optimization of the energy functional does not necessarily lead to a globally optimal solution. The second class are graph cuts [GPS89, BVZ01] with two main advantages: the computation time is generally very short and the solution is approximately globally optimal. The main disadvantage of this approach is the discrete formulation on a graph, leading to discretization errors. A combination of the benefits of those two methods constitutes the third class: total variation (TV) minimization. Chan et al. [CEN04] proposed this method for image segmentation of intensity-based images using a transformed Mumford-Shah model. Additionally in [SHRW09] and [UPCB08], it was shown how this approach can be extended to color images. In this paper, we will rely on [SHRW09] but use a different color space for segmentation.

2.1 TV-Segmentation

The segmentation of an image $I : \Omega \to [0,1]^3 \subset \mathbb{R}^3$ with $\Omega \subseteq \mathbb{R}^2$ can be seen as separation of the image plane Ω into disjoint regions $\Omega_1, \Omega_2, ..., \Omega_n$, with $\Omega = \Omega_1 \cup, ..., \cup \Omega_n \cup \Gamma$, where Γ denotes the contour of the segmentation. In the case discussed here, there will be only two regions Ω_{obj} and Ω_{bgd}, so we are looking for a binary image $u : \Omega \to \{0,1\}$.

In [SHRW09] the authors present a convex energy functional based on total variation. In their work they use the HSV color space. In order to be independent of illumination changes this color space is the correct choice. However, since the definition of the hue channel is done in polar coordinates, euclidean distances are not applicable on every channel identically, which is preferable. To obtain an independency of illumination changes and the ability to use the euclidean distance we will use the I1I2I3 color space, proposed by Hafner [Haf99]. The transformation of a RGB pixel value to an I1I2I3 pixel value can be denoted with:

$$I_{123}(\mathbf{x}) = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{2} & 0 & -\frac{1}{2} \\ -\frac{1}{4} & \frac{2}{4} & -\frac{1}{4} \end{pmatrix} \cdot I(\mathbf{x}). \tag{1}$$

The first channel contains the illumination information. The second and third channel contain the color information. With this linear transform, we can derive a convex energy

functional for color image segmentation in the proposed color space:

$$E(u, \boldsymbol{\mu}_{\text{obj}}, \boldsymbol{\mu}_{\text{bgd}}) = \int_{\Omega} \left(f(I_{123}(\mathbf{x}), \boldsymbol{\mu}_{\text{obj}}) - f(I_{123}(\mathbf{x}), \boldsymbol{\mu}_{\text{bgd}}) \right) u(\mathbf{x}) \mathrm{d}\mathbf{x} + \lambda \int_{\Omega} |\nabla u(\mathbf{x})| \mathrm{d}\mathbf{x}, \tag{2}$$

with

$$f(I_{123}(\mathbf{x}), \boldsymbol{\mu}) = w_1([I_{123}(\mathbf{x})]_{\text{I1}} - \boldsymbol{\mu}_{\text{I1}})^2 + w_2([I_{123}(\mathbf{x})]_{\text{I2}} - \boldsymbol{\mu}_{\text{I2}})^2 + w_3([I_{123}(\mathbf{x})]_{\text{I3}} - \boldsymbol{\mu}_{\text{I3}})^2 \tag{3}$$

denoting a weighted squared sum of the individual channels. For the results presented in this paper we use $w_{\text{I1}} = 0.1$ and $w_{\text{I2}} = w_{\text{I3}} = 0.45$. As additional input we use mean values for the foreground $\boldsymbol{\mu}_{\text{obj}}$ and background $\boldsymbol{\mu}_{\text{bgd}}$ and a smoothing parameter λ. The proof of convexity is straight forward and will not be presented at this point. Using the Euler-Lagrange equation and a local optimization scheme (e.g successive over-relaxation) we can find the global minimum of (2), which is the desired segmentation.

3 Probabilistic Classification

The next step in the proposed algorithm is a classification of all pixels that were labeled as part of a leaf by the procedure from the previous section. Each classification algorithm has an offline and online phase. In the offline phase the classification model is learned. The actual classification is then performed in the online phase, where the measurements are checked against the learned model [Bis07].

Model Learning
In order to learn a non-over fitted model, we take several images from healthy leafs. Then, we perform a segmentation and save all leaf pixel values (several millions). To be independent from illumination changes we only use the second and third channel of the I1I2I3 color space, leading to two dimensional data points. We cluster the data points into M clusters (e.g. $M = 3$) using the k-means algorithm. Finally, for each cluster we compute its mean value $\boldsymbol{\mu}_i$ and covariance matrix Σ_i, with $i = 1, ..., M$. By using this multimodal color distribution we can provide a probabilistic model $\mathcal{M}$ for a healthy leaf. Since this step is quite time consuming, this type of model learning can be done offline before the actual classification task.

Model Checking
Given a probabilistic model $\mathcal{M}$ representing a healthy plant we can now efficiently check for each labeled pixel $\mathbf{x}$ if it belongs to this model. For this purpose we compute the following likelihood for every labeled pixel.

$$p(\mathbf{x}|\mathcal{M}) = \max_{i=1,...,M} \exp\left(-0.5 \cdot ([I_{123}(\mathbf{x})]_{I2,I3} - \boldsymbol{\mu}_i)^T \Sigma_i^{-1} ([I_{123}(\mathbf{x})]_{I2,I3} - \boldsymbol{\mu}_i)\right) \tag{4}$$

Herein, $[I_{123}(\mathbf{x})]_{I2,I3} \in \mathbb{R}^2$ denotes a vector which consists of the I2 and I3 channel information. Given a labeled pixel $\mathbf{x}$, we test the following condition:

$$1.0 - p(\mathbf{x}|\mathcal{M}) \geq \tau. \tag{5}$$

If (5) is true, $\mathbf{x}$ is classified as unhealthy, otherwise as healthy. A typical value for τ is 99.995%.

4 Results

In this section, we present some experimental results achieved with the proposed algorithm. In Figure 3, one can see a screenshot of the graphical user interface (GUI) developed for this task. The big benefit of this GUI is its simplicity and clarity. Users without knowledge about the underlaying algorithms can use them efficiently to classify. Additionally, one can see some classification and segmentation results for a given input image. Some further results are displayed in Figure 4. As it can be easily recognized, the automatic classification results match the visual perception of a human observer. These examples make clear that the proposed algorithm shows reliable results. Unfortunately, we did not have a ground truth for this data to intensively analyze the algorithm, but we can say with fair certainty that this work is a good basis for further development.



Figure 3: Screenshot of the algorithm GUI. In the center the marked pixels are displayed. On the right the input image and the segmentation result can be seen. On the left and at the bottom control parameters can be set.

5 Conclusion

In this work, we present a probabilistic algorithm for classification of disease symptoms in *Arabidopsis thaliana*, caused by *Salmonella*. First, a detection of leafs in the input image is performed. This is achieved by a globally optimal color segmentation strategy based on total variation. Second, all leaf pixels are classified using a learned multimodal color distribution model and a likelihood function. In practical experiments, we could show a good performance. The presented algorithm can simplify the quantitative evaluation of plant defense reaction to bacterial infection, because he provides an objective rate of disease symptoms.

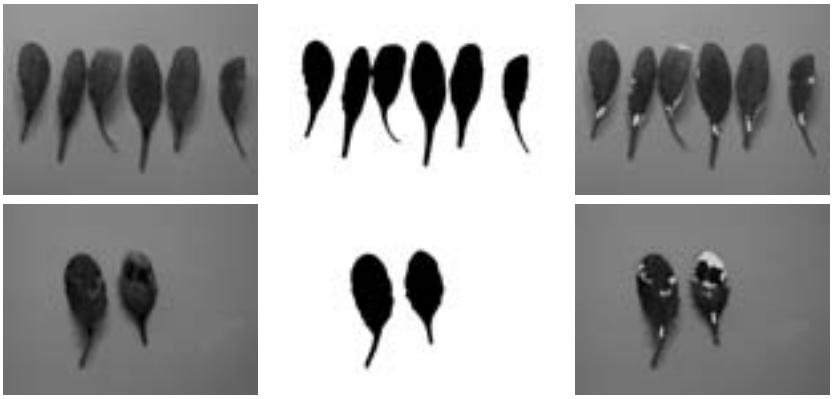


Figure 4: Classification results of the proposed algorithm. Left images: input; middle images: segmentation result; right images: classification visualization, with cyan marking.

References

[Bis07] C.M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2 edition, 2007.

[BVZ01] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(11):1222–1239, 2001.

[CEN04] T.F. Chan, S. Esedouglu, and M. Nikolova. Algorithms for finding global minimizers of image segmentation and denoising models. Technical report, UCLA, September 2004.

[CV01] T.F. Chan and L.A. Vese. Active contours without edges. *IEEE Transactions on Image Processing*, 10(2):266–277, 2001.

[GPS89] D.M. Greig, B.T. Porteous, and A.H. Seheult. Exact maximum a posteriori estimation for binary images. *Journal of Royal Statistical Society, Series B*, 51(2):271–279, 1989.

[Haf99] W. Hafner. *Segmentierung von Video-Bildfolgen durch Adaptive Farbklassifikation*. PhD thesis, Computer Science Department, TU Munich, 1999.

[OS88] S. Osher and J.A. Sethian. Fronts propagation with curvature-dependent speed: Algorithms based on hamilton-jacobi formulation. *Journal of Computational Physics*, 79:12–49, 1988.

[SHRW09] M. Schikora, M. Häge, E. Ruthotto, and K. Wild. A convex formulation for color image segmentation in the context of passive emitter localization. In *12th International Conference of Information Fusion*, pages 1424–1431, July 2009.

[UPCB08] M. Unger, T. Pock, D. Cremers, and H. Bischof. TVSeg - Interactive Total Variation Based Image Segmentation. In *British Machine Vision Conference (BMVC)*, September 2008.

A tripartite architecture based on referee function for generic implementations of evidence combination rules

Frederic Dambreville

http://email.fredericdambreville.com

Abstract: An approach for a generic implementation of combination rules of evidence is proposed. This approach implies a tripartite architecture, with respective parts implementing the logical framework (complete distributive lattice, Boolean algebra), the combination definition (referee function), and the belief-related processes (basic belief assignment, belief and plausibility computation, computations of the combinations according to their definitions). Referee functions are decisional arbitrament conditionally to basic decisions provided by the sources of information. Two generic processes are proposed for computing combination rule defioned by referee functions: a sampling method and a deterministic method based on an adaptive reduction of the set of focal elements. The proposed generic implementation makes possible the construction of several rules simply on the basis of a referee function extension.

Notations

- $I[boolean]$ is defined by $I[\mathbf{true}] = 1$ and $I[\mathbf{false}] = 0$.
- G^{Θ} denotes a *complete distributive lattice* or *Boolean algebra* generated by Θ ,
- $x_{1:n}$ and $\{x_j\}_{j=1:n}$ are abbreviations for the sequence $x_1, \cdots, x_n$.

1 Introduction

Evidence theory [1, 2] is often promoted as an alternative approach for fusing informations, when the hypotheses for a Bayesian approach cannot be precisely stated. When manipulating evidence-based informations, two issues may be encountered. First, there are potential combinatorics arising from the multiplication of the focal elements implied by the combination processes. Secondly, the interpretation of the evidence combination rules may be difficult, when conflicts are notably involved. In the recent litterature, there has been a large amount of work devoted to the definition of new fusion rules [3, 4, 6] , in order to handle the conflict efficiently. The choice for a rule is often dependent of the applications and there is not a systematic approach for this task. Somehow, it appears also that this choice of a rule implies the choice of decision paradigm in order to handle the conflict. Facing such variety of rule, it appears that there is a need for tools as well as common frameworks, in order to evaluate and compare these various rules in regards to

the applications.

This paper focuses on the definition of a system for the generic implementation of the combination rules. It allows an implementation of rules by means of simple and interpretable extensions of the system. All these extensions share common combination engines, which compute relaxation of the combinations and thus handle the combinatorics efficiently. Three keywords are thus the guideline of this work, *i.e* generic, efficient and interpretable.

Our approach is based on a tripartite architecture, with respective parts implementing the logical framework *(logical component)*, the combination definition *(referee component)*, and the belief-related processes *(belief component)*.

The logical framework represents the information while considered without its uncertainty. Typical logical frameworks are powersets, which are Boolean algebras. However, lattices are also useful frameworks and actually generalize the Boolean algebras. A common and incremental implementation is made by the logical component, adressing yet several frameworks.

The combination definition is obtained by implementing *referee functions*. Referee functions are decisional arbitraments conditionally to basic decisions provided by the sources of information. It is shown that referee functions [7] are sufficient for a definition of most combination rules. Then, a generic implementation of the rule is made possible on the basis of an algorithmic extension implementing the referee function.

The belief-related processes involve the encoding of basic belief assignment (bba), the computation of belief and plausibility from bba, but also the computations of the combinations according to their definitions. Two processes are proposed for computing the combination rules, both based on a definition of the combinations by means of referee functions: a sampling method and a summarization method, which provides an adaptive reduction of the set of focal elements. The purpose of these approaches is to avoid the combinatorics which are inherent to the definition of fusion rules of evidences.

Section 2 defines the logical component of our architecture. Section 3 defines the referee component. Section 4 introduces the combination processes of the belief component. Section 5 concludes.

2 Logical component

The logical component constitutes one of the three part of our system. While the combination of Dempster-Shafer [1, 2] is generally defined over powersets, the combination rules are defined on the basis of logical operators and properties which are not necessary inherent to powersets, nor Boolean algebras. Extensions of the theory have been proposed on the basis of distributive lattices [9]. Hyperpowersets are also distributive lattices. Most varients of evidence theory are built on logical structures known as *complete* (in general finite) *distributive lattices*. The logical component handles these structures in a generic manner, especially considering the logical notions from the general viewpoint of lattice.

These notions are defined now.

It is given Θ, a set of atomic propositions, which constitutes the logical frame of our knowledge representation. The set Θ is thus called the *frame of discernment*.

Distributive lattice: An algebraic structure G^Θ generated by Θ and operators $\cap$ and $\cup$ is a distributive lattice if it verifies the properties:

- $X \cup (Y \cup Z) = (X \cup Y) \cup Z$; $X \cap (Y \cap Z) = (X \cap Y) \cap Z$ (associativity)
- $X \cup Y = Y \cup X$; $X \cap Y = Y \cap X$ (commutativity)
- $X \cup (X \cap Y) = X$; $X \cap (X \cup Y) = X$ (absorption)
- $X \cap (Y \cup Z) = (X \cap Y) \cup (X \cap Z)$ (distributivity)

Implied order: The partial order $\subseteq$ is implied by $X \subseteq Y \overset{\Delta}{\iff} X \cap Y = X \iff X \cup Y = Y$.

Complete distributive lattice: A distributive lattice G^Θ is *complete* if any subset $\Xi \subset G^\Theta$ has a greatest lower bound and a least upper bounds. This property is implied by the finiteness of the structures in use.

Complement and co-complement: The complement, $\overline{X}$, and the co-complement, $\underline{X}$, are defined by:

$$\overline{X} = \bigcup_{Y : Y \cap X = \emptyset} Y \quad \text{and} \quad \underline{X} = \bigcap_{Y : Y \cup X = \Omega} Y \ .$$

Complete Boolean algebra: A complete distributive lattice G^Θ is a complete Boolean algebra if $\underline{X} \cap X = \emptyset$ and $\overline{X} \cup X = \Omega$ for any $X \in G^\Theta$. As a consequence, the complement and co-complement operators are the same for a complete Boolean Algebra.

Examples: A powerset, 2^Θ, generated by Θ is a complete Boolean algebra. An *hyper-power set*, D^Θ, generated by a *finite* Θ is a complete distributive lattice.

3 Referee component

The referee component constitutes the part of our system, which implements the definition of the combination rules. The definition of the combination rules is made by means of *referee function*. Referee functions have been defined in [7] and allow a simple, general and computational interpretation of the combination rules. This interpretation was inspired

first by works on probabilistic restrictions of evidence combinations. This work led to a conditionnal interpretation of the combination, in terms of *Referee Function*.

3.1 Referee function

Definition A referee function over G^Θ for s sources of information and with context γ is a mapping $X, Y_{1:s} \mapsto F(X|Y_{1:s};\gamma)$ defined on propositions $X, Y_{1:s} \in G^\Theta$, which satisfies for any $X, Y_{1:s} \in G^\Theta$:

$$F(X|Y_{1:s};\gamma) \geq 0 \text{ and } \sum_{X \in G^\Theta} F(X|Y_{1:s};\gamma) = 1 \ ,$$

A referee function for s sources of information is also called a s-ary referee function. The quantity $F(X|Y_{1:s};\gamma)$ is called a *conditional arbitrament* between $Y_{1:s}$ in favor of X. Notice that X is not necessary one of the propositions $Y_{1:s}$; typically, it could be a combination of them. The case $X = \emptyset$ is called the *rejection case*.

Fusion rule Let be given s basic belief assignments (bba) $m_{1:s}$ and a s-ary referee function F with context $m_{1:s}$. Then, the fused bba $m_1 \oplus \cdots \oplus m_s[F] \triangleq \oplus[m_{1:s}|F]$ based on the referee F is constructed as follows:

$$\begin{aligned} &\oplus[m_{1:s}|F](X) = \frac{I[X \neq \emptyset]}{1-z} \sum_{Y_{1:s} \in G^\Theta} F(X|Y_{1:s};m_{1:s}) \prod_{i=1}^{s} m_i(Y_i) \ , \\ &\text{where } z = \sum_{Y_{1:s} \in G^\Theta} F(\emptyset|Y_{1:s};m_{1:s}) \prod_{i=1}^{s} m_i(Y_i) \ . \end{aligned} \tag{1}$$

The value z is called the *rejection rate*.

3.2 Examples of referee functions

Dempster-shafer: The definition of a referee function for Dempster-Shafer combination is immediate:

$$m_{\mathrm{DS}} = \oplus[m_{1:s}|F_{\mathrm{DS}}] \ , \text{ where } F_{\mathrm{DS}}(X|Y_{1:s};m_{1:s}) = I\left[X = \bigcap_{k=1}^{s} Y_k\right] \ .$$

Disjunctive: The definition of a referee function for the disjunctive combination is:

$$m_{\mathrm{d}} = \oplus[m_{1:s}|F_{\mathrm{d}}] \ , \text{ where } F_{\mathrm{d}}(X|Y_{1:s};m_{1:s}) = I\left[X = \bigcup_{k=1}^{s} Y_k\right] \ .$$

PCR6: The proportional conflic redistribution rules (PCRn) have been introduced By Dezert and Smarandache [6] and the rule PCR6 has been proposed by Martin and Osswald in [5] . The original definition of the rule could be found there. A formulation of PCR6 by means of a referee function is derived in [7]:

$$m_{\mathrm{PCR6}} = \oplus[m_{1:s}|F_{\mathrm{PCR6}}] \ ,$$

where the referee function F_{PCR6} is defined by:

$$F_{\mathrm{PCR6}}(X|Y_{1:s};m_{1:s}) = I\left[X = \bigcap_{k=1}^{s} Y_k \neq \emptyset\right] + I\left[\bigcap_{k=1}^{s} Y_k = \emptyset\right] \frac{\sum_{j=1}^{s} I[X = Y_j]\ m_j(Y_j)}{\sum_{j=1}^{s} m_j(Y_j)} \ . \tag{2}$$

This referee function implies an interpretation of PCR6 as a two-cases process:

- The entry informations are compatible; then, the conjunctive consensus is decided,
- The entry informations are not compatible; then, a mean decision is decided, weighted by the relative beliefs of the entries.

PCR6 only considers consensus or no-consensus cases. But for more than 2 sources, there are many cases of *intermediate consensus*. By construction, PCR6 is not capable to manage intermediate consensus. This limitation is addressed by rule PCR# introduced in [7] and which is defined in the long version of the paper.

4 Belief component: combination

Two combination processes are defined, which are by sampling and by summarization. Both processes work on a same definition of the combination based on referee functions.

Process based on sampling. The process produces an approximation of the fused bba by means of a cloud of particles, which are elements of G^{Θ}.

The algorithm is described in the long version of the paper.

Process based on summarization. The principle is to limit the size of the set of focal elements by reducing this size by a simple summarization process[8]. In our approach, the less significant focal elements in term of basic belief assignment are both relaxed to their union, and the process is repeated as many times it is needed.

The algorithm is described in the long version of the paper.

5 Conclusion

This paper describes an architecture for generic implementation of combination rules of evidence. This architecture is tripartite, that is it is composed of a logical part, a definition part and a processing part. The logical part specifies on which kind of logical framework the rule is defined. The definition part specifies the combination itself, and is implemented by means of referee functions, which are arbitrament processes conditionally to the contributions of the sources of information. Altogether, these components make the architecture easily expandable and interpretable. This architecture is actually implemented under the form of a java package, and is made available at address:

`http://refereefunction.fredericdambreville.com/releases`

Future releases will focus on performance improvement and on the implementation of more combinations rules.

Literatur

[1] Dempster A.P., *Upper and Lower probabilities induced by a multivalued mapping*, Annals of Mathematical Statistics, vol. 83, pp. 325–339, 1967.

[2] Shafer G., *A mathematical theory of evidence*, Princeton University Press, 1976.

[3] Florea M.C., Dezert J., Valin P., Smarandache F. and Jousselme A.L., *Adaptative combination rule and proportional conflict redistribution rule for information fusion*, COGnitive systems with Interactive Sensors, Paris, France, March 2006

[4] Lefevre E., Colot O., Vannoorenberghe P., *Belief functions combination and conflict management*, Information Fusion Journal, Elsevier Publisher, Vol. 3, No. 2, pp. 149–162, 2002.

[5] Martin A. and Osswald C., *A new generalization of the proportional conflict redistribution rule stable in terms of decision*, Applications and Advances of DSmT for Information Fusion, Book 2, American Research Press Rehoboth, F. Smarandache and J. Dezert, pp. 69–88 2006.

[6] Smarandache F., Dezert J., *Proportional Conflict Redistribution Rules for Information Fusion*, Applications and Advances of DSmT for Information Fusion, Book 2, American Research Press Rehoboth, F. Smarandache and J. Dezert, pp. 3–68, 2006.

[7] Frédéric Dambreville, *Chap. 6: Definition of evidence fusion rules based on referee functions*, in Smarandache F. & Dezert J., Editors, *Applications and Advances on DSmT for Information Fusion (Collected Works)*, Vol. 3, American Research Press, 2009.

[8] Harmanec H., *Faithful approximations of belief functions*, in Uncertainty in Artificial Intelligence 15 (UAI99), Stockholm, 1999.

[9] Besnard P., Jaouen P., Perin J. Ph., *Extending the transferable belief model for inconsistency handling*, Information Processing and Management of Uncertainty, 1996.

Multiple ground target tracking and classification with DSmT

Benjamin Pannetier, Jean Dezert

benjamin.pannetier@onera.fr, jean.dezert@onera.fr

Abstract: Based on our previous work we propose to track multiple ground targets with GMTI (Ground Moving Target Indicator) sensors as well as with imagery sensors. The scope of this paper is to fuse the attribute type information given by heterogeneous sensors with DSmT (Dezert Smarandache Theory) and to introduce the type results in the tracking process to improve its performances.

1 Introduction

Data fusion for ground battlefield surveillance is more and more strategic in order to create the situational assessment or improve the precision of fire control system. For this, we develop new ground target tracking algorithms adapted to GMTI (Ground Moving Target Indicator) sensors. In fact, GMTI sensors are able to cover a large surveillance area during few hours or more if several sensors evolve on the same operational theatre. Several references exist for the MGT (Multiple Ground Tracking) with GMTI sensors [**?**, 8] whose fuse contextual informations with MTI reports. The main results are the improvement of the track precision and track continuity. Our algorithm [6] is built with several reflexions inspired of this literature. The proposed VS-IMMC (Variable Structure Interacting Multiple Models) filter is extended in a multiple target context and integrated in a SB-MHT (Structured Branching - Multiple Hypotheses Tracking).

One way to enhance data associations is to fused data obtained by several sensors. The most easily approach is to consider the centralized fusion between two or more GMTI sensors. Another way is to introduce heterogeneous sensors in the centralized architecture in order to improve the data associations (by using the reports in location and its classification attribute) and palliate the poor GMTI sensor classification. In our previous works [6], the classification information of the MTI segments and IMINT segments (IMagery INTelligence) has been introduced in the target tracking process. The idea was to maintain aside each target track a set of ID hypotheses. Their committed belief are revised in real time with the classifier decision through a very recent and efficient fusion rule called proportional conflict redistribution (PCR).

In this paper, in addition to the measurement location fusion, we illustrate on a complex scenario our approach to fuse MTI classification type with image classification type associated to each report.

2 Motion & observation models

2.1 Constrained motion model

The target state $\mathbf{x}(k)$ at the current time t_k is defined in a local horizontal plane (O, X, Y) of a Topographic Coordinate Frame denoted *TCF*. The target state on the road segment s is defined by $\mathbf{x}_s(k)$ where the target position $(x_s(k), y_s(k))$ belongs to the road segment s and the corresponding heading $(\dot{x}_s(k), \dot{y}_s(k))$ is in its direction. The event that the target is on road segment s is noted $e_s(k) = \{\mathbf{x}(k) \in s\}$. Given the event $e_s(k)$ and according to a motion model M_i, the estimation of the target state can be improved by considering the road segment s. The constrained motion model M_i^s is build in such a way that the predicted state is on the road segment s and the gaussian noise is defined under the road segment constaint [6]. After the state estimation obtained by a Kalman filter, the estimated state is then projected according to the road constraint $e_s(k)$. This process is detailed in [6].

2.2 GMTI measurement model

According to the NATO GMTI format [5], the MTI reports received at the fusion station are expressed in the WGS84 coordinates system. The MTI reports must be converted in the *TCF*. A MTI measurement z at the current time t_k is given in the *TCF*. Each MTI report is characterized both with the location and velocity information (range radial velocity) and also with the attribute information and its probability that it is correct. We denote C_{MTI} the frame of discernment on target ID based on MTI data. C_{MTI} is assumed to be constant over the time and consists in a finite set of exhaustive and exclusive elements representing the possible states of the target classification. In this paper, we consider only 3 elements in C_{MTI} defined as $C_{MTI} = \{$*Tracked vehicle*, *Wheeled vehicle*, *Rotary wing aircraft*$\}$.

We consider also the probabilities $P\{c(k)\}$ $(\forall c(k) \in C_{MTI})$ as input parameters of our tracking systems characterizing the global performances of the classifier. The vector of probabilities $[P(c_1)\ P(c_2)\ P(c_3)]$ represents the diagonal of the confusion matrix of the classification algorithm assumed to be used. Let $\mathbf{z}^{\star}_{MTI}(k)$ the extended MTI measurements including both kinematic part and attribute part expressed by the herein formula:

$$\mathbf{z}^{\star}_{MTI}(k) \triangleq \{\mathbf{z}_{MTI}(k), c(k), P\{c(k)\}\} \tag{1}$$

2.3 IMINT motion model

For the imagery intelligence (IMINT), we consider two sensor types : a video EO/IR sensor carried by a Unmanned Aerial Vehicle (UAV) and a EO sensor fixed on a Unattended Ground Sensor (UGS). We assume that the IMINT reports $\mathbf{z}_{video}(k)$ at the current time t_k are expressed in the reference frame (O, X, Y) and give a location information and type

target. We assume that the video information given by both sensor types are processed by their own ground stations and that the system provides the video reports of target detections with their classification attributes. For the last point, a human operator selects targets on a movie frame and is able to choose its attribute with a HMI (Human Machine Interface). Based on the military symbology called 2525C [3], we build the frame of discernment for an EO/IR source denoted C_{video}. Each video report is associated to the attribute information $c(k)(\forall c(k) \in C_{video})$ with its probability $P\{c(k)\}$ that it is correct. As C_{MTI}, C_{video} is assumed to be constant over the time and consists in a finite set of exhaustive and exclusive elements representing the possible states of the target classification.

Let $\mathbf{z}^{\star}_{video}(k)$ be the extended video measurements including both kinematic part and attribute part expressed by the following formula $(\forall c(k) \in C_{video})$:

$$\mathbf{z}^{\star}_{video}(k) \triangleq \{\mathbf{z}_{video}(k), c(k), P\{c(k)\}\} \tag{2}$$

The attribute type of the image sensors belongs to a different and better classification than the MTI sensors.

2.4 Taxonomy

In our work, the symbology 2525C [3] is used to describe the links between the different classification sets C_{MTI} and C_{video}. Figure 1 represents a short part of the 2525C used in this paper. The red elements underlined in italic style are the atomic elements of our taxonomy. Each element of both sets can be placed in 1. For example, the "wheeled vehicle" of the set C_{MTI} is placed at the level "Armoured $\rightarrow$ Wheeled" or the "Volkswagen Touareg" given by the video is placed at the levels "Armoured $\rightarrow$ Wheeled$\rightarrow$ Medium" and "Civilan Vehicle $\rightarrow$ Jeep $\rightarrow$ Medium".

3 Tracking with road constraints

3.1 VS IMM with a road network

The IMM is an algorithm for combining state estimates arising from multiple filter models to get a better global state estimate when the target is under maneuvers. In section 2.1, a constrained motion model i to a road segment s, noted $M^i_s(k)$, was defined. We extend the segment constraint to the different dynamic models (among a set of $r+1$ motion models) that a target can follow. The model indexed by $r = 0$ is the stop model. It is evident that when the target moves from one segment to the next, the set of dynamic models changes according to the road network configuration. The steps of the IMM under road segment s constraint are the same as for the classical IMM as described in [1].

In real applications, the predicted state could also appear onto another road segment, because of a road turn for example, and we need to introduce new constrained motion models.

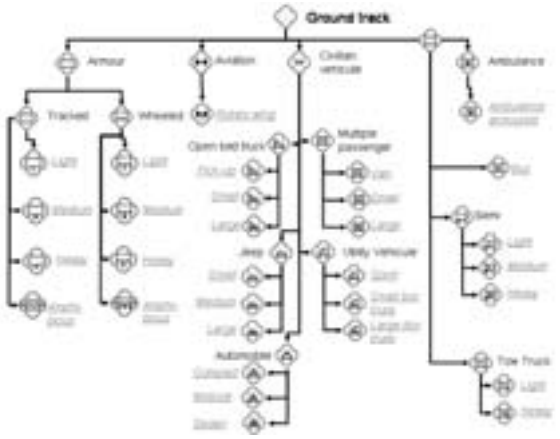


Figure 1: 2525C (light version).

In such case, we activate the most probable road segments sets depending on the local predicted statelocation of the track $T^{k,l}$[6]. We consider $r+1$ oriented graphs which depend on the road network topology. For each graph i, $i = 0, 1, \ldots, r$, each node is a constrained motion model M_s^i. The nodes are connected to each other according to the road network configuration and one has a finite set of $r+1$ motion models constrained to a road section. The selection of the most probable motion model set, to estimate the road section on which the target is moving on, is based on Wald's sequential probability ratio test (SPRT) [9].

3.2 Multiple target tracking

For the MGT problem, we use the SB-MHT (Structured Branching Multiple Hypotheses Tracking) presented in [2]. When the new measurements set $\mathbf{Z}(k)$ is received, a standard gating procedure is applied in order to validate MTI reports to track pairings. The existing tracks are updated with VS-IMMC and the extrapolated and confirmed tracks are formed. More details can be found in chapter 16 of [2]. In order to palliate the association problem, we need a probabilistic expression for the evaluation of the track formation hypotheses that includes all aspects of the data association problem. It is convenient to use the log-likelihood ratio (LLR) $L^l(k)$ or a track score of a track $T^{k,l}$ expressed at current time t_k.

4 Target type tracking

Our approach consists to use the belief on the identification attribute to revise the LLR with the *posterior* pignistic probability on the target type. We recall briefly the Target Type

Tracking (TTT) principle and explain how to improve VS-IMMC SB-MHT with target ID information. TTT is based on the sequential combination (fusion) of the predicted belief of the type of the track with the current "belief measurement" obtained from the target classifier decision. The adopted combination rule is the so-called Proportional Conflict Redistribution rule no 5 (PCR5) developed in the DSmT (Dezert-Smarandache Theory) framework since it deals efficiently with (potentially high) conflicting information. A detailed presentation with examples can be found in [4, 7].

4.1 Principle of the target type tracker

To estimate the true target type $type(k)$ at time k from the sequence of declarations $c(1)$, $c(2), \ldots c(k)$ done by the unreliable classifier up to time k. To build an estimator $\widehat{type}(k)$ of $type(k)$, we use the general principle of the Target Type Tracker (TTT) developed in [4] which consists in the following steps:

1. Initialization step (*i.e.* $k = 0$). Select the target type frame $C_{Tot} = \{\theta_1, \ldots, \theta_n\}$ and set the prior bba $m^-(.)$ as vacuous belief assignment, i.e $m^-(\theta_1 \cup \ldots \cup \theta_n) = 1$ since one has no information about the first observed target type.

2. Generation of the current bba $m_{obs}(.)$ from the current classifier declaration $c(k)$ based on attribute measurement. At this step, one takes $m_{obs}(c(k)) = P\{c(k)\} = C_{c(k)c(k)}$ and all the unassigned mass $1 - m_{obs}(c(k))$ is then committed to total ignorance $\theta_1 \cup \ldots \cup \theta_n$. $C_{c(k)c(k)}$ is the element of the known confusion matrix $\mathbf{C}$ of the classifier indexed by $c(k)c(k)$.

3. Combination of current bba $m_{obs}(.)$ with prior bba $m^-(.)$ to get the estimation of the current bba $m(.)$.

4. Estimation of True Target Type is obtained from $m(.)$ by taking the singleton of Θ, *i.e.* a Target Type, having the maximum of belief (or eventually the maximum Pignistic Probability).

5. Set $m^-(.) = m(.)$; do $k = k + 1$ and go back to step 2).

Naturally, in order to revise the LLR in our GMTI-MTT system for taking into account the estimation of belief of target ID coming from the Target Type Trackers, we transform the resulting bba $m(.) = [m^- \oplus m_{obs}](.)$ available at each time k into a probability measure.

4.2 Data attributes in the VS IMMC

To improve the target tracking process, the introduction of the target type probability is done in the likelihood calculation. For this, we consider the measurement $\mathbf{z}_j^*(k)(\forall j \in$

$\{1, \ldots, m_k\}$) described in (1) and (2). With the assumption that the kinematic and classification observations are independant, it is easy to prove that the new combined likelihood Λ_N^l associated with a track $T^{k,l}$ is the product of the kinematic likelihood.

5 Illustration

In the extended version of this paper, we will illustrate our algorithm by using a complex scenario generated with a powerful simulator developed at ONERA. The area of interest is located in a fictive country called North Atlantis. In this scenario, the goal is to detect and track several targets with 2 GMTI sensors (JSTARS, SIDM), 18 UGS and 4 UAV (SDTI), in oder to build the situation assessment and evaluate the threat in order to protect the coalition forces. On the operation theater, 250 targets evolve, they can maneuver on and out the road network. The set of target type is significant, we can have for instance civilian vehicles (as 4x4, cars, bus, truck,...) and military vehicles as well (T−62, AMX 30, Kamakov,...). llustrations and conclusion of our algorithm will be presented in the extended version of this paper.

References

[1] Y. Bar-Shalom and D. Blair, *Multitarget multisensor tracking : Applications and Advances*, Vol. III, Artech House, pp. 523, 2000.

[2] S.S. Blackman and R. Popoli, *Design and analysis of modern tracking systems*, Artech House, 1999.

[3] Defense Information Systems Agency, *Common Warfighting Symbology, MIL-STD 2525C*, Technical document, IPSC, Nov. 2008.

[4] J. Dezert, A. Tchamova, F. Smarandache and P.Konstantinova, *Target Type Tracking with PCR5 and Dempster's rules: A Comparative Analysis*, in Proc. of Fusion 2006, Firenze, Italy, July 2006.

[5] NATO, *STANAG 4607 JAS (Edition 2) - NATO ground moving target indicator GMTI) format*, NSA0749(2007)-JAS/4607, Aug. 2007.

[6] B. Pannetier, V. Nimier and M. Rombaut, *Multiple ground target tracking with a GMTI sensor*, in Proc. of MFI 2006, Sept. 2006.

[7] F. Smarandache, J. Dezert, *Advances and applications of DSmT for information fusion (Collected works)*, Vols. 1-3, American Research Press, 2004–2009. http://www.gallup.unm.edu/˜smarandache/DSmT.htm

[8] M. Ulmke, W. Koch, *Road-map assisted ground moving target tracking*, IEEE Trans. on AES, Vol. 42, No. 4, pp. 1264–1274, Oct. 2006.

[9] A. Wald, *Sequential Tests of Statistical Hypotheses*, Annals of Mathematical Statistics, Vol. 16, No. 2, pp. 117–186, June 1945.

Different tools for clutter mapping

Isabel Schlangen and Martina Daun
Fraunhofer FKIE, Dept. Sensor Data and Information Fusion
Neuenahrer Str. 20, 53343 Wachtberg, Germany
{isabel.schlangen, martina.daun}@fkie.fraunhofer.de

Abstract: Stationary clutter arises in a wide area of different radar or sonar applications. Without adaption to a non-uniform clutter distribution, tracking algorithms try to initialize tracks in these areas repeatedly while - in most cases - there is no target to be found.

The intention of this paper is to discuss different ways to enable clutter mapping in order to use this as an efficient tool for target tracking algorithms. The methods will be applied to the Range/Doppler tracking task in a multistatic passive radar application, and the performances of the different algorithms are discussed.

1 Introduction

Passive radar systems use illuminators of opportunity, no additional emitting source is necessary. This provides the advantage that the sensor does not reveal its own position, and in view of the application to air surveillance it is an important aspect that a passive system does not need a frequency permission. However, the signal waveforms are usually not perfectly suited for radar applications. Therefore passive radar systems usually deliver less precise measurements than active radar systems. Moreover, the measurements are often ambiguous. Thus, a fusion/tracking algorithm is needed to display the target information to the operator reliably.

In this paper we focus on DAB and DVB-T transmitters as illuminators of opportunity. A characteristic property of passive radar systems using DAB/DVB-T is the arrangement of the illuminators in so called single frequency networks (SFNs). As a consequence we cannot distinguish between measurements of individual illuminators within the same SFN. This means that the association of measurement and illuminator is a priori unknown and has to be solved by the tracking algorithm.

To handle the association problem in SFNs, a multi-stage tracking strategy is proposed in [DK08]. The first tracking stage operates in target parameters, i.e. bistatic range, bistatic Doppler and (if available) also azimuth. The main task of Range/Doppler (R/D) tracking is to remove false alarms before handling the association problem. A good performance of next tracking stage, the Cartesian tracking, is therefore highly dependent on the performance of the R/D tracking.

In the multistatic passive radar application that is studied here, clutter regions take shape as thin stripes spreading in Doppler. This paper shall not give an explanation for their appearance, but discusses ways to deal with this problem in context of R/D tracking.

Different approaches will be shown. In the first part of this paper, the Gaussian mixture reduction (GMR) algorithms by M. West [Wes93] and A. R. Runnalls [Run07] shall be shortly presented. Thereafter, there will be a revision of the Expectation Maximization (EM) algorithm, which is used in order to optimize the results of the other algorithms. The next part treats a greedy EM algorithm presented by Vlassis and Likas in 2000 [VL00]. After this, we will see different ways to use and combine the presented solutions, so that we can conclude the effect of clutter map usage on tracking algorithms. The algorithms are applied to real data of the Fraunhofer FHR [KMH07, KHSO09, OKH+10] and the resulting improvements with respect to R/D tracking are pointed out.

Due to space limitations we only give an outline of the algorithms and achieved results here, for details we refer to the long version of this paper which can be found in the electronic version of the proceedings.

2 Problem formulation

A common approach to create a clutter map is to divide the observation area into predefined clutter cells and to count the number of measurements appearing in the respective cell on a given time interval [MSMM05].

The focus of this paper is the representation of a clutter map by a Gaussian Mixture. The idea is that each measurement is represented by an expectation and a covariance matrix. Thus, from the view of a clutter statistic, a single measurement increases the clutter probability at the position of the expectation and its vicinity, whereas the expansion is described by the covariance.

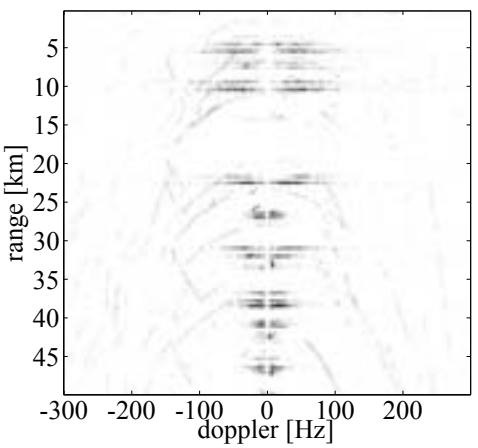


Figure 1: Representation of the observed area by a Gaussian mixture over 200 time scans. The stationary clutter regions appear as stripes of similar range.

In our specific application, a measurement at time t_i contains measurements of the bistatic range and Doppler given by $\mathbf{z}^{(i)} = (r^{(i)}, d^{(i)})$. The set of measurements at time t_i is further given by $\mathcal{Z}^i = \{\mathbf{z}_1^{(t_i)}, \ldots, \mathbf{z}_{N^{t_i}}^{(t_i)}\}$. A single measurement is by assumption Gaussian distributed, thus the set of measurements can be visualized by Gaussian mixture (GM), where for each $\mathbf{z}_j^{(t_i)}$ one summand $w_j^{t_i}\mathcal{N}(\mathbf{y}, \mathbf{z}_j^{(t_i)}, \mathbf{R}_j^{t_i})$ with weights $w_j^{t_i}$ and equally chosen covariances $\mathbf{R}_j^{t_i} = \mathrm{diag}(\sigma_r^2, \sigma_d^2)$ is defined for each pair i, j, so that the calculation of

$$f(\mathbf{y}) = \sum_{i=1}^{t_{max}} \sum_{j=1}^{N^{t_i}} w_j^{t_i}\mathcal{N}(\mathbf{y}, \mathbf{z}_j^{(t_i)}, \mathbf{R}_j^{t_i}), \quad \mathbf{y} \in \mathbb{R}^d. \tag{1}$$

To visualize (1) we place a grid on the section of interest and derive the $f(\mathbf{y})$-values for all grid points seperately; the result is shown in figure 1. One can observe some horizontal oriented lines which are interpreted as the unwanted clutter, as well as tracks of

some targets which cause the parabolic curves in the current R/D system[1]. Accumulating the data of more than a few time steps increases the complexity within the tracking algorithm. An efficient algorithm has to be found to decrease the information to its substance. With this motivation, we will compare and combine different algorithms and contrast their performance with respect to reduction of the Gaussian components.

3 Two GMR algorithms, EM algorithm and greedy EM.

To approximate clusters of points by only a few Gaussians, we use different methods to reduce the number of components of the Gaussian mixture in (1), i.e. we need to find a similar expression with a reduced number of K summands to approximate the old sum.

GMR algorithms We use and compare two algorithms for Gaussian mixture reduction following the approaches of M. West [Wes93] and A. R. Runnalls [Run07], referred to as GMR West and GMR Runnalls in the follwing.

The basic idea of GMR West is to delete the summand with lowest weight by merging it to its nearest neighbour, and then go on inductively, so that summands with little relevance will be sorted out first. Summands with weight 0 will be totally ignored.

The GMR Runnalls [Run07] has a more complicated structure than the rather straightforward solution by West. Runnalls' thought is to compare all possible pairs of summands in each merging step and to combine the two summands with highest compatibility. The comparison factor in this case is based on the so called *Kullback-Leibler discrimination*, which describes a divergence measure between two general probability density functions.

The EM and greedy EM algorithms The GMR techniques of Runnalls and West give us quite promising tools to reduce the number of GM components. However, it is recommendable to smooth the results of the GMR algorithms afterwards. We therefore use the Expectation Maximization (EM) algorithm as given in [VL00] and [OT96].

One more tool to build a clutter map is the greedy EM algorithm presented by Vlassis and Likas [VL00]. The strategy behind it has a slightly other view on our task since it does not try to merge certain summands like Runnalls and West proposed, but initializes by just one single Gaussian and later tries to reconstruct the old cluster inductively. This is provided by sequentially adding new components to the initialization, which fit best on the underlying old setting and simultaneously optimizing the Gaussians by the classical EM.

4 Application to multistatic passive radar data

Due to the wide multi-functionality of Gaussian mixtures, the introduced algorithms are very suitable for a vast number of circumstances. In this part of the paper we would like to present the tools in context of clutter map computation which occurs as a problem of target tracking. The clutter map is set up in Range/Doppler coordinates to assist the R/D tracking [DK08]. All algorithms will be tested with real sensor data from passive radar experiments with the passive radar demonstrator CORA, conducted at Fraunhofer FHR. A detailed description of CORA is given in [KHMA08].

[1] The little gap at Doppler zero is caused by an *a priori* clutter filtering at the receiver.

The data used for the clutter map analysis were obtained during a military exercise in Germany in 2007, where numerous aircraft operations and electronic warfare applications were trained. Illumination was provided by three local DAB single frequency networks.

Independent of the algorithms currently used, we will always start based on the following setting: The data that are to be analysed are stored in range and Doppler coordinates with units $[km]$ and $[Hz]$. The calculations are processing along 200 time steps, and for each time step, there will be approximately 100 measurements. Here we assume that the clutter is stationary during 200 time steps. In each step, the new data is added to the Gaussian and then optimized in different ways.

4.1 Creating the clutter map

Default setting The data that are to be analysed are stored in range and Doppler coordinates with units $[km]$ and $[Hz]$. The calculations are processing along 200 time steps $t_0, \ldots, t_{199}$, and for each time step t_i, there will be approximately $N^{t_i} \approx 100$ measurements. Here we assume that the clutter is stationary during 200 time steps. In each step, the new data is added to the Gaussian and then optimized in different ways, where one or two of the presented algorithms will be combined.

Implementation problems of the EM and greedy EM algorithms The EM algorithm assumes i.i.d. samples on the distribution that is to be approximated. In our application, the measurements are interpreted as Gaussian samples of a true target, a Clutter target or a false alarm, described by an expectation and a covariance. We apply the EM on two different settings, on the one hand as an intermediate step in the greedy EM algorithm, and on the other hand independently, where it needs i.i.d. samples of the Gaussian mixture that is to be optimized. In the latter case, it is necessary to add a covariance matrix $\mathbf{R} = \text{diag}(\sigma_r^2, \sigma_d^2)$ to the covariances of the Gaussians, as justified in the large version of this paper. With this approach we overcome difficulties of the EM due to small sampling numbers (causing non positive definite covariances) and therefore overcome computational problems.

The initialization of the greedy EM algorithm requires a very sensitive choice. In our case, accuracy is improved by choosing just one element of the original data set at random, setting the weight of it to 1.

Results of clutter map generation In Fig. 2, the result of the combination of greedy EM algorithm and the classical EM is shown.

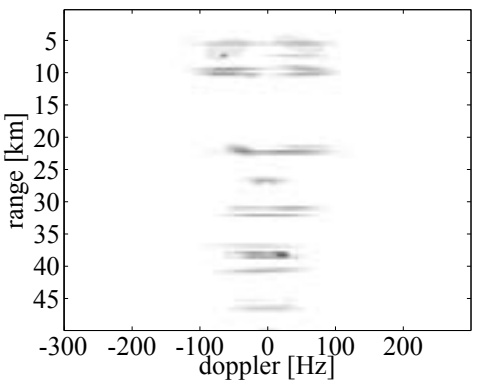


Figure 2: greedy EM algorithm and classical EM

The greedy EM algorithm gives an initialization based on the first 20 time steps and then, the classic EM algorithm is applied in every of the following time steps.

This scenario performs with high accuracy and displays the clutter map with only 63 Gaus-

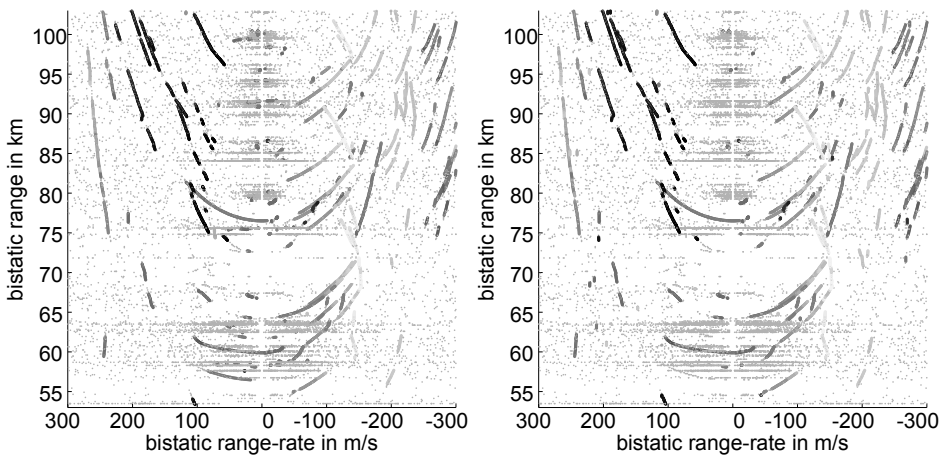
sian summands; hence we will use the Gaussian sum produced by the two EM routines for the tracking algorithm. Please see the full version of this paper for more results on clutter map generation.

4.2 The clutter map as a tool for target tracking

After the discussion of the different tool combinations, we will now explore the efficiency of using clutter maps in tracking algorithms. For generating R/D tracks we use Multi-hypothesis Tracking (MHT) [KKU06]. It is applied with different parameter settings of the Clutter value ϱ_F. The data deliver additional measurements of the azimuth direction, that is processed in the R/D tracking algorithm, but not considered in the R/D clutter map generation. For the azimuth we assume here a uniform false measurement background.

In the standard target tracking scenario we assume a uniform false alarm background, which means that ϱ_F is chosen to be fixed. This assumption is not suited for modelling the clutter distribution, which is reflected in the tracking results, see Fig. 3(a). A large number of false tracks is extracted in the regions of enhanced clutter intensity, since the tracker notices a deviation from the background model.

This effect can be counteracted by calculating a clutter map over several time scans. The results of the clutter map (with only 63 Gaussian components) generated by combination of Greedy EM initialization and EM (fig. 3(b)) give a better performance, generation of false tracks is significantly reduced. For the other tracks the tracking performance is quite comparable, but we can note some slight degradation for the clutter map approach. A track at range between 55km and 60km (lying close to clutter region) is not found by the tracker, when estimating the false alarm intensity using the clutter map.



(a) fixed clutter intensity (b) clutter map usage for clutter intensity estimation

Figure 3: Doppler-Range Tracking Results of real data of the Fraunhofer FHR/PSK: The measurements are shown by black dots, whilst the tracks are illustrated in different colors (representing the value of the azimuth estimate).

5 Conclusion and Acknowledgements

As the results in the last section show, clutter map computation is of great importance in order to reduce the amount of false alarms by a tracking algorithm. Based on the underlying Gaussian mixture representation of the clutter map, different algorithms were shown that reduce Gaussian mixtures and optimize the reduction results. It could be observed that, despite the loss of a few tracks that coincide with high appearance of clutter, the usage of a clutter map generated by the greedy EM algorithm can improve the R/D tracking by sorting out most of the generated false tracks.

The authors would like to thank the department *Passive Sensor Systems, Electronic Countermeasures and Classification* (PSK) of the Fraunhofer FHR located in Wachtberg (Germany) for the provision of the real passive radar data.

References

[DK08] M. Daun and W. Koch. Multistatic target tracking for non-cooperative illumination by DAB/DVB-T. In *Proc. IEEE Radar Conference RADAR '08*, pages 1–6, 26–30 May 2008.

[KHMA08] H. Kuschel, J. Heckenbach, S. Mueller, and R. Appel. On the potentials of passive, multistatic, low frequency radars to counter stealth and detect low flying targets. In *in IEEE Radar Conf.*, pages 1443–1448, may 2008.

[KHSO09] H. Kuschel, J. Heckenbach, J. Schell, and D. O'Hagan. Passive Radar for Homeland Defence. *Proceedings of the IRS*, 2009.

[KKU06] W. Koch, J. Koller, and M. Ulmke. Ground target tracking and road map extraction. *ISPRS Journal of Photogrammetry and Remote Sensing*, 61:197–208, 2006.

[KMH07] H. Kuschel, M.Glende, and J. Heckenbach. Experimental passive radar systems using digital illuminators (DAB/DVB-T). *Proceedings of the IRS*, 2007.

[MSMM05] D. Mušicki, S. Suvorova, M. Morelande, and B. Moran. Clutter Map and Target Tracking. *Proceedings of the 8th International Conference on Information Fusion 2005*, 2005.

[OKH+10] D. W. O'Hagan, H. Kuschel, J. Heckenbach, M. Ummenhofer, and J. Schell. Signal reconstruction as an effective means of detecting targets in a DAB-based PBR. *Proceedings of the IRS*, 2010.

[OT96] D. Ormoneit and V. Tresp. Improved Gaussian Mixture Density Estimates Using Bayesian Penalty Terms and Network Averaging. *Advances in Neural Information Processing Systems 8*, pages 542–548, 1996.

[Run07] A. R. Runnalls. Kullback-Leibler Approach to Gaussian Mixture Reduction. *IEEE Trans. on Aerospace and Electronic Systems, vol. 43, no. 3*, pages 989–999, 2007.

[VL00] N. Vlassis and A. Likas. A Greedy EM Algorithm for Gaussian Mixture Learning. *Neural Processing Letters 15*, pages 77–87, 2000.

[Wes93] M. West. Approximating Posterior Distributions by Mixtures. *Royal Statistical Society, vol. 55, no. 2*, pages 409–422, 1993.

Spatial Design of Transmitter-Receiver-Constellations for PCL-Systems using GIS based Visibility Analysis

Martin Rick
Fraunhofer FKIE, Dept. Sensor Data and Information Fusion
Neuenahrer Str. 20, 53343 Wachtberg, Germany
martin.rick@fkie.fraunhofer.de

Abstract: Passive Coherent Location (PCL) requires detailed knowledge about the geometric transceiver-receiver-constellations and the technical parameters for simulating the effective field of view. Geographic Information Systems (GIS) provide tools to fuse spatial information with technical metadata to evaluate individual geometric constellations and obtain optimal radar coverage. This paper presents a GIS-based model procedure for planning a passive radar system using base stations of the Global System for Mobile Communications (GSM) as illuminator. The implementation of GIS-technology allows analyzing spatial information about illuminated areas in combination with the underlying technical metadata.

1 Introduction

During the last decades PCL radar has been established as a subset of bistatic radar with various types of signals for illumination (e.g. DAB, DVB, FM signals). The high global distribution of the GSM technology dedicates this system as an ideal illuminator for passive radar [ZNW09, LTS07]. Furthermore, all Base Transceiver Stations (BTS) send individual signals with constant power which makes it easy to identify each signal [TSL+05a]. A central task when installing such a system is the geometric evaluation of variable receiver locations in relation to the BTS which can be considered constant in location, transmitting frequency, orientation and antenna characteristics.
Mobile telecommunication companies use GIS-based visibility analyses in the field of network planning [vL02]. For the considered application we extend the classical visibility analysis methods by combining the illuminated areas with the technical parameters of the original transmitter station. Finally, the results are checked against the observation space of receiving antenna to determine the effective field of view. The results of this investigation proof the usability of GIS-based spatial analysis in the field of data fusion for the geometric planning of a GSM-based passive radar system.

2 GSM based passive radar

In contrary to the conventional (active) radars, PCL systems use electromagnetic signals of multiple foreign, non-cooperative transmitters as illuminators. With its multiple transmitters PCL is considered as a subtype of a bistatic radar constellation and should be treated as a multi-static constellation (see fig. 1) [Nic10]. The receiver antenna compares the direct reference signal from the transmitter stations and the target echo for locating the target position. The whole system is controlled by the receiver configuration as variable part of the system [TSL05b].

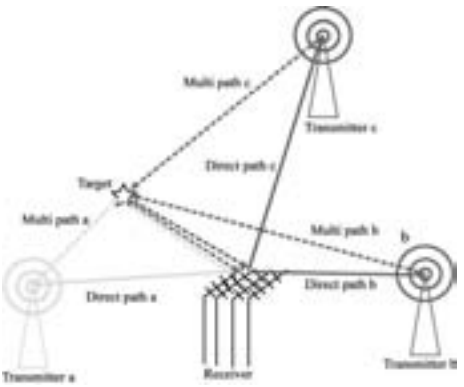


Figure 1: Passive radar schema

Each sending unit serves one grid cell within the telecommunication network. Arriving signals get identified by their individual Broadcast Channel (BCCH), the served network cell (Cell ID) and the Local Area Code (LAC). It requires extensive knowledge about all sending parameters of the serving station in order to interpret the reference signal and the run-time-difference to the echo signal correctly. To analyse the potentials of GSM-based PCL within a dedicated area three tasks have to be fulfilled:

1. Simulation of BTS illuminated areas
2. Simulation of observed areas for individual receiver positions
3. Intersection of both to calculate the effective field of view

3 Visibility analysis for PCL planning

With regard to passive radar applications, the knowledge about the accessible field of view in a network of multiple independent transmitters is essential. The determination of visible areas from one or more given observation points is a key feature of geographic data processing [Coo05, KRW97, LTS07].

3.1 Viewshed analysis

ESRI's ArcGIS 9.3.1 provides a viewshed tool that determines visible areas within the continuous surface of a Digital Elevation Models (DEM) to a set of observer features [ESR10b]. The tool calculates a Line-of-Sight (LOS) between each raster cell and the observer location. To verify if an object can be seen from a certain location, the elevation angle from the observer to the object is checked against all elevation points of the surface on a straight line between the observer and the target. If only one profile point has a greater elevation angle than the object, it has to be classified as not visible (see fig. 2) [KRW97].



Figure 2: Let P equal a set of profile points $\{p1, p2, p3, ..., pn\}$. If ∃ a point pi$_{\in}$ P where θ_{pi} ¿ θ_{object} the object is not line-of-sight (taken from [KRW97]).

The fact that radio waves get absorbed by intervening topography makes visibility analysis to an appropriate method for simulations in the field of radar applications [EA08]. It is favourable to use a Digital Terrain Model (DTM) instead of the DEM because it takes all surface objects into account that could blockade the radio wave propagation. In the case of radar applications, visibility can be interpreted as illumination in space [Ric10].
When simulating radio wave propagations the different behaviour of electromagnetic and optical waves has to be taken into account. Depending on the frequency (or wave length) this is often achieved with a correction of the earth diameter and the refractivity coefficient of visible light in the air. The default settings for these two parameters in ArcGIS are 12740 km as diameter and 0.13 as refractivity coefficient. The formula for the correction of surface and elevation units reads [ESR10a]:

$$Z_{actual} = Z_{Surface} - \frac{Dist^2}{Diam_{Earth}} + R_{refr} \cdot \frac{Dist^2}{Diam_{Earth}} \tag{1}$$

$Dist$ is the planimetric distance between the observation feature and the observed location, while $Diam_{Earth}$ and R_{refr} denote the diameter of the earth and the refractivity coefficient of visible light in the air. The implementation of the Fresnel zone could improve the approximation and deliver more precise simulation results.

3.2 Extended visibility analysis

Whereas classical visibility tools only return a raster dataset that divides the surface into visible and non-visible areas, the requirements for radio wave propagations are more complex. Passive radar systems need information which transmitters illuminate a particular part of the observation room. To extend the results of the visibility analysis it is necessary to fuse the results with additional information about the transmitter location, frequencies and other metadata. Due to the limited information capacity of a raster, it is advisable to use vector formats instead. Therefore, a new GIS algorithm was developed in Python to extend the capabilities of the standard viewshed tool (see fig. 3) [Ric10].

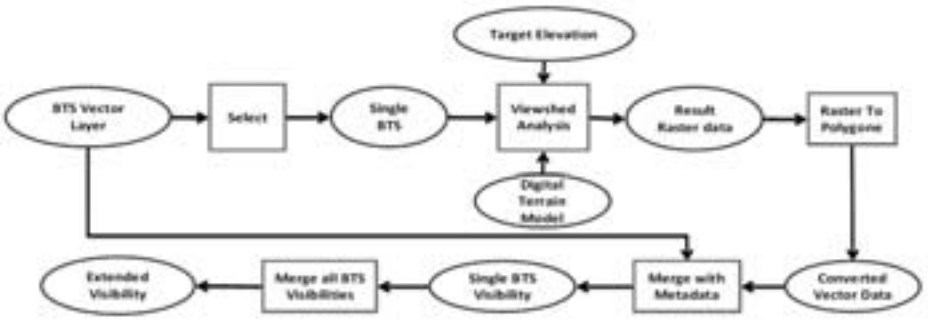


Figure 3: Function scheme for the extended viewshed analysis

The fundamental problem of visibility analysis is to identify the locations on the surface which can be seen from a single observer position [KRW97]. This would generally exclude the detection of air targets. The new extended viewshed tool involves a ”Target Elevation” parameter to consider air targets. While the transmitter parameters in a GSM network are considered as constant, the tool can be used in a batch process to simulate the visibility for multiple target elevations in one turn. Thus, it is possible to build a database of illuminations for different target elevations. With these results, it is possible to evaluate the entire observation room with regard to all parameters from multiple transmitters including possible intersections of illuminations (see fig. 4). The tool performance strongly depends on the number of transmitters, the spatial extent and the raster resolution of the DTM.

3.3 Effective observation space

Although the transmitter illumination is a key factor for PCL observation, the total system also strongly depends on the physical settings of the receiving antenna. Only the combination of transmitters and receiver allows the determination of the effective spatial observation space [Ric10]. Therefore, another tool was developed that simulates the view-

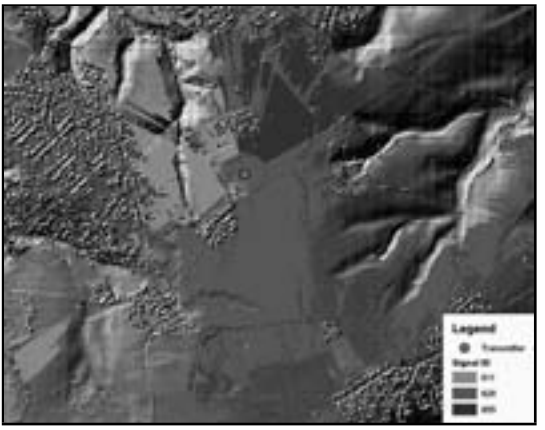


Figure 4: Visibility analysis for a single base station with three transmitter antennas

shed for individual receiver locations and intersects the results with the polygons from the extended visibility tool). The user sets the parameters for the receiver position, observation direction and target elevation. The resulting polygons contain the transmitter parameters and represent the effective observation space for the chosen receiver settings.

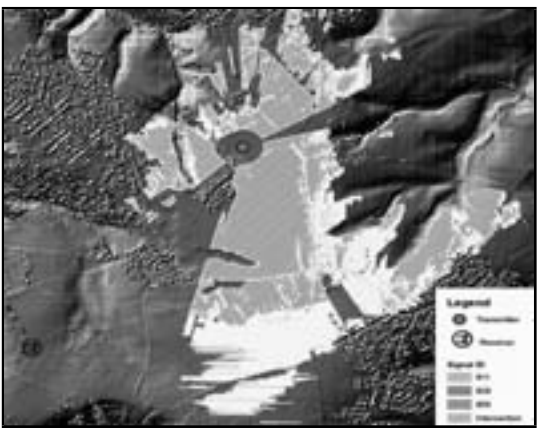


Figure 5: Effective passive radar observation potential

4 Results

This paper shows the potentials of GIS-based visibility analysis as an approach to radio wave propagations for PCL applications. Due to the particular requirements of PCL, it

was necessary to extend the classical methods to fuse multiple transmitters with individual technical parameters for comprehensive evaluations. The results of the new extended viewshed tool intersected with the receiver's observation space returns the effective field of view for passive radar applications for a given spatial extent. Even if the common visibility tool does not simulate the electromagnetic wave propagation it is sufficient for preliminary investigations on PCL. The case study uses GSM base stations as illuminators but the overall concept could also be adapted to other frequencies (e.g. FM radio, DVB-T, etc.) as well. Further investigations could involve reflection characteristics or the approximation to the real radio wave propagation.

References

[Coo05] G.R.J. Cooper. Analysing potential field data using visibility. *Computers and Geosciences*, 31:877–881, 2005.

[EA08] Thomas R. Etherington and Shelley M. Alexander. Using Viewsheds to Determine Area Sampled by Ground-Based Radiotelemetry. *Journal of Wildlife Management*, 72:1043–1046, 2008.

[ESR10a] ESRI, Redlands. *How visibility works*, 2 2010.

[ESR10b] ESRI, Redlands. *Viewshed*, 4 2010.

[KRW97] David B. Kidner, Philip J. Rallings, and J. Andrew Ware. Parallel Processing for Terrain Analysis in GIS Visibility as a Case Study. *GeoInformatica*, 1:183–207, 1997.

[LTS07] Y. L. Lu, D. K. P. Tan, and H. B. Sun. Air Target Detection and Tracking Using a Multi-Channel GSM Based Passive Radar. In *2007 International Waveform Diversity and Design Conference*, pages 122–126, 2007.

[Nic10] Dr. Ulrich Nickel. Features and Design Parameters of Passive Localization Based on GSM Base Station Illuminators. Technical report, Fraunhofer-FKIE, 2010.

[Ric10] Martin Rick. Einsatz Geographischer Informationssysteme bei der Sensordatenfusion, Wachtberg. Master's thesis, University Trier, Wachtberg, 4 2010.

[TSL+05a] D.K.P. Tan, H. Sun, Y. Lu, M. Lesturgie, and H.L. Chan. Passive radar using Global System for Mobile communication signal: theory, implementation and measurements. *IEE Proceedings - Radar, Sonar and Navigation*, 152(3):116 – 123, June 2005.

[TSL05b] D.K.P Tan, Hongbo Sun, and Yilong Lu. Sea and Air Moving Target Measurements Using a GSM Based Passive Radar. In *Radar Conference, 2005 IEEE International*, pages 783 – 786, Singapore, May 2005.

[vL02] Kees van Loo. Telekommunikationsmarkt. *arcaktuell*, 1:6–8, 2002.

[ZNW09] Reda Zemmari, Dr. Ulrich Nickel, and Dr. Wulf-Dieter Wirth. GSM Passive Radar for Medium Range Surveillance. In *Proceedings of the 6th European Radar Conference*, 2009.

Extended Object and Group Tracking: A Comparison of Random Matrices and Random Hypersurface Models

Marcus Baum[◇], Michael Feldmann[†], Dietrich Fränken[‡],
Uwe D. Hanebeck[◇], and Wolfgang Koch[†]

[◇]Intelligent Sensor-Actuator-Systems Laboratory (ISAS), Institute for Anthropomatics, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany
[†]Dept. Sensor Data and Information Fusion, Fraunhofer FKIE, Wachtberg, Germany
[‡]Data Fusion Algorithms & Software, EADS Deutschland GmbH, Ulm, Germany
marcus.baum@kit.edu, {michael.feldmann | wolfgang.koch}@fkie.fraunhofer.de
dietrich.fraenken@eads.com, uwe.hanebeck@ieee.org

Abstract: Based on previous work of the authors, this paper provides a comparison of two different tracking methodologies for extended objects and group targets, where the true shape of the extent is approximated by an ellipsoid. Although both methods exploit usual sensor data, i.e., position measurements of varying scattering centers, the distinctions are a consequence of the different modeling of the extent as a symmetric positive definite random matrix on the one hand and an elliptic random hypersurface model on the other. Besides analyzing the fundamental assumptions and a comparison of the properties of these tracking methods, simulation results are presented based on a static tracking environment to highlight especially the differences in the update step for the extension estimate.

Short Summary

The explicit consideration of target extents is becoming more and more important in the development of modern tracking systems, where an object is considered as extended if it is the source of multiple measurements at the same time. The specific set of problems arises from different scattering centers, named measurement sources, which give rise to several distinct radar detections varying, from scan to scan, in both number as well as relative origin location. In order to improve robustness and precision of estimation results, it is felt desirable to track the target extent in addition to the kinematic state of the object. More than these quantities cannot safely be estimated as well in the (opposite) case, where limited sensor resolution causes a fluctuating number of detections for a group of closely spaced targets and thus, prevents a successful tracking of (all of) the individual targets. In this case, it is suitable to consider this group of point targets as a *single* extended object.

Several suggestions for dealing with this problem can be found in literature. For an early work, see [DBP90]. An overview of existing work up to 2004 is given in [WD04]. In general, implicit and explicit models for the scattering centers, i.e., the measurement sources, can be distinguished. Explicit target models [VIG04, VIG05, IG03] aim at estimating the

location of the measurement sources in addition to kinematics of the object. These approaches require data association and are not treated in this paper. An implicit model for the target extent assumes that the measurement sources are generated independently according to an internal state of the object that reflects the extent. An example for an implicit model is a so-called spatial distribution model, where each measurement source is assumed to be an independent random draw from a probability distribution [GGMS05, GS05].

In this paper, we discuss two different tracking methods for extended object tracking, where the measurement sources are modeled implicitly and the true physical extension is approximated by an ellipsoid. Ellipsoidal shapes are highly relevant for real world applications, as an ellipsoid provides useful information about the target orientation and spatial extent. Here, the focus is on track maintenance, while estimation under uncertain observation-to-track association in the possible presence of missed detections and false alarms is considered out of scope.

The first tracking method considered in this paper is based on the use of symmetric positive definite (SPD) random matrices and has been introduced in [Koc08]. Therein, an SPD random matrix to describe the ellipsoid is the counterpart of a random vector representing the centroid. The decisive point for tracking of extended objects in [Koc08] is the special interpretation of usual sensor data, whereupon each measurement is considered as a measurement of the centroid scattered over object extension. This implicit neglect of any statistical sensor error has caused some further investigations and developments to honor the fact that both sensor error and extension contribute to the measurement spread [FF09]. Hence, we employ two approaches: the original one of [Koc08] as well as the advanced one of [FF09]. The second tracking method is based on a random hypersurface model (RHM), which has been introduced in [BNH10]. An *Elliptic RHM* specifies the relative Mahalanobis distance of a measurement source to the center of the target object by means of a one-dimensional random scaling factor. In order to avoid the treatment of an SPD matrix for the ellipsoid, the kinematic state of the extended object is supplemented with the non-zero entries of the Cholesky factorization of the inverse SPD matrix. By this means, all parameters describing the corresponding ellipsoid are contained in one random vector.

This paper analyzes the fundamental assumptions of these three tracking approaches and provides a comparison of the properties of the two different methods. It is completed by simulation results based on a static tracking environment to highlight especially the differences in the update step for the extension estimate.

References

[BNH10] M. Baum, B. Noack, and U. D. Hanebeck. Extended Object and Group Tracking with Elliptic Random Hypersurface Models. In *Proc. 13th Int. Conf. on Information Fusion (Fusion 2010)*, Edinburgh, United Kingdom, 2010.

[DBP90] O. Drummond, S. Blackman, and G. Petrisor. Tracking clusters and extended objects with multiple sensors. In *Proc. SPIE Conf. Signal & Data Processing of Small Targets*, Orlando, Florida, 1990.

[FF09] M. Feldmann and D. Fränken. Advances on Tracking of Extended Objects and Group Targets using Random Matrices. In *Proc. 12th Int. Conf. on Information Fusion (Fusion 2009)*, Seattle, Washington, 2009.

[GGMS05] K. Gilholm, S. J. Godsill, S. Maskell, and D. Salmond. Poisson Models for Extended Target and Group Tracking. In *Proc. SPIE Conf. Signal & Data Processing of Small Targets*, San Diego, California, 2005.

[GS05] K. Gilholm and D. Salmond. Spatial Distribution Model for Tracking Extended Objects. *IEE Radar Sonar Navig.*, 152(5):364–371, 2005.

[IG03] N. Ikoma and S. J. Godsill. Extended Object Tracking with Unknown Association, Missing Observations, and Clutter using Particle Filters. In *IEEE Workshop on Statistical Signal Processing*, St. Louis, Missouri, 2003.

[Koc08] W. Koch. Bayesian Approach to Extended Object and Cluster Tracking using Random Matrices. *IEEE Trans. Aerosp. Electron. Syst.*, 44(3):1042–1059, 2008.

[VIG04] J. Vermaak, N. Ikoma, and S. J. Godsill. Extended object tracking using particle techniques. In *Proc. IEEE Aerospace Conf.*, Big Sky, Montana, 2004.

[VIG05] J. Vermaak, N. Ikoma, and S. J. Godsill. Sequential Monte Carlo Framework for Extended Object Tracking. *IEE Radar Sonar Navig.*, 152(5):353–363, 2005.

[WD04] M. J. Waxman and O. Drummond. A Bibliography of Cluster (Group) Tracking. In *Proc. SPIE Conf. Signal & Data Processing of Small Targets*, Orlando, Florida, 2004.

Bearings Only Versus Bearings and Extent Tracking for Missile Guidance Systems Utilising Particle Methods

Alexander Watson[1], Lyudmila Mihaylova[1], Dave Vorley[2]
[1] Lancaster University, UK, [2] MBDA UK, Filton, Bristol
Email: {a.watson2, mila.mihaylova}@lancaster.ac.uk, Dave.Vorley@mbda-systems.com

Abstract: This paper addresses the problem of bearings only and bearings and extent tracking for closed loop missile guidance systems using particle methods compared with derivative free methods. Lack of observability is a typical phenomenon for the bearings only tracking problem. We demonstrate that exploiting angular extent information does help mitigating the lack of observability issues. The simulation results show that the particle filter ourperforms the derivative free algorithms.

1 Motivation

In recent years there has been considerable interest in the application of particle filters and derivative-free filters to target tracking problems, e.g. [ASS03, FRB02]. However, with a few exceptions [Gat09, SEG01], the applications are essentially open-loop; control is largely ignored. Here the application is a closed loop missile homing guidance problem.

For missile homing guidance, an estimator is required in order to calculate states relating to the relative motion of the target with respect to the missile. Traditionally Kalman Filters, or Extended Kalman Filters (EKFs), are used. However, when significant system dynamics or measurement non-linearities exist, and non Gaussian system or measurement noise is present, it is well known that Kalman Filtering has its limitations. Such conditions exist in homing guidance applications. Seeker measurements can have significant non-linearities, and target manoeuvres, can often be best described by non-linear models with non-Gaussian noise. In addition what a missile sees affects what it does, and what it does affects what it sees. Thus any 'parasitic' errors within the system cause missile body-motion to corrupt seeker measurements. In this situation it is not the open loop accuracy of the estimation that is of prime importance, but how errors build up. Closed-loop stability becomes of primary importance. Generally this limits the innovation gains and associated bandwidth of conventional estimators. A further restriction on bandwidth is that noise propagation onto the missile control surfaces must be kept within bounds. However, if the filter model is well matched to the 'real-world' dynamics/measurements good estimation may be obtained using much lower filter gains (or their equivalent). Alternative approaches such as particle filtering or derivative-free techniques are worth exploring.

This work investigates the performance of particle filters (sequential Monte Carlo based methods) within such a closed loop environment, and to what extent the resulting system can be characterised and tuned to be robust to parameter changes. In particular a generic

particle filter and the application of so called derivative-free methods in efficient and stable square root form is investigated. Such a technique was developed by MBDA in the late 80's [Vor92]. There now exist a wide variety of similar methods which also aim to improve statistical accuracy. Different approaches to the derivation of these filters are proposed such as use of the unscented transformation leading to the Square Root Unscented Kalman Filter (SRUKF) [WM01] or use of a divided-difference formulations [NPR04]. Here the SRUKF is used.

The problem considered is strap-down sightline-rate estimation for missile homing guidance. The missile carries a seeker and inertial instruments fixed to the missile body (gyros and accelerometers) that in effect allow it to measure the angular bearing of a target. The further case is considered where it is also assumed that onboard signal processing allows the angular extent of some fixed feature on the target to be measured. This is potentially useful extra information particularly if, as here, the problem is restricted to that of homing onto a fixed asset. Without extent information the problem is a bearings only problem. Bearings only tracking has been a problem widely studied in the literature in different contexts [ASS03, FRB02, BC07, CVY05], but mainly for open loop systems.

The remaining part of this paper is organised as follows. Section 2 presents the state equations for the missile guidance system and the measurement model. The proposed approximate solutions to the filtering problem for closed loop guidance are outlined in Section 3. Results are presented in Section 4. Finally, Section 5 concludes the results.

2 System Dynamics and Observation Model

The state vector is chosen as follows: $\boldsymbol{x} = \left(\psi_s - \psi_b, \dot{\psi}_s, 1/R, v_c/R, d/R\right)'$, where ψ_s is the sightline angle, ψ_b is the missile body angle, $\dot{\psi}_s$ is the sightline rate, R is the relative distance of the missile with respect to the target, v_c is the closing speed, d is a fixed distance and $'$ is the transpose operation. Figure 1 a) shows the angles in the engagement geometry. The ratio d/R represents the angular extent of some feature on the target (d is unknown however is 10 metres in the example). Figure 1 b) illustrates this. The system dynamics may be described by the following equations

$$\frac{d(\psi_s - \psi_b)}{dt} = \dot{\psi}_s - \dot{\psi}_b, \tag{1}$$

$$\frac{d\dot{\psi}_s}{dt} = 2\left(\frac{v_c}{R}\right)\dot{\psi}_s - a_m \cos(\psi_s - \psi_b))\left(\frac{1}{R}\right), \tag{2}$$

$$\frac{d\left(\frac{1}{R}\right)}{dt} = \frac{1}{R}\left(\frac{v_c}{R}\right), \tag{3}$$

$$\frac{d\left(\frac{v_c}{R}\right)}{dt} = \left(\frac{v_c}{R}\right)^2 + (a_m \sin[(\psi_s - \psi_b)])\frac{1}{R}, \tag{4}$$

$$\frac{d\left(\frac{d}{R}\right)}{dt} = \left(\frac{v_c}{R}\right)\left(\frac{d}{R}\right). \tag{5}$$

A modified polar coordinate system is used which tends to prevent sightline rate estimates needed for guidance being corrupted when range is uncertain or unobservable. Note that

$\left(\frac{v_c}{R}\right) \approx \left(\frac{1}{t_{go}}\right)$ where $t_{go,k}$ is the time to go. A digital approximation to the state equations over a sample time T (here 0.01 sec) is used within the filtering. $\dot{\psi}_b T$ and $a_m T$ are provided by the gyro and accelerometer (here assumed accurate). The seeker measurement

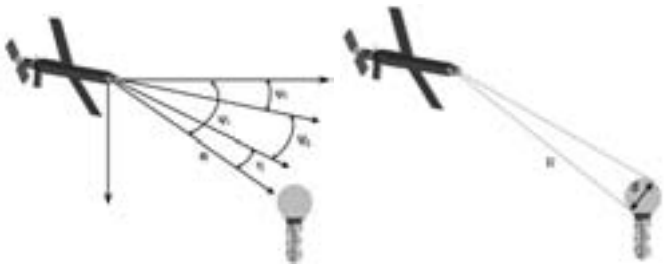


Figure 1: a) Definition of Angles for Sightline Estimation, b) Angular Extent

vector is defined by $\boldsymbol{z}_k \in \mathbb{R}^4$ consisting of: $\boldsymbol{z}_k = (\eta, R, v, d/R)'$, where η is the boresight error. The measurement equations are written in the form:

$$z_{1,k} = \eta + \left(\frac{R}{R_{ref}}\right)^{stype} \nu_{1,k}, \tag{6}$$

$$z_{2,k} = R + \nu_{2,k}, \tag{7}$$

$$z_{3,k} = v_c + \nu_{3,k}, \tag{8}$$

$$z_{4,k} = \left(\frac{d}{R}\right) + \nu_{4,k}. \tag{9}$$

The primary seeker measurement is of boresight error $\eta = (\psi_s - \psi_b) - \psi_g$ where ψ_g is the seeker gimbal angle, assumed to be accurately determined for each measurement time and $\left(\frac{R}{R_{ref}}\right)^{stype} \nu_{1,k}$ is a range dependent boresight error measurement noise term, depending upon whether the seeker type (stype) is passive, semi-active or active assuming values (0,1 and 2) respectively, and ν_k is a vector of noise inputs such that $\nu_k \in \mathbb{R}^4$ and $\nu_k \sim \mathcal{N}(0, \Sigma_{\nu_k})$ that is, normally distributed with zero mean and covariance Σ_{ν_k}.

In this study two types of measurement are considered: *i*) boresight error measurement only (B); *ii*) boresight error and an angular extent measurement d/R (BE). The extent characterises the size of the object and is provided by imagery or radar data.

3 Proposed Solutions

3.1 The Particle Filter

Particle filters [GSS93, AMGC02] have been successfully applied to many nonlinear estimation problems [Gat09, SEG01]. The strength of particle methods, which subsume the sigma point filters, consists in their ability to represent multi-modal posterior densities by

a finite set of point masses (particles). The particles are propagated through the true system state and measurement equations rather than approximation of these functions based on linearisation. The posterior probability density function is represented by a weighted statistical sample, from which the distribution's statistics can be approximated. The way in which the point masses are selected and how the weights are ascribed are the principal difference between the algorithms in this class of filter. Hence, the main steps of this approach are prediction and update of the samples, followed by a resampling stage aimed at introducing diversity in the samples.

3.2 The Square Root Unscented Kalman Filter

The Square Root Unscented Kalman Filter (SRUKF) [WM01] is a deterministic Minimum Mean Square Error estimator which propagates the square root of the *a posteriori* covariance rather than the full covariance matrix. QR decomposition and Cholesky updating replace weighted sums in the one step prediction and measurement update equations. The SRUKF has numerical stability as well as computational complexity advantages over non-square root form, with a reduced dynamic range and assured positive definiteness of the covariance matrix. Sigma points filters select the regression points deterministically, and are therefore not Monte Carlo methods.

4 Performance Analysis

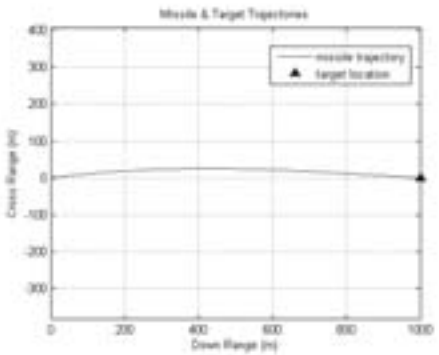


Figure 2: Trajectory of the missile with a static target

The simulated engagement geometry is shown in Figure 2. The seeker type utilised in these simulations is passive with no range dependency in the boresight error. The target is **static**. The filters are initialised with 10% perturbed initial conditions for the range and range rate compared with their true values.

4.1 Results Comparative Analysis of Filter Performance

Figure 3 shows the Root Mean Square Error (RMSE) between the estimated state value and the true values over 200 independent Monte Carlo runs for: a) boresight error and b) sightline rate. Similarly Figure 4 shows RMSE results for: a) range and b) time-to-go, respectively. In respect of boresight error measurement the SRUKF(B) performs worst. The addition of extent information promotes the SRUKF to the top position. However, the PF(B and BE) give similar results. Notice that with bearings only information the PF gives the best results. Similar conclusions apply based on sightline rate estimation.

Results for time-to-go estimation show that the PF(BE) starts to converge towards the

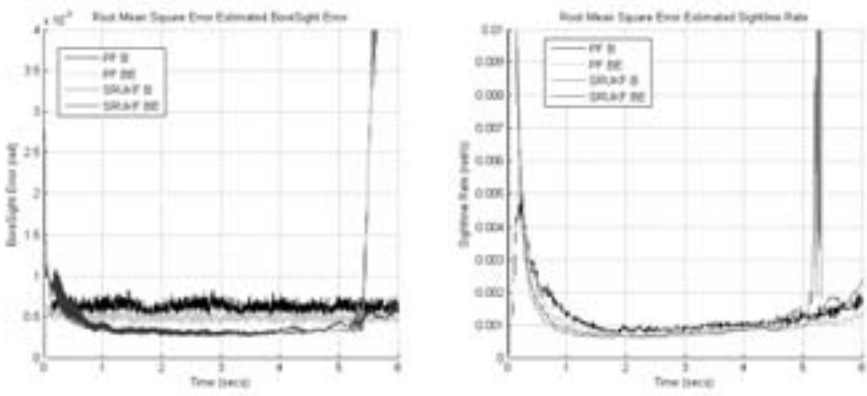


Figure 3: a) RMSE for boresight error, b) RMSE for sightline rate for PF and SRUKF

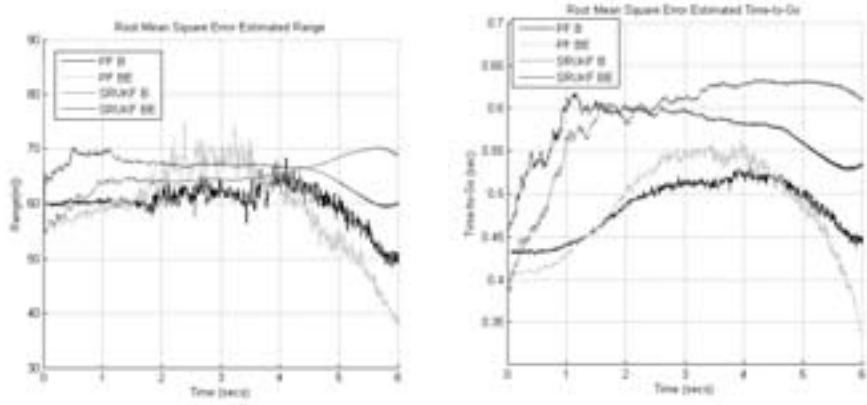


Figure 4: a) RMSE for range, b) RMSE for time to go for PF and SRUKF

end of the engagement. The filters PF(B) and SRUKF(B) do not converge as one might expect because range and closing velocity are not observable. For range and time-to-go estimation the PF(B and BE) clearly outperform the derivative methods.

5 Conclusions

Bearing and extent measurements can help in mitigating the lack of observability phenomenon. The angular extent information enhances by performance of all filters (PFs and UKFs). The experiments show that boresight error and sightline rate estimation are comparable for PF and SRUKF. Particle methods and other derivative free filters provide a computationally efficient approximate solution to the filtering problem.

Acknowledgements. The authors are grateful to the support from the EPSRC, MBDA and ITP project UKG 6473 "Particle Methods for Estimation and Control".

References

[AMGC02] M. Arulampalam, S. Maskell, N. Gordon, and T. Clapp. A Tutorial on Particle Filters for Online Nonlinear/Non-Gaussian Bayesian Tracking. *IEEE Trans. on Signal Proc.*, 50(2):174–188, 2002.

[ASS03] D.S. Angelova, I. Simeonova, and T. Semerdjiev. Monte Carlo Algorithm for Ballistic Object Tracking with Uncertain Drag Parameter. In *LSSC*, volume 2907, 2003.

[BC07] T. Bréhard and J.P. Le Cadre. Hierarchical particle filter for bearings only tracking. *IEEE Trans. on Aero. and Electronic Systems*, 43(4), 2007.

[CVY05] J. M. C. Clark, R. B. Vinter, and M. Yaqoob. The Shifted Rayleigh Filter: a New Algorithm for Bearings Only Tracking. *IEEE Trans on Aero and Elec Systems*, 2005.

[FRB02] A. Farina, B. Ristic, and D. Benvenuti. Tracking a Ballistic Target: Comparison of Several Nonlinear Filters. *IEEE Trans. on Aero. and Elec. Systems*, 38(3), 2002.

[Gat09] M.S Gate. The Use of Particle Based Technology to Produce Optimal Guidance. In *AIAA Guidance, Navigation, and Control Conference*, August 2009.

[GSS93] N. Gordon, D. Salmond, and A. Smith. A Novel Approach to Nonlinear / Non-Gaussian Bayesian State Estimation. *IEE Proceedings-F*, 140:107–113, April 1993.

[NPR04] M. Nørgaard, N.K. Poulsen, and O. Ravn. Advances in Derivative-Free State Estimation for Nonlinear Systems. Technical report, Technical University of Denmark, 2004.

[SEG01] D. Salmond, N. Everett, and N. Gordon. Target tracking and guidance using particles. In *Proc. of the American Control Conf.*, volume 6, pages 4387 – 4392, 2001.

[Vor92] D.H Vorley. Square Root EKF with Numerical Linearisation. *IMIA and Royal Aero Soc. Collected Papers from a Conference on Aero Vehicle Dynamics and Control*, 1992.

[WM01] E. Wan and R. Van Der Merwe. Chapter 7 The Unscented Kalman Filter. In *Kalman Filtering and Neural Networks*, pages 221–280. Wiley, 2001.

Multi-Target Tracking with Partially Unresolved Measurements

Daniel Svensson
Dept. Signals and Systems
Chalmers Univ. of Technology
Gothenburg, Sweden
daniel.svensson@chalmers.se

Martin Ulmke
Dept. SDF
Fraunhofer FKIE
Wachtberg, Germany
martin.ulmke@fkie.fraunhofer.se

Lars Danielsson
Active Safety Electronics
Volvo Car Corporation
Gothenburg, Sweden
ldanie25@volvocars.com

Abstract: The problem of limited sensor resolution, although usually ignored in target tracking, occurs in multi-target scenarios whenever the target distance falls below the size of the sensor resolution cell. Typical examples are the surveillance of aircraft in formation, and convoy tracking for ground surveillance. Ignoring the limited sensor resolution in a tracking system may lead to degraded tracking performance, in particular unwanted track-losses. In this paper, an extension of the resolution model by Koch and van Keuk to the case of arbitrary object numbers is discussed. The model is incorporated into the Joint Probabilistic Data Association Filter (JPDAF) and applied to a simulated scenario with partially unresolved targets.

1 Introduction

Generally, target tracking is separated into data assignment and kinematic filtering (see e.g. [BP99]). For both steps appropriate sensor models have to be developed which are simple enough to be feasible for a tracking algorithm, however, still cover the essential features important to tracking such as measurements errors and resolution properties. When the distance between targets is small compared to the measurement error, the assignment between targets and detections becomes ambiguous, leading to a fast growing number of possible data assignments. When, on the other hand, the target distance becomes smaller than the size of the resolution cell, there is an additional source of ambiguity as a given detection may result from the measurement of two or more targets. Usually, this resolution problem is neglected in the design of target tracking algorithms. In many situations, this is a reasonable assumption, but there are important cases when the resolution capabilities of the sensor cannot be ignored [DF94]. Typical examples of when objects are closely spaced in relation to the sensor resolution is the tracking of aircraft in formation, and convoy tracking for ground surveillance. For such applications, ignoring the limited sen-

sor resolution may lead to degraded tracking performance, in particular due to premature deletion of tracks.

The ability of a sensor to resolve several objects can be described by the resolution probability. An important aspect of sensor resolution is its dependence on the sensor–target distance. This property has been covered in the resolution model by Koch et al. [KvK97] but is absent in traditional grid-based approaches [CBS84, MCC87]. Until recently, to the best of our knowledge, the existing resolution models have been limited to only two targets. The present paper describes a generalization of the resolution model [KvK97] to arbitrary target numbers and the incorporation of the model into the Joint Probabilistic Data Association Filter (JPDAF) [FBSS83]. We also present results of a simulated scenario with resolution limitations, where the JPDA filter with the proposed resolution model is shown to yield better tracking results than the ordinary JPDAF.

2 Problem formulation

The ultimate goal of a tracking filter is to calculate the posterior density $p\big(\mathbf{x}_k \big| \mathbf{Z}^k\big)$ of the joint target state vector $\mathbf{x}_k$, given all measurements $\mathbf{Z}^k$ up to, and including, the current time index k. For the N-target case, the joint target states is described as

$$\mathbf{x}_k = \Big[\big(\mathbf{x}_k^{(1)}\big)^T \quad \big(\mathbf{x}_k^{(2)}\big)^T \quad \dots \quad \big(\mathbf{x}_k^{(N)}\big)^T \Big]^T, \tag{1}$$

where $\mathbf{x}_k^{(i)}$ is the state of target i. Further, the collection of measurements up to time index k is given by the set $\mathbf{Z}^k = \big\{\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_k\big\}$.

In order to perform the calculation of the posterior density in the presence of resolution uncertainty, we need to model the measurements from an unresolved group of targets (henceforth called group target), and to describe the probability of resolution for that group. We also need to model the motion of the targets. At time index k, a sensor produces M_k observations which are either target-generated or spurious. As is common in most tracking algorithms, we assume that the clutter measurements are described by a spatially homogeneous Poisson process with intensity λ.

Group measurement model and process model: We assume a simple model for the measurement of an unresolved group of targets (henceforth referred to as a group measurement). The model does not capture the true nature of a group measurement, but serves the purpose of illustrating the proposed framework. In general, more refined group measurement models can be used within the framework.

The assumed model states that a group measurement can be described as a measurement of the group center, where the center point is given by the arithmetic mean of the involved target states. That is, for an unresolved group of n_g targets (possibly one), whose state vectors are gathered in the joint vector $\mathbf{x}_k^g$, their group measurement $\mathbf{z}_k^{t,(j)}$ is described by

$$\mathbf{z}_k^{t,(j)} = \mathbf{H}_{n_g} \mathbf{x}_k^g + \mathbf{u}_k^{g,n_g}, \tag{2}$$

where $\mathbf{H}_{n_g} = [\mathbf{H}, \cdots, \mathbf{H}]/n_g$ and where $\mathbf{u}_k^{g,n_g} \sim \mathcal{N}\big(\mathbf{0}, \mathbf{R}_k^{n_g}\big)$ is the measurement noise for an n_g-target group. Typically, the measurement noise covariance matrix increases with increased number of targets in the group.

The (potentially correlated) motion of the multiple targets are described, as usual, by a discrete time Gauss-Markov process model.

Resolution model: The multi-target resolution model of [SUD10] describes the resolution probability for a group of arbitrary, but known, number of targets, and it is a generalization of the two-target model in [KvK97]. For a certain pair of targets $\mathbf{x}_k^{(i)}$, $\mathbf{x}_k^{(j)}$, the probability that they are unresolved is, according to [KvK97],

$$P_u(\mathbf{x}_k^{(i)}, \mathbf{x}_k^{(j)}) = \prod_{l=1}^{N_{\text{res}}} e^{-\ln(2)\left(\frac{\Delta^{i,j}x_l}{\alpha_{x_l}}\right)^2}, \tag{3}$$

where N_{res} is the dimension of the measurement space (2 for range and azimuth), $\Delta^{i,j}x_l$ is the distance between the predicted positions of targets i and j in dimension x_l, and α_{x_l} describes the resolution capabilities of the sensor in dimension x_l. The dimension can for example be $x_l = r$ for range, or $x_l = \varphi$ for azimuth angle. The probability $P_u(\mathbf{x}_k^{(i)}, \mathbf{x}_k^{(j)})$ can also be written as a scaled multivariate Gaussian [KvK97]

$$P_u(\mathbf{x}_k^{(i)}, \mathbf{x}_k^{(j)}) = \left|2\pi\mathbf{R}_{u,N_{\text{res}}}\right|^{1/2}\mathcal{N}\big(0; \Delta^{i,j}\mathbf{x}_k, \mathbf{R}_{u,N_{\text{res}}}\big), \tag{4}$$

where $\Delta^{i,j}\mathbf{x}_k$ is a vector with the distances between the target positions in each dimension, and $\mathbf{R}_{u,N_{\text{res}}} \propto \text{diag}\{\alpha_{x_1}^2, \ldots, \alpha_{x_{N_{\text{res}}}}^2\}$. For range and azimuth measurements, and small $\Delta\varphi$, this can easily be generalized to Cartesian coordinates by a rotation by the average angle $\bar{\varphi} = (\varphi^{(1)} + \varphi^{(2)})/2$. The Gaussian shape of the resolution probability will allow Kalman filter-like updates of the probability density function (pdf).

To extend the model to arbitrary target numbers, we assume the resolution probability to be pairwise independent. Resolution events can be described by a graph representation [SUD10], where each node (vertex) in the graph represents a target. Two targets are then mutually unresolved if they are connected by an edge in the graph. Further, a group of targets is unresolved if there exists a walk in the graph through all target vertices. Consequently, there are several graphs which generate the same target group. The probability for a group of unresolved targets as well as for the whole graph, or "subgroup pattern", $\mathcal{V}$, can be expressed in terms of products of the probabilities for being unresolved or resolved, P_u or $(1 - P_u)$, respectively [SUD10]. As one exploits a "negative" sensor output for estimating the multi-target state, it is sometimes denoted as "negative information" update.

3 Data Assignment

In standard tracking approaches [BP99], the unknown assignment between targets and detection are treated as hidden variables, and the updated pdf is marginalized over all

feasible data assignments. In the case of possible unresolved measurements, the sum over assignments is split into two steps: first we sum over all possible target subgroup patterns $\mathcal{V}$, defined above. Applying Bayes theorem, the sum is:

$$p\big(\mathbf{x}_k\big|\mathbf{Z}^k\big) \propto \sum_{\mathcal{V}} \Pr\big\{\mathcal{V}\big|\mathbf{x}_k\big\} p\big(\mathbf{Z}_k\big|\mathcal{V},\mathbf{x}_k\big) p\big(\mathbf{x}_k\big|\mathbf{Z}^{k-1}\big) \tag{5}$$

In a second step, we marginalize the likelihood function in the sum (5) above over all data assignments $d \in \mathcal{D}(\mathcal{V})$ for a given subgroup pattern $\mathcal{V}$:

$$p\big(\mathbf{Z}_k\big|\mathcal{V},\mathbf{x}_k\big) = \sum_{d\in\mathcal{D}(\mathcal{V})} p\big(\mathbf{Z}_k\big|\mathcal{V},d,\mathbf{x}_k\big) \Pr\big\{d|\mathcal{V},\mathbf{x}_k\big\}. \tag{6}$$

Explicit expression for the pdf and likelihood function are given in [SUD10].

4 JPDA filter incorporation

Under the assumptions mentioned above (linear, Gaussian measurement and process models) the updated multi-target pdf, eventually, is given as:

$$p\big(\mathbf{x}_k\big|\mathbf{Z}^k\big) = \sum_{\mathcal{V}} \sum_{u\in\mathcal{U}(\mathcal{V})} \sum_{d\in\mathcal{D}(\mathcal{V})} c^{\mathcal{V},u,d} \mathcal{N}\big(\mathbf{x}_k; \mathbf{x}_{k|k}^{\mathcal{V},u,d}, \mathbf{P}_{k|k}^{\mathcal{V},u,d}\big), \tag{7}$$

where the normalized scaling factors $c^{\mathcal{V},u,d}$ depend on the data association probability $\Pr\{d\big|\mathcal{V},\mathbf{x}_k\}$, the graph and state vector likelihood $p\big(\mathbf{Z}_k\big|\mathcal{V},d,\mathbf{x}_k\big)$ and the graph probability $\Pr\{\mathcal{V}\big|\mathbf{x}_k\}$. The sum over $u \in \mathcal{U}$ comes from the so-called "negative-information" update, discussed in Sec. 2, where each update with a resolved pair of targets doubles the number of Gaussian components.

The standard JPDA filter makes a Gaussian approximation of a Gaussian mixture at each time step. Similarly, a JPDA filter with a resolution model should make a Gaussian approximation of the mixture in (7). In principle, JPDA approximations on different level are possible [SUD10]. Here, we first perform a moment matching over the data association hypotheses for a given graph, and then make a second moment matching over the set of graphs, after negative information update.

5 Simulation Results

We study a simple simulated tracking example with two targets. In the scenario, the targets are first approaching each other, then move in parallel, and finally separate (see Fig. 1). Due to limited sensor resolution, the two targets are not always resolved. Since the targets are closely spaced, there is a big risk of track coalescence in the ordinary JPDA filter. Example output from the JPDA filter without and with resolution model incorporation can be found in Fig. 1. As seen in the figures, the tracks from the JPDA filter with resolution

model are more separated, due to a better description of the received measurements. The targets move with a constant speed of 5m/s, and the true separation of the targets is 40m in the parallel section. The sensor resolution cell is chosen as a square with edge length 40m in Cartesian coordinates.

For the evaluation of the tracking performance, we use as a first criterion the Mean Optimal Subpattern Assignment (MOSPA) measure [SVV08], which is a measure that disregards target identity and thus only considers the estimation of where there are targets. The second measure is the Root Mean Square Error (RMSE) of the position error. In Fig. 2 (left), the MOSPA performance, averaged over 500 Monte Carlo runs, is shown for the JPDA filter with and without resolution model, and for the case of perfect resolution, i.e., when the targets are always resolved. The results clearly show that the JPDA filter with resolution model performs better than the filter without resolution model. The performance in RMSE sense, averaged over 500 Monte Carlo runs, is shown in Fig. 2 (right). Obviously, also the RMSE performance is improved when the resolution model is incorporated in the JPDA filter. The increased RMSE of the JPDA filter in the end of the scenario is due to the fact that there is a larger probability of track switch for that filter, compared to the filter with resolution model.

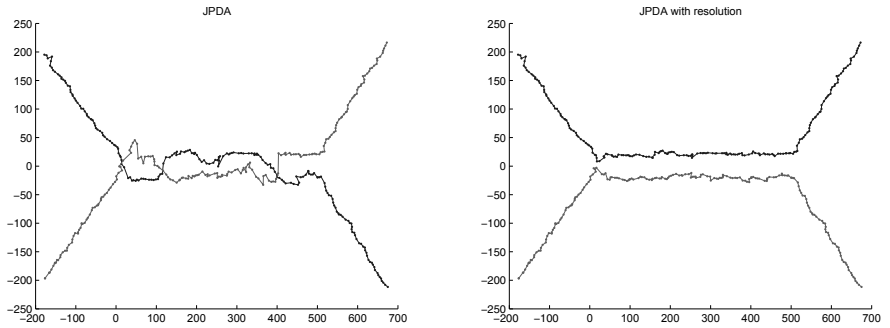


Figure 1: Typical tracking result for the JPDAF without (left) and with (right) resolution model.

6 Summary and Outlook

We have proposed a sensor resolution model for arbitrary target numbers, which is an extension of the two-target model by Koch and van Keuk [KvK97]. The resolution model leads to additional data association possibilities and to a multi-target likelihood function that takes missed detections due to merged measurements into account. As the filter update, in general, is infeasible, the extensive Gaussian mixture has been approximated by a JPDA filter. The application to a simple target tracking scenario shows a significantly improved tracking performance if the sensor resolution model is used. While these first results are encouraging, more simulations are needed to confirm the practicability of the presented approach. In particular, the stability against resolution model mismatch needs

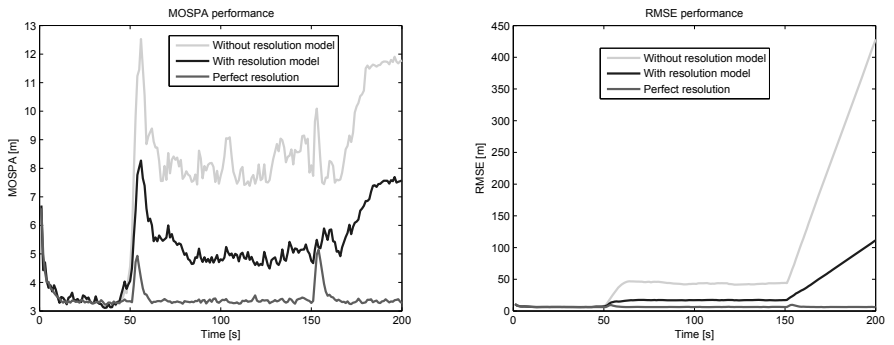


Figure 2: OPSA (left) and RMSE (right) performance of the JPDAF without resolution model (top), with resolution model (middle) in comparison to the case of perfect sensor resolution (buttom).

to be investigated. Furthermore, the resolution model and the corresponding likelihood function can also be integrated into more sophisticated tracking filters such as the Multi Hypothesis Tracker or the Probability Hypothesis Filter.

References

[BP99] Samuel Blackman and Robert Popoli. *Design and Analysis of Modern Tracking Systems*. Artech House, Norwood, MA, 1999.

[CBS84] K.C. Chang and Y. Bar-Shalom. Joint probabilistic data association with possibly unresolved measurements and maneuvers. *IEEE Transactions on Automatic Control*, 29(7):585–594, July 1984.

[DF94] F.E. Daum and R.J. Fitzgerald. Importance of resolution in multiple-target tracking. In *Signal and Data Processing of Small Targets*, volume 2235 of *Proc. SPIE*, 1994.

[FBSS83] T.E. Fortmann, Y. Bar-Shalom, and M. Scheffe. Sonar tracking of multiple targets using joint probabilistic data association. *IEEE Journal of Oceanic Engineering*, 8(3):173–183, July 1983.

[KvK97] W. Koch and G. van Keuk. Multiple hypothesis track maintenance with possibly unresolved measurements. *IEEE Transactions on Aerospace and Electronic Systems*, 33(3):883–892, July 1997.

[MCC87] Shozo Mori, K.C. Chang, and C.Y. Chong. Tracking Aircraft by Acoustic Sensors; Multiple Hypothesis Approach Applied to Possibly Unresolved Measurements. In *American Control Conference*, pages 1099–1105, June 1987.

[SUD10] D. Svensson, Martin Ulmke, and Lars Danielsson. Multitarget sensor resolution model for arbitrary target numbers. In *Signal and Data Processing of Small Targets*, volume 7698 of *Proc. SPIE*, 2010.

[SVV08] D. Schuhmacher, B.-T. Vo, and B.-N. Vo. A Consistent Metric for Performance Evaluation of Multi-Object Filters. *IEEE Transactions on Signal Processing*, 56(8):3447–3457, August 2008.

Tutorien

Erfolgreiche Anfängerausbildung im Programmieren

Marcus Crestani
Universität Tübingen
crestani@informatik.uni-tuebingen.de

Michael Sperber
DeinProgramm
sperber@deinprogramm.de

Abstract: Das Tutorial gibt einen Überblick über Inhalte und Techniken zur effektiven Anfängerausbildung im systematischen Programmieren. Der im Tutorial vermittelte Ansatz baut auf den *design recipes* bzw. *Konstruktionsanleitungen* des TeachScheme!- bzw. DeinProgramm-Projekts auf – einem System expliziter Anleitungen zur systematischen Konstruktion von Programmen. Diese Anleitungen werden durch eine Progression spezieller auf Anfänger zugeschnittener Programmiersprachen ermöglicht, die in einer Programmierumgebung, die ebenfalls speziell auf Anfänger zugeschnitten ist. Ein über viele Jahre entwickeltes Übungskonzept bringt das Konzept zur vollen Entfaltung. Dieser Artikel beschreibt Motivation und wichtige Bestandteile des Lehransatzes und demonstriert die Anwendung der Konstruktionsanleitungen an einem Beispiel. Zum Lehransatz sind ein universitäres Lehrbuch, *Die Macht der Abstraktion*, ein wachsendes Schulbuch, umfangreiche Software sowie weiteres Lehrmaterial (darunter eine umfangreiche Aufgabensammlung) verfügbar.

1 Einleitung

Programmieranfänger haben andere Bedürfnisse als fertig ausgebildete professionelle Programierer. Traditionelle Programmierkurse tragen diesen Bedürfnissen nur unzureichend Rechnung: Das Ergebnis sind oft Absolventen, die nicht oder nur unsystematisch und fehlerhaft programmieren können [BDH+08a].

Der im Tutorial vermittelte Ansatz baut auf den *design recipes* [FFFK04] bzw. *Konstruktionsanleitungen* des TeachScheme!/DeinProgramm-Projekts [KS07] auf. Das Konzept ist die Grundlage für die Vorlesungen „Informatik I“ an den Universitäten Tübingen und Freiburg und wird dort seit mehreren Jahren erfolgreich praktiziert und verbessert [BDH+08b]. Es hat sich durch mehrere Evaluationen über verschiedene Dozenten als robust erwiesen und wird kontinuierlich aktualisiert und verbessert.

Die Konstruktionsanleitungen fangen mit einer Analyse der Daten eines Problems an und liefern Schritt für Schritt die Elemente des Programms, die sich aus der Struktur der Daten ergeben. Die systematische Konstruktion wird durch eine Progression spezieller, auf Anfänger zugeschnittener Programmiersprachen ermöglicht, die eine seichte Lernkurve mit direktem Anschluss an den Mathematikunterricht über den Umgang mit Daten hin zu

umfangreicheren Programmen (z.B. reaktiven Animationen) ermöglichen. Diese laufen in einer Programmierumgebung, die speziell auf Anfänger zugeschnitten ist. Ein über viele Jahre entwickeltes Übungskonzept bringt den Ansatz zur vollen Entfaltung.

2 Traditionelle Programmierausbildung

Die traditionelle Programmierausbildung tut sich oft schwer, das Programmieren erfolgreich zu vermitteln:

- Professionelle Programmiersprachen (wie z.B. C++, Java und Pascal/Delphi) sind sehr umfangreich; die Behandlung ihrer Sprachelemente, der Syntax und idiosynkratischen Aspekte der Semantik nimmt einen Großteil der Ausbildung in Anspruch. Damit bleibt wenig Zeit, die Grundkonzepte des Programmierens zu behandeln [Köl99a].
- Programmierumgebungen, die für professionelle Software-Entwickler entwickelt wurden, setzen sie vieles voraus, das Anfänger erst noch lernen müssen. Außerdem liefern sie für Anfänger ungenügendes Feedback, wenn es Probleme gibt. [Köl99b, FCF+02].
- Damit bleiben frühe Erfolgserlebnisse aus, die für die Freude am eigenständigen Lernen wichtig sind. Anfänger, welche auf dem Weg zur Lösung nicht schnell sichtbare Fortschritte machen, geben oft auf und schreiben ab [BDH+08a].
- Da Konzepte nicht systematisch vermittelt werden, sind diese nur autodidaktisch zugänglich. Damit verkümmert der Unterricht zu einer reinen Polarisation zwischen den Autodidakten – die zudem oft nur unsystematisch programmieren lernen – und denjenigen Studierenden, die den Unterricht ohne Gewinn wieder verlassen. (Obwohl das Problem schon lange bekannt ist [Gri74], ist die heutige Ausbildung erstaunlich frei von Lösungen.)

3 Effektive Programmierausbildung für Anfänger

Die folgenden, aufeinander abgestimmten Bausteine, lösen gemeinsam die Probleme der traditionellen Ausbildung:

Explizite Anleitung Die *design recipes* in der Terminologie von Felleisen [FFFK01] bzw. *Konstruktionsanleitungen* bei Klaeren/Sperber [KS07] bieten *explizite* Anleitungen zur systematischen Programmieren. Abschnitt 4 beschreibt, wie die Konstruktionsanleitungen das systematische Problemlösen ermöglichen.

Programmiersprache Der Ansatz verwendet eine Lehrsprache basierend auf der funktionalen Sprache Scheme [SDFvS10], die es erlaubt, im direkten Anschluss an den

Mathematikunterricht alle grundlegenden Konzepte des Programmierens nachvollziehbar darzustellen. Vollständige Programme werden bereits in der ersten Unterrichtsstunde geschrieben. Die Sprachen sind sehr klein, womit Dozenten nur wenig Zeit der spezifischen Behandlung der Sprachen widmen müssen.

Programmierumgebung Die Programmierumgebung *DrRacket* (früher *DrScheme*) ist speziell auf Anfänger zugeschnitten [FCF+02]: Sie ist interaktiv und ermöglicht das schrittweise sofortige Testen und visualisiert den Programmablauf. DrRacket liefert präzise Fehlermeldungen, die Anfänger aktiv bei der Fehlersuche unterstützen [CS10]. In DrRacket ist aber auch die Entwicklung anspruchsvoller Programme mit grafischer Benutzeroberfläche bis hin zu professionellen Applikationen möglich.

Übungen Im Rahmen der traditionellen Übungsorganisation bieten wir eine umfangreiche, annotierte Aufgabensammlung. Zudem haben wir neue Ansätzen zur Betreuung („Betreutes Programmieren“) und innovative Konzepte für Gruppenarbeit („Gruppentestate“) [BDH+08a] entwickelt, die den Lernstoff besonders effektiv vermitteln. Ein Online-Kursinformationssystem erleichtert die organisatorische Umsetzung und unterstützt den Übungsbetrieb.

4 Systematische Problemlösung nach Konstruktionsanleitung

Die Konstruktionsanleitungen unterscheiden unseren Ansatz von traditioneller Programmierausbildung, die mit Beispielen funktionierender Programme oder „design patterns“ arbeiten. (Die mächtigen Abstraktionsmöglichkeiten unserer Lehrsprachen machen „design patterns“ weitgehend überflüssig.) Die Konstruktionsanleitungen trennen die systematischen von den kreativen Aspekten des Problemlösungsprozesses und erlauben damit Anfängern, sich auf den kreativen Teil zu konzentrieren. Im Folgenden stellen wir die Problemlösung mit Konstruktionsanleitung anhand einer einfachen Beispielaufgabe vor. Die Erarbeitung der Aufgabenlösung folgt den Schritten der Konstruktionsanleitungen:

> Ein Computer besteht vereinfacht gesehen aus Prozessor, Hauptspeicher und Festplatte. Die Taktfrequenz des Prozessors gibt man in Ghz, die Größe des Hauptspeichers und der Festplatte in GB an.
> Programmieren Sie Computer! Schreiben Sie ein Programm, in dem es möglich ist, den Gesamtspeicher eines Computers zu berechnen!

Kurzbeschreibung Die Kurzbeschreibung fasst die Aufgabe der Prozedur in einer einzelnen Zeile zusammen:

```
; Gesamtspeicher eines Computers berechnen
```

Datenanalyse Die Datenanalyse ist der zentrale Teil der Konstruktionsanleitung. Sie klassifiziert die Arten von Informationen, die in der Aufgabenstellung auftauchen: In der Aufgabenstellung taucht die Größe „Computer“ auf. Der Begriff „Computer“ kann anhand der Aufgabenstellung weiter ausgeführt werden:

Ein Computer hat einen Prozessor mit Taktfrequenz in Ghz, einen Hauptspeicher mit Größe in GB und eine Festplatte mit Größe in GB. Diese Formulierung ist eine *Datendefinition*. Eine Formulierung der Form „*x* hat“ identifiziert ein Objekt als *zusammengesetzte Daten*: Es besteht aus mehreren Bestandteilen. Solch eine Datendefinition wird als Kommentar formuliert und dann direkt in Code in Form einer Record-Definition übersetzt:

```
; Ein Computer besteht aus:
; - Prozessor (in Ghz)
; - Hauptspeicher (in GB)
; - Festplatte (in GB)
(define-record-procedures computer
  make-computer
  computer?
  (computer-processor computer-ram computer-drive))
(: make-computer (rational rational rational -> computer))
```

Diese Form definiert die Signatur `computer` für Computer sowie mehrere Prozeduren für den Umgang mit Record-Werten: den Konstruktor `make-computer`, um Computer herzustellen, das Prädikat `computer?`, um Computer von anderen Sorten von Werten zu unterscheiden und die Selektoren `computer-processor`, `computer-ram` und `computer-drive`, die für einen Computer jeweils den Prozessor, den Hauptspeicher und die Fesplatte liefern.

Schließlich wird noch die Signatur für die Konstruktorprozedur `make-computer` deklariert: `make-computer` akzeptiert drei Zahlen (Taktfrequenz des Prozessors und Größen von Hauptspeicher und Festplatte). Die Signatur `rational` für rationale Zahlen ist eingebaut.

Signatur Die Signatur ähnelt einer Typsignatur der Prozedur. Sie enthält den Namen der Prozedur, welche Arten von Daten die Prozedur erwartet und welche Art sie zurückgibt. Aus der Aufgabenstellung ist ersichtlich, dass die Prozedur, die den Gesamtspeicher berechnet, einen Computer konsumiert und eine Zahl zurückgibt:

```
(: total-memory (computer -> rational))
```

Testfälle Die Testfälle werden vor der Prozedur geschrieben und erlauben den Studierenden, erste Korrektheitskriterien aufzustellen. Der erste Schritt für die Testfälle ist die Erstellung von Beispiele für die auftretenden Daten. Diese Beispiele sollten durch Kommentare die Beziehung zwischen den Daten und der repräsentierten Information klarstellen. In diesem Fall könnten Beispiele so aussehen:

```
; Computer mit 1 Ghz, 2 GB RAM, 500 GB Festplatte
(define c1 (make-computer 1 2 500))
; Computer mit 2 Ghz, 4 GB RAM, 1000 GB Festplatte
(define c2 (make-computer 2 4 1000))
```

Mit diesen Beispielen lassen sich folgende Testfälle formulieren:

```
(check-expect (total-memory c1) 502)
(check-expect (total-memory c2) 1004)
```

Die Form `check-expect` akzeptiert zwei Operanden, deren Werte gleich sein sollen. Sind sie es nicht, werden Testfall-Fehlschläge in einem Protokoll festgehalten.

Gerüst Der erste Schritt zur Konstruktion der eigentlichen Prozedur ist das Gerüst, das sich direkt aus der Signatur ergibt: Die Studierenden müssen lediglich die Namen für die Parameter wählen.

```
(define total-memory
  (lambda (c)
    ...))
```

Das `define` benennt einen Wert mit dem Namen `total-memory` – der `lambda`-Ausdruck liefert eine Prozedur mit dem Parameter `c` (für den Computer). Die Ellipse steht für den noch zu schreibenden Rumpf der Prozedur.

Schablone Die Schablone leitet sich rein aus der Form der Daten her, ohne die Bedeutung oder Zweck der Prozedur in Betracht zu ziehen. Die Studierenden können an der Form der Daten die Programmelemente ablesen, die im späteren Rumpf der Prozedur auftauchen. In den meisten Fällen kommen mehrere Schablonen zum Einsatz:

1. Der Parameter `c` muss im Rumpf auftauchen:
   ```
   (define total-memory
     (lambda (c)
       ... c ...))
   ```
2. Da es sich bei `c` um zusammengesetzte Daten handelt, wird eine Prozedur die Bestandteile des Computers benötigen. Die Schablone wird also erweitert:
   ```
   (define total-memory
     (lambda (c)
       ... c ...
       ... (computer-processor c) ...
       ... (computer-ram c) ...
       ... (computer-drive c) ...))
   ```

Rumpf Erst in diesem Schritt ist problemspezifisches Wissen nötig, um die Lücken der Schablone im Rumpf der Prozedur zu füllen – bis hierher ist das Programm systematisch aus der Datenanalyse der Problemstellung entstanden. Zunächst spielt für die Berechnung des Gesamtspeichers eines Computers die Prozessor-Komponente keine Rolle. Dann geht es noch darum, die Verknüpfung der im Rumpf verbliebenen Größen zu finden, in diesem Fall die Addition. Damit ist die Konstruktion des Rumpfes abgeschlossen, der vollständig so aussieht:

```
(define total-memory
  (lambda (c)
    (+ (computer-ram c)
       (computer-drive c))))
```

Testen Der letzte Schritt ist das Testen: Die vorher erstellten Testfälle werden ausgeführt und Fehler werden behoben, die durch die Testfälle aufgedeckt wurden.

5 Aufruf zur Zusammenarbeit

Wir laden alle Interessenten ein, den Ansatz auszuprobieren und gemeinsam zu verbessern.

Literatur

[BDH+08a] Annette Bieniusa, Markus Degen, Phillip Heidegger, Peter Thiemann, Stefan Wehr, Martin Gasbichler, Marcus Crestani, Herbert Klaeren, Eric Knauel und Michael Sperber. Auf dem Weg zu einer robusten Programmierausbildung. In Andreas Schwill, Hrsg., *3. Workshop des GI-Fachbereichs Hochschuldidaktik der Informatik*, Potsdam, Germany, Dezember 2008.

[BDH+08b] Annette Bieniusa, Markus Degen, Phillip Heidegger, Peter Thiemann, Stefan Wehr, Martin Gasbichler, Marcus Crestani, Herbert Klaeren, Eric Knauel und Michael Sperber. HtDP and DMdA in the Battlefield. In Frank Huch und Adam Parkin, Hrsg., *Functional and Declarative Programming in Education*, Victoria, BC, Canada, September 2008.

[CS10] Marcus Crestani und Michael Sperber. Growing Programming Languages for Beginning Students. In Stephanie Weirich, Hrsg., *Proceedings International Conference on Functional Programming 2010*, Baltimore, Maryland, USA, September 2010. ACM Press, New York.

[FCF+02] Robert Bruce Findler, John Clements, Cormac Flanagan, Matthew Flatt, Shriram Krishnamurthi, Paul A. Steckler und Matthias Felleisen. DrScheme: A Programming Environment for Scheme. *Journal of Functional Programming*, Seiten 159–182, Marz 2002.

[FFFK01] Matthias Felleisen, Robert Bruce Findler, Matthew Flatt und Shriram Krishnamurthi. *How to Design Programs*. MIT Press, 2001.

[FFFK04] Matthias Felleisen, Robert Bruce Findler, Matthew Flatt und Shriram Krishnamurthi. The TeachScheme! Project: Computing and Programming for Every Student. *Computer Science Education*, Marz 2004.

[Gri74] David Gries. What should we teach in an introductory programming course? In *Proceedings of the Fourth SIGCSE Technical Simposium on Computer Science Education*, Seiten 81–89, Detroit, Michigan, USA, 1974.

[KS07] Herbert Klaeren und Michael Sperber. *Die Macht der Abstraktion*. Teubner Verlag, 1st. Auflage, 2007.

[Köl99a] Michael Kölling. The Problem of Teaching Object-Oriented Programming, Part 1: Languages. *Journal of Object-Oriented Programming*, 11(8):8–15, January 1999.

[Köl99b] Michael Kölling. The Problem of Teaching Object-Oriented Programming, Part 2: Environments. *Journal of Object-Oriented Programming*, 11(9):6–12, April 1999.

[SDFvS10] Michael Sperber, R. Kent Dybvig, Matthew Flatt und Anton van Straaten, Hrsg. *Revised[6] Report on the Algorithmic Language Scheme*. Cambridge University Press, 2010.

Semi-strukturierte Interviews im Software-Engineering: Indikationsstellung, Vorbereitung, Durchführung und Auswertung – Ein Fall-basiertes Tutorium

Christa Weßel

Dr. Christa Weßel MPH | Organisationsentwicklung & Informationstechnologie
Kaiserstraße 79
60329 Frankfurt am Main
mail@christa-wessel.de

Abstract: Zu den Aufgaben in IT-Projekten gehören das Kennenlernen der Anwenderbedürfnisse, die Diskussion von Analyse, Design und Implementierungsergebnissen mit dem Auftraggeber, sowie die Durchführung von Evaluationen. Die dabei eingesetzten Methoden des Software-Engineerings und der statistischen Analyse werden ideal ergänzt durch qualitative Forschungsmethoden aus den Sozialwissenschaften. Von diesen Methoden sind vor allem Beobachtungen und Einzel- und Gruppeninterviews für den Einsatz in IT-Projekten geeignet. Das Leitfaden-gestützte Interview ist eine in Seminaren gut erlernbare und handhabbare Methode, die eine zuverlässige, nachvollziehbare, klar strukturierte und zielorientierte Gewinnung von Informationen, insbesondere von neuen Erkenntnissen zu einer Problemstellung zulässt. Die Resultate können unmittelbar den Software-Engineering-Prozess einfließen. Die Teilnehmer können Erkenntnisse zu den Situationen und Bedürfnissen des Unternehmens, des Unternehmers und der Beschäftigten und ihrer Arbeit gewinnen, die in dieser Form anderen Ansätzen, beispielsweise der Datenerhebung mittels quantitativer Methoden (Fragebögen etc.), nicht zugänglich sind. Von besonderem Interesse ist dabei das Kennenlernen der Hintergründe und Ursachen. Hieraus lassen sich Lösungsansätze formulieren. Das Interview fördert die Einbeziehung der Interviewten, beispielsweise der Nutzer, in den Entwicklungsprozess. Diese Methode können die Teilnehmer auf andere Anwendungsbereiche übertragen. Zu diesen zählen beispielsweise die Vorbereitung, Durchführung und Auswertung von Besprechungen und Beobachtungen. – Schlüsselwörter: Software-Engineering, Projekt, Nutzerintegration, interdisziplinäre Kommunikation, qualitative Forschungsmethoden, Interview, Weiterbildung

1 Den Nutzer kennen lernen und einbeziehen

Für das Gelingen von IT-Projekten ist die frühzeitige Einbindung des Nutzers und eine gute Kommunikation mit dem Auftraggeber ein entscheidender Erfolgsfaktor. Auftragsstellung, Arbeitssituationen, Anwenderbedürfnisse und –ideen müssen zielgerichtet und umfassend zu Beginn und im Projektverlauf erhoben und überprüft werden. Dies trägt zur effizienten und effektiven Erstellung eines Produktes und Durchführung einer Dienstleistung und damit zur Qualitätssicherung bei [Wi96, Kl00, So06]. Auftraggeber und Nutzer sollen von Beginn an in den Entwicklungs- und Implementierungsprozess einbezogen werden. Dies findet zunehmend Berücksichtigung in Vorgehens- und Managementmodellen für das Software-Engineering, wie beispielsweise den Agilen Methoden [Be01, CLC04].

Die dabei eingesetzten Methoden des Software-Engineerings und der statistischen Analyse werden ideal ergänzt durch qualitative Forschungsmethoden (QFM) aus den Sozialwissenschaften [SC98]. Von diesen Methoden sind vor allem Beobachtungen und Einzel- und Gruppeninterviews für den Einsatz in IT-Projekten geeignet [AIM03, WS07]. Ziel ist es, methodengeleitet und für dritte nachvollziehbar Zusammenhänge zu verstehen und daraus Maßnahmen abzuleiten.

Informatiker und Domänen-Experten brauchen Anleitungen und ein geeignetes Werkzeugportfolio, um Beobachtungen und Einzel- und Gruppeninterviews angemessen und methodisch fundiert anzuwenden [ABN04].

2 Interviews als wissenschaftlich fundierte Explorationsmethode

Interviews gehören zu den qualitativen Forschungsmethoden und sind eine in den Sozialwissenschaften seit Jahrzehnten entwickelte und angewendete Methode [BD06, SC98, FW09]. Die Gewinnung neuer Erkenntnisse, die in dieser Form anderen Ansätzen, beispielsweise der Datenerhebung mittels quantitativer Methoden (Fragebögen etc.), nicht zugänglich sind, soll der Formulierung von Hypothesen und Lösungsansätzen dienen. Hypothesen sind mittels weiterer Untersuchungen zu belegen oder zu widerlegen. Qualitatives Forschen zielt ab auf Verstehen und Erklären und ist charakterisiert durch Offenheit für Neues, strukturiertes Vorgehen und Flexibilität. Letztere ist eine große Stärke in der qualitativen Forschung. Hier ist es erlaubt und gefordert, auf neue Erkenntnisse unmittelbar einzugehen. Dies erfolgt beispielsweise mit einer Anpassung des Erhebungsmethoden. Subjektivität wird zur Stärke der Methode, wenn Reflektion, Explikation, Methodenkontrolle in der Durchführung verankert sind. Dies gilt auch für die Anwendung von Interviews im Rahmen von IT-Projekten. Es handelt sich um einen iterativen Prozess, der durch den Austausch mit Kollegen und den Interviewten eine solide Datenerhebung und –Interpretation ermöglicht [BD06, SC98].

In der qualitativen Forschung kommen verdeckte und offene, neutrale und teilnehmende Beobachtungen, Einzel- und Gruppeninterviews, und die Analyse von Dokumenten zur Anwendung [Ma02]. Auf IT-Projekte angewendet entspricht dies Beobachtungen, Gesprächen und Aktenanalysen. Das Datenmaterial besteht aus Texten, die bei Beobachtungen durch Protokolle, bei Interviews durch Notizen und Transkripten von Aufnahmen und bei Dokumenten durch diese selbst entstehen.

Interviews können offen, semi-strukturiert oder strukturiert durchgeführt werden. Offene Interviews beginnen mit einer einleitenden Frage, zu der der Interviewte sich äußert ohne weitere größere Eingriffe des Interviewers. Sie finden vor allem im psychologischen Umfeld Anwendung. Strukturierte Interviews basieren auf einem Katalog von offenen und geschlossenen Fragen, deren Reihenfolge festgelegt ist. Damit entspricht diese Interviewform der Anwendung eines Fragebogens im Gespräch.

Semi-strukturierte Interviews werden unter Anwendung eines Leitfadens durchgeführt, und daher auch Leitfaden-gestützte Interviews genannt. Der Interviewte wird als Experte für sein Gebiet betrachtet. Daraus leitet sich das Synonym Experten-Interviews für semi-strukturierte Interviews ab. Diese ermöglichen zweierlei: (1) Auf der Basis eines Interview-Leitfadens kann der Interviewer Themen und Ablauf steuern und (2) der Interviewte kann frei sprechen. Dies gibt Freiraum für die Beschreibung von Problemen und Lösungsideen. Vor allem gibt es Freiraum für Aspekte, die der Interviewer so bislang nicht gesehen hat, also für neue Erkenntnisse.

In IT-Projekten dienen Interviews der Exploration der Arbeitsorganisation und -Prozesse, der Arbeitsinhalte sowie der Untersuchung von vorhandenen und möglichen Umsetzungen der Software-Entwicklung und -Implementierung.

3 Vorbereitung, Durchführung und Auswertung eines semi-strukturierten Interviews

Tabelle 1 zeigt die acht Schritte der Vorbereitung, Durchführung und Analyse des oder der Interviews. Diese werden – je nach Verlauf des IT-Projekts – teilweise mehrfach durchlaufen. Jeder dieser Schritte ist als ein Meilenstein zu sehen, an dem der Interviewer – allein und im Gespräch mit Kollegen – Fortschritte reflektiert und offene Fragen sowie eine eventuell erforderliche Änderung des Leitfadens diskutiert und beschließt. Identifikation der Interviewpartner, Vertraulichkeit und Nutzen für den Interviewten, Interviewleitfaden, Vorverständnis, Feldnotizen und Auswertung werden im Folgenden näher betrachtet.

Schritt	Teile
Identifikation der Interviewpartner	Organisationsanalyse und Auswahl
Terminvereinbarung	Hintergrund und Ziel des Interviews, Verabredung, Dauer, Ort, Vertraulichkeit, Nutzen für den Interviewten
Vorbereitung	Interviewleitfaden, Explikation des Vorverständnisses, Materialien (Aufnahmegerät, Uhr etc.)
Durchführung des Interviews	Begrüßung, Aufnahme, Abschied
Feldnotizen	Ablauf, zusätzliche Informationen (nach der Aufnahme), Eindrücke, neue Ideen und Erkenntnisse
Dokumentation	Transkription: Aufnahme durch eine Schreibkraft, Feldnotizen durch den Forschenden
Datenauswertung	Iterativ, computer-gestützt
Bericht	Übertragung der Auswertungsergebnisse in das IT-Projekt

Tabelle 1: Interview-Vorbereitung, -Durchführung, -Auswertung

3.1 Identifikation der Interviewpartner

Diese basiert auf einer Organisationsanalyse [Cu08]. Der zu Interviewende sollte mehrere Jahre Mitarbeiter der Organisation sein, die Organisation, ihre Arbeitsabläufe und die Kooperation mit externen Organisationen gut kennen. Dabei liegt der Schwerpunkt auf Berührungspunkten zu dem zu untersuchenden Informationssystem. Des Weiteren sollte der zu Interviewende bereit sein, Visionen und Wünsche zu entwickeln und die Zeit für ein oder auch mehrere Interviews aufzubringen.

3.2 Vertraulichkeit und Nutzen

Bei der Kontaktaufnahme und der Durchführung des Interviews muss der Forschende die Vertraulichkeit und den Nutzen für den Interviewten deutlich machen.

3.3 Interview-Leitfaden

Der Ablauf eines Interviews lässt sich gliedern in Aufwärmphase, Behandlung des Themas und Abschluss. Für leitfaden-gestützte Interviews in IT-Projekten bieten sich die Phasen Einleitung, Exploration der gegenwärtigen Situation, Exploration zukünftiger Situationen und Abschluss an. Tabelle 2 zeigt Aspekte, die in den einzelnen Phasen berücksichtigt werden sollten. Zu beachten ist, dass der Interviewer den Leitfaden entwickelt und die hier genannten Aspekte Beispiele sind, die als Basis herangezogen werden können.

Phase	**Teile**
Einleitung	Die Organisation (Visionen, Ziele, Strukturen, Geschäftsaktivitäten)
	Der Interviewte (beruflicher Hintergrund, Aufgaben, Arbeit)
	Begriffsklärungen
Exploration der gegenwärtigen Situation	Relevanz
	IT und andere Werkzeuge
	Arbeitskontext (Prozesse, Kommunikation, ...)
	Vorbedingungen, Einschränkungen, Probleme
Exploration zukünftiger Situationen	Erwartungen und Wünsche
	Wünsche zu neuen Werkzeugen und Anwendungen
	Weitere Ideen
Abschluss	Zusammenfassung durch den Interviewer
	Feedback zur Vollständigkeit durch den Interviewten
	Nutzen für den Interviewten
	Dank und Abschied

Tabelle 2: Interview-Phasen

3.4 Vorverständnis

Die schriftliche Formulierung des eigenen Vorverständnisses zum zu untersuchenden Feld, Problem und Forschungsziel schließt die Beschreibung des eigenen Vorwissens und eigener positiver wie negativer Vorurteile ein [Se06]. Hierdurch ist es möglich, bei der Auswertung neue Aspekte, die durch den Interviewten eingebracht werden, zu erkennen und Aspekte nicht zu vernachlässigen, die durch eigene Vorurteile verdeckt werden könnten.

3.5 Feldnotizen

Der Interviewer fertigt während und vor allem in unmittelbarem Anschluss an das Interview an einem ungestörten Platz kurze Notizen an. Diese Notizen nennen Sozialwissenschaftler „Feldnotizen“, da der Interviewer sie „im Feld“, also in der zu untersuchenden Umgebung erstellt. Diese Feldnotizen dokumentieren den Ablauf, zusätzliche Informationen (nach der Aufnahme), Eindrücke, neue Ideen und Erkenntnisse und fließen in die spätere Textanalyse und Interpretation ein [Ma02].

3.6 Auswertung

Das Datenmaterial, also der transkribierte Interviewtext und die Feldnotizen werden induktiv und deduktiv analysiert [Ma02]. Aus dem Text entwickelt der Auswertende induktiv Kategorien ("Codes") und Subkategorien, die Aspekte des Interviews mit Schlagwörtern beschreiben, bspw. "Layout" und "Farbe". In der deduktiven Analyse nutzt der Auswertende zuvor festgelegte Kategorien. Der Interview-Leitfaden bildet mit seinen Themen die Basis für diese Kategorien. Die Analyse schließt mit der Untersuchung zuvor noch nicht markierter Textabschnitte. Die den Codes zugeordneten Textabschnitte werden „Codings“ genannt.

Diese Analysen wiederholt der Auswertende mehrfach. Dadurch gewinnt die Auswertung an Schärfe und Klarheit. Die Analyse kann computergestützt erfolgen, bspw. mit MAXqda(TM) [Ve10]. Dies erleichtert die Dokumentation und Nachvollziehbarkeit durch Dritte. Die mögliche Visualisierung unterstützt die Untersuchung besonders häufig oder mit anderen Kategorien zusammen angesprochener Aspekte. Dies kann Hinweise auf gemeinsame Probleme oder Lösungsvorschläge aufdecken. Die Quantität eines angesprochenen Aspekts darf jedoch nicht als Äquivalent für ihre Relevanz gesehen werden: oft werden wichtige Punkte nur ein oder zweimal angesprochen. Hier ist die Reflektionsfähigkeit des Auswertenden gefordert.

3.7 Übertragung der Auswertungsergebnisse in das IT-Projekt

Die Texte der Interviews und Feldnotizen bieten Datenmaterial an, das nach der Analyse direkt und abstrahiert sowohl für Projektplanungen als auch für das Software-Engineering verwendet werden können (Abbildung 1).

In Projekten im Allgemeinen finden die Codings Eingang in Reports mit der Beschreibung von Ziel, Ressourcen, Risikofaktoren, Aufgaben und Meilensteinen und in abstrahierter Form in das Projektdesign. Im Software-Engineering können Codings in Szenarios und Persona Models einfließen, die wiederum die Basis für Use Cases bilden. Gemeinsam werden sie dann in der Anforderungsanalyse genutzt [So06]. Wichtig ist die Wahrung der Vertraulichkeit und des Datenschutzes in der direkten und abstrahierten Verwendung der Codings.

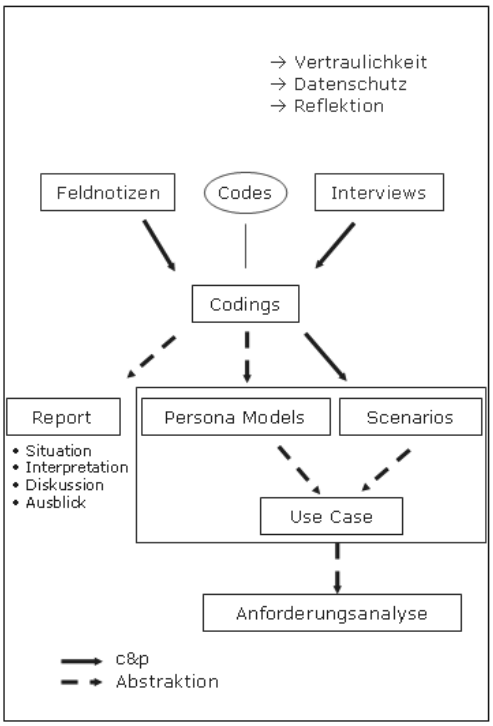


Abbildung 1: Interviewergebnisse – Übertragung ins Software-Engineering

4 Fall-basiertes Erlernen und Training

4.1 Universitäre und Hochschulausbildung

In der Regel erlernen Studierende oder Forschende qualitative Forschungsmethoden in ein- bis zweisemestrigen Seminaren oder im Zusammenhang mit Forschungsprojekten, zu denen auch eigene Abschlussarbeiten (Diplom, Bachelor, Master) oder Dissertationen gehören können. Die Curricula enthalten [WS07]:

- Einführung in qualitative Forschungsmethoden
- Abgrenzung zu und Kombination mit quantitativen Methoden (Triangulation)
- Charakteristika Leitfaden-gestützter Interviews
- Anwendungsgebiete
- Vorbereitung, Durchführung und Nachbereitung

- Erstellung eines Interviewleitfadens
- Bedeutung von Vorverständnis und kontinuierlicher Reflektion
- Dokumentation
- Auswertung: Datenanalyse und Interpretation
- Grundzüge der (computer-gestützten) Auswertung
- Übertragung der Ergebnisse auf das untersuchte Gebiet (Schlussfolgerungen, Empfehlungen, offene Fragen)
- Erstellung des Berichtes (Abschlussarbeit, Dissertation, Projektdokumentation, Projektbericht, Publikation)

4.2 Weiterbildung

Im Rahmen ihrer kontinuierlichen Weiterbildung können IT-Experten Leitfaden-gestützte Interviews in Seminaren erlernen. Der Inhalt orientiert sich an dem beschriebenen Curriculum. Wie dort kommen auch hier Methoden des Fall- und Projekt-basierten Lernens anhand eigener oder vorgeschlagener Beispiele zur Anwendung [OMN05, WS09].

Die Dauer solcher Seminare sollte bei zwei bis drei Tagen liegen. Seminare, die einige Stunden oder einen Tag umfassen, bieten den Teilnehmern die Möglichkeit, sich mit ersten Grundlagen vertraut zu machen, und vor allem die Erstellung eines Interviewleitfadens zu erlernen.

4.3 Bisherige Ergebnisse und Erfahrungen

Die Entwicklung und Durchführung von Weiterbildungen zur Anwendung semi-strukturierter Interviews im Software-Engineering, in Gesundheitswissenschaften und Medizin beruht auf dem Curriculum „Qualitative Forschungsmethoden in der Medizinischen Informatik“, an dem zwischen Oktober 2004 und August 2006 am Institut für Medizinische Informatik der RWTH Aachen fünfzehn Studierende, Doktoranden und Wissenschaftler aus der Informatik und der Medizin teilnahmen [WS07]. Die Teilnehmer verwendeten leitfadengestützte Einzel- und Gruppeninterviews für die Bereiche Exploration, Anforderungsanalyse und formative und summative Evaluationen für das Wissensmanagement in Klinik und Forschung, Krankenhausmanagement, Datenbanken, Web-basierte Informationsplattformen und ihre Applikationen und Mobile Computer in der Notfallmedizin [z.B. Ka05, WWS06, SSG07, Ra09].

Die Evaluation zeigte die große Zufriedenheit der Teilnehmer mit der Erlernbarkeit und Anwendbarkeit von qualitativen Forschungsmethoden (QFM) in ihren Forschungs- und Entwicklungsprojekten. Da die Ergebnisse unmittelbar in Berichte und Anforderungsanalysen einfließen konnten, schätzten die Teilnehmer den Aufwand für QFM als angemessen ein. Vor allem betonten sie den strukturierten und nachvollziehbaren Ansatz von QFM. Hierzu einige Zitate aus der Evaluation: „Durch leitfaden-gestützte Interviews ist es leicht für mich, einen Eindruck von und Informationen zum Problem insgesamt und mit vielen wichtigen Details in kurzer Zeit zu bekommen." – „Ich kann die Texte und Ergebnisse auf einer klaren Basis mit den anderen Studenten und meinen Betreuern diskutieren." – „Der erste Schritt der Anforderungsanalyse – den Anwender, seine Situation und seine Bedürfnisse kennen zu lernen – ist jetzt viel strukturierter." – „Die Perspektive des Informatikers wechselt von der IT hin zum Nutzer. Ist wirklich Nutzer-orientiert." [siehe auch WS07].

Die in den Folgejahren durchgeführten Seminare richteten sich vor allem an Experten aus den Gebieten Informatik und Software-Engineering, Gesundheitswissenschaften und Arbeitsmedizin[1].

Die Teilnehmer zeigten eine hohe Zufriedenheit mit der Strukturiertheit, Nachvollziehbarkeit und praktischen Anwendbarkeit der Methode. Wichtig und als schwierig umsetzbar eingeschätzt wurde es, das Vertrauen der Interviewkandidaten für eine Aufnahme zu gewinnen. Die hierzu in den Seminaren vorgestellten Vorgehensweisen schätzten die Teilnehmer als sehr hilfreich ein. Die für ein semi-strukturiertes Interview erforderlichen Vorbereitungen planten die Teilnehmer auch in anderen Zusammenhängen einzusetzen. Hierzu zählten die Entwicklung eines Gesprächsleitfadens im Sinne einer Agenda mit offenen Fragen und der Reflektion mittels der Formulierung des eigenen Vorverständnisses. Als mögliche Einsatzgebiete nannten die Teilnehmer Projektübergaben und andere Gespräche mit Kollegen und Kunden.

5 Fazit

Mittels Leitfaden-gestützter Interviews können IT-Experten Erkenntnisse zu den Situationen und Bedürfnissen des Unternehmens, des Unternehmers und der Beschäftigten und ihrer Arbeit gewinnen, die in dieser Form anderen Ansätzen, beispielsweise der Datenerhebung mittels quantitativer Methoden (Fragebögen etc.), nicht zugänglich sind [AIM03]. Von besonderem Interesse ist dabei das Kennenlernen der Hintergründe und Ursachen. Hieraus lassen sich Lösungsansätze formulieren. Die Resultate fließen unmittelbar in Projekte im Allgemeinen und in das Software-Engineering im engeren Sinn ein.

[1] Publikationen zu auf Fachkonferenzen durchgeführten Workshops, Tutorien und Seminaren siehe http://www.christa-wessel.de/index.php/Publikationen.html

Das Interview fördert die Einbeziehung der Interviewten in das IT-Projekt [FW09, Ul05]. Damit unterstützt es Ansätze, wie sie beispielsweise in der Philosophie der Agilen Methoden postuliert sind [Be01, CLC04]. Das Leitfaden-gestützte Interview ist eine gut erlernbare und handhabbare Methode, die eine zuverlässige, nachvollziehbare, klar strukturierte und zielorientierte Gewinnung von Informationen, insbesondere von neuen Erkenntnissen zu einer Problemstellung zulässt [BD06].

Adressaten von Seminaren zur Erlernung der Methode sind IT-Experten, die ihr Portfolio der Instrumente in Analyse, Design, Implementierung und Evaluation im Software-Engineering um ein Nutzer und Kunden orientiertes Instrument erweitern möchten. Sie erlernen am konkreten Fall Indikationsstellung, Vorbereitung, Durchführung und Auswertung von Leitfaden-gestützten Interviews und die Übertragung der Ergebnisse in den Software-Engineering-Prozess. Sie können diese Methode auf andere Anwendungsbereiche übertragen. Zu diesen zählen beispielsweise die Vorbereitung, Durchführung und Auswertung von Besprechungen, Reviews und Beobachtungen.

Die Lehrenden müssen mit qualitativen Methoden in Theorie und eigener Praxis vertraut sein, Methoden des Fall- und Projekt-basierten Lernens und Lehrens beherrschen und das Feld, in dem die Lernenden sich bewegen, kennen. Dies bedeutet, dass die Lehrenden sowohl eine sozialwissenschaftliche Ausbildung als auch Kenntnisse und Berufspraxis in IT-Projekten haben sollten [WS07].

Literaturverzeichnis

[ABN04] Ammenwerth E, Brender J, Nykänen P, Prokosch H-U, Rigby M, Talmon J. Visions and strategies to improve evaluation of health information system - reflections and lessons based an the HIS-EVAL workshop in Innsbruck. Int J Med Inf 2004; 73 (6): 479-91

[AIM03] Ammenwerth E, Iller C, Mansmann U. Can evaluation studies benefit from triangulation? A case study. Int J Med Inform 2003; 70: 237-248

[BD06] Bortz J, Döring N. Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler. 4., überarbeitete Auflage. Berlin, Springer 2006

[Be01] Beck K et al. Manifesto for Agile Software Development - The Mainfesto - Principles - History. http://agilemanifesto.org/ 2001. 01/07/2010

[CLC04] Cohen D, Lindvall M, Costa P. An Introduction to Agile Methods. Advances in Computers 2004; 62: 1-66

[Cu08] Cummings TG (Ed.) Handbook of Organization Development. Los Angeles, Sage Publications 2008

[Fi05] Fichter K. Modelle der Nutzerintegration in den Innovationsprozess. WerkstattBericht Nr. 75. Berlin, IZT Institut für Zukunftsstudien und Technologiebewertung 2005

[FW09] Friedman C, Wyatt J. Evaluation Methods in Biomedical Informatics. New York, Springer 2009

[Gl10] Gloger B. SCRUM. Der Paradigmenwechsel im Projekt- und Produktmanagement - Eine Einführung. Informatik Spektrum 2010; 33: 195-200

[Ka05] Karakas G. Automatisches Generieren von webbasierten und benutzungsfreundlichen Informationsseiten über Krankenhäuser aus einer objektorientierten Datenbank. Diplomarbeit. Aachen, RWTH Aachen, Fakultät 1 - Mathematik -Informatik - Naturwissenschaften 2005

[Kl00] Kling R. Learning About Information Technologies and Social Change: The Contribution of Social Informatics. The Information Society, 2000; 16: 217–232

[Ma02] Mayring P. Einführung in die qualitative Sozialforschung. Beltz Studium, Weinheim 2002

[OMN05] O'Neill G, Moore S, McMullin B (eds). Emerging Issues in the Practice of University Learning and Teaching. Dublin, AISHE 2005. - http://www.aishe.org/readings/2005-1/collection.pdf 01/07/2010

[Ra09] Rams J. Informationen ueber Krankenhaeuser im Internet - Evaluation des Web-Frontends einer Objektdatenbank. Dissertation. Aachen, RWTH Aachen, Medizinische Fakultät 2009

[Re04] Reinicke T. Möglichkeiten und Grenzen der Nutzerintegration in der Produktentwicklung. Eine Systematik zur Anpassung von Methoden zur Nutzerintegration. Dissertation. Berlin, Technische Universität Berlin, Fakultät V – Verkehrs- und Maschinensysteme 2004

[SC98] Strauss A, Corbin JM. Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory. Thousands Oaks, CA: Sage Publications 1998

[Se06] Senge PM. The Fifth Discipline. The art and practice of the learning organization. 2nd edition. London, Doubleday 2006

[So06] Sommerville I. Software Engineering. 8th edition. Boston, Addison Wesley 2006

[SSG07] Soboll M, Spreckelsen C, Geisler S, Spitzer K, Weßel C. Exploration and Requirement Analysis for Mobile Data Communication for Emergency Medical Services - A Case Study. In: Leimeister JM, Mauro C, Krcmar H, Eymann T, Koop A (Hrsg.). Mobiles Computing in der Medizin. Proceedings zum 7. Workshop der GMDS Arbeitsgruppe Mobiles Computing in der Medizin, Augsburg, 20. September 2007. Aachen, Shaker 2007: 85 - 93

[Ul05] Ulich E. Arbeitspsychologie. 6., überarbeitete und erweiterte Auflage. Stuttgart, Schäffer-Poeschel 2005

[Ve10] Verbi. MAXqda - The Art of Text Analysis. http://www.maxqda.de/ 01/07/2010

[Wi96] Wiegers KE. Creating a Software Engineering Culture. New York (NY), Dorset House Publishing Company 1996

[WS07] Weßel C, Spreckelsen C. Qualitative Forschungsmethoden auf den Punkt gebracht – Ein Trainingsprogramm für leitfaden-gestützte Interviews in der Medizinischen Informatik. PrInterNet - Zeitschrift für Pflegewissenschaft 2007 (9), 9: 542-548

[WS09] Weßel C, Spreckelsen C. Continued Multidisciplinary Project-Based Learning – Implementation in Health Informatics. Methods Inf Med. 2009; 48 (6): 558-563

[WWS06] Weymann F, Weßel C, Spreckelsen C. Temporal extension of an object-oriented information system about hospitals. In: Engelbrecht R, Hasman A (Ed.). European Notes in Medical Informatics. Ubiquity: Technologies for Better Health in Aging Societies. Proceedings of MIE2006. CD-ROM. ISSN 1861-3179. Munich, EFMI: ENMI European Notes in Medical Informatics Vol II No 2, 2006: MIE 2006 pp 203-208

Software-Engineering Culture. Vom Umgehen miteinander und mit externen Partnern in IT-Projekten – Seminare in Aus- und Weiterbildung

Christa Weßel

Dr. Christa Weßel MPH | Organisationsentwicklung & Informationstechnologie
Kaiserstraße 79
60329 Frankfurt am Main
mail@christa-wessel.de

Abstract: Ziel eines Projekts ist die effiziente und effektive Erstellung eines Produkts oder Durchführung einer Dienstleistung. Eine gute Projektkultur kann wesentlich zum Projekterfolg beitragen. Das Wissen um zwischenmenschliche Interaktionen, Grundlagen der Organisations- und Projektkultur, Methoden der Teambildung und der Personalentwicklung und die Einbettung dieser Grundlagen und Methoden in Qualitäts- und Projektmanagement bilden dafür eine wichtige Grundlage. Dieses Wissen gilt es Studierenden der Informatik und IT-Experten in der kontinuierlichen Weiterbildung zu vermitteln. Das Projekt-basierte Lernen ist eine seit Jahrzehnten etablierte Methode zur nachhaltigen Vermittlung von fachlichen, methodischen und sozialen Kompetenzen. Seminare aus der universitären und Hochschulausbildung bilden die Grundlage für die Entwicklung von mehrtägigen Seminaren und Workshops in der Weiterbildung von Experten. - Schlüsselwörter: Software-Engineering, Projekt, Projektkultur, soziale Interaktion, Projekt-basiertes Lernen, Didaktik, Weiterbildung.

1 Warum Kultur?

Der Erfolg von IT-Projekten hängt von der Zusammenarbeit der aus unterschiedlichen Disziplinen stammenden Beteiligten und einem qualitätsgesicherten Vorgehen ab [Wi96, DL99, Kl00, So06]. Welche Kultur des Umgehens miteinander und mit externen Partnern sich in einem IT-Projekt entwickelt, hängt entscheidend von den Fähigkeiten, Kenntnissen und Zielen der Beteiligten ab.

Ziel des Aufbaus und der Pflege einer guten Projektkultur ist die effiziente und effektive Durchführung des Projekts mit der termingerechten und den Anforderungen entsprechenden Erstellung eines Produkts [Ke06]. Gleichzeitig fördert eine gute Projektkultur die Motivation und das Commitment der Teammitglieder und der externen Partner, und trägt damit zur Pflege der wichtigsten Ressource in Projekten bei: den Mitarbeitern [Ul05, Cu08].

Die Charakteristika einer guten Projektkultur und die erforderlichen Kenntnisse und Fähigkeiten zum Aufbau und zur Pflege können und müssen IT-Experten erlernen, die ihre Projekte erfolgreich planen, aufbauen, durchführen und abschließen wollen [Wi96, DL99]. Darum findet das Thema Software-Engineering Culture zunehmend Eingang in die universitäre und Hochschulausbildung und in die Weiterbildung von Informatikern aller Teildisziplinen (Informatik, Wirtschaftsinformatik, Medieninformatik, ...) [GI05].

Im Folgenden verwende ich für die Kultur eines IT-Projekts den von Wiegers geprägten Begriff „Software-Engineering Culture“ (SWEC) [Wi96].

2 Rahmenbedingungen einer guten Projektkultur

Projekte sind alltägliches Geschehen in der Arbeitswelt. Gemeinsame Werte, Ziele und Regeln bilden die Basis für die Zusammenarbeit und die erfolgreiche Durchführung [Wi96, DL99, Ke06]. Eine gute Projektkultur zeichnet sich durch Transparenz, Zuverlässigkeit, Termintreue, Zusammenarbeit, Commitment und Motivation der Teammitglieder aus. Dies zu realisieren ist davon abhängig, welche Kultur die Organisation pflegt, in die ein Projektteam eingebettet ist.

Abbildung 1 skizziert die Rahmenbedingungen guter Projektkultur. Im Zentrum stehen die Ansatzfelder zur Verwirklichung. Im weiteren Umfeld sind Aspekte aufgeführt, die Bezug und Einfluss auf ein oder mehrere Ansatzfelder haben. Auf die hier dargestellten Ansatzfelder gehe ich im Folgenden näher ein.



Abbildung 1: Projektkultur – Rahmen

2.1 Kultur

Kultur (lateinisch cultura (f) „Bebauung“, „Pflege“ (des Körpers und Geistes) und „Ausbildung“) ist nach [Me01] „die Gesamtheit der typischen Lebensformen größerer Gruppen einschließlich ihrer geistigen Aktivitäten, besonders der Werteinstellung. ...“ Hieraus wird deutlich, dass der Umgang der Menschen in einer Gruppe entscheidend von Werteinstellungen und Regeln abhängig ist. Je höher der Konsens der Beteiligten dazu ist, desto wahrscheinlicher ist es, dass die Mitglieder einer Gruppe einen für alle befriedigenden Umgang miteinander pflegen können. Auf die Arbeitswelt bezogen bedeutet dies: die gemeinsame erfolgreiche Arbeit wird durch eine gute Kultur gefördert.

2.2 Die Organisation

Grundcharakteristika einer guten Organisationskultur sind die Entwicklung von Unternehmenszielen, Respekt vor den Mitarbeitern, den Kunden und der Umwelt, eine solide strategische und taktische Geschäftspolitik [Cu08]. Dies schlägt sich nieder in einer transparenten und fairen Kommunikation, guter Pflege der Ressourcen und der Beziehungen zu Kunden und Partnern. Die European Foundation for Quality Management hat dies und die Messung der Ergebnisse in ihrem EFQM Excellence Model beschrieben [EFQM09].

2.3 Personalentwicklung

Mitarbeiter sind die entscheidende Ressource für den Erfolg eines Unternehmens. Personalentwicklung ist eine Hauptaufgabe im Personalmanagement (Human Resources Management – HRM) und muss von Personalabteilungen und Führungskräften durchgeführt werden [Hu97, Ul05, Se06, Cu08]. Personalentwicklung beinhaltet die Rekrutierung fachlich und sozial kompetenter Menschen und die Förderung einer kontinuierlichen Fortbildung der Mitarbeiter in fachlicher und methodischer Hinsicht. Auch soziale Kompetenzen können und sollten im Verlauf eines Berufslebens trainiert werden. Dies kann vor allem „on the job“ durch eine entsprechende Arbeitskultur erfolgen. Hinsichtlich der Motivation sind neben Anerkennung und Feedback auch Punkte wie Gehalt, Aufstiegsmöglichkeiten und Bonussysteme zu berücksichtigen. Bei den Bonussystemen kommen auch nicht-finanzielle Boni in Frage. Dies können beispielsweise Freistellungen für besondere Fortbildungen oder Sabbatmonate sein.

Das Software Engineering Institute (SEI) will mit dem People Capability Maturity Model (P-CMM®) Organisationen in ihrer Personalentwicklung unterstützen [CHM09]. Damit bildet das P-CMM® eine Ergänzung zu den Modulen CMMI-DEV (Entwicklung), CMMI-ACQ (Beschaffung), CMMI-SVC (Dienstleistungen) des Capability Maturity Model® Integration [SEI10]. Die Personalentwicklung mit Rekrutierung, Aus- und Weiterbildung, Förderung und Forderung soll in den Organisationsalltag integriert werden. In den beim CMMI üblichen fünf Levels schlagen Curtis et al. folgende Schritte zur Umsetzung vor [CHM09]:

- Level 1: Bewusstsein schaffen

- Level 2 und 3: Prozesse implementieren
- Level 4: Strukturen, Prozesse, Ergebnisse messen
- Level 5: Fortlaufende Verbesserung

Die Konzipierung und Durchführung der Personalentwicklung ist Aufgabe der Organisationsleitung und der Führungskräfte.

2.4 Führungskompetenz – Leadership

Führungskräfte leiten Menschen in einer Organisation, Abteilung, Gruppe oder einem Team [Wi96, DL99, Ke06, Le07, Cu08]. Im Weiteren verwende ich für diese Gruppierungen den Begriff Team. Aufgabe der Führungskraft ist, diese Menschen dazu zu befähigen, effizient und effektiv zu arbeiten. Dies bedeutet: sie braucht die entsprechenden Fähigkeiten und Vollmachten.

Die Führungskraft ist verantwortlich für Aufgabenbeschreibungen, Teamzusammensetzung, kurz-, mittel- und langfristige Planungen, Gestaltung der internen Kommunikation (technisch und sozial) und Controlling. Die Kommunikation zur Organisationsleitung, zu anderen Teams und Organisationsabteilungen und zu externen Partnern und Kunden zählen ebenso zu den Aufgaben der Führungskraft. In all diesen Bereichen sowie für Zwischen- und Abschlussberichte ist sie für die Dokumentation verantwortlich. Je nach der Größe ihres Teams muss und wird eine Führungskraft Teile dieser Aufgaben delegieren.

Eine Führungskraft muss fachliche Kompetenz besitzen: sie muss wissen, wo Risiken und Chancen auftreten, damit sie auf diese frühzeitig reagieren kann. Wichtiger als fachliche Kompetenz werden von vielen Autoren, z.B. von [Wi96, DL99, Ul05 Ke06, Le07, Cu08], soziale und Methodenkompetenz eingeschätzt. Methodenkompetenz ermöglicht der Führungskraft, Management-Instrumente, beispielsweise zur Planung, zum Controlling und im Risikomanagement, effizient und effektiv einzusetzen.

Soziale Kompetenz ist die Basis für eine gute Mitarbeiterführung. Eine Führungskraft muss gut mit Menschen kommunizieren können, auf sie eingehen, ihnen auch Grenzen aufzeigen und sie unterstützen und loben können. Diese Fähigkeiten kann sie durch Ausbildungen in Moderation, Konfliktlösung und Mitarbeitergesprächen ausbauen. Sie muss mit Druck von innen (interne zwischenmenschliche und technische Probleme) und von außen (Termine, Kommunikation mit der Organisationsleitung, mit anderen Abteilungen und dem Kunden) umgehen können. Die Führungskraft muss ihr Team zusammenhalten und gegensteuern, wenn sich das Team oder Teile des Teams in einer Sackgasse verrennen. Bei aller Flexibilität darf sie das Ziel, die erfolgreiche Arbeit, nicht aus den Augen verlieren. Die Arbeit muss im Fluss bleiben. Ein wichtiger Baustein der Arbeit einer Führungskraft ist also der Aufbau und die Pflege von Teams.

2.5 Teamentwicklung

Die multidisziplinäre Zusammensetzung von Projektteams ist mittlerweile Routine [So06, Ke06, Le07, Cu08]. So arbeiten in IT-Projekten nicht nur Vertreter der Subdisziplinen der Informationstechnologie zusammen, sonders es sind auch, je nach Unternehmensphilosophie oder Domäne, für die ein Produkt entwickelt wird, Ökonomen, Psychologen, Soziologen, Biologen, Mediziner oder andere im Team vertreten. Hinzu kommt, dass die Mitarbeiter einen akademischen oder auch einen nicht-akademischen Hintergrund haben und oft aus verschiedenen Kulturen und Ländern kommen.

Diese Diversität ist eine große Chance, stellt aber auch an den Projektleiter und an jedes Teammitglied Anforderungen. Neben fachlicher Expertise und Methodensicherheit kommt Fähigkeiten, die auch in vermeintlich homogenen Teams erforderlich sind, hier besondere Bedeutung zu [Ke06][1]. Zu diesen Fähigkeiten zählen beispielsweise zuverlässige Mitarbeit, Respekt, Aufgeschlossenheit, Reflexionsfähigkeit, Kritikfähigkeit, Bereitschaft zur kontinuierlichen Weiterbildung und persönlichen und beruflichen Weiterentwicklung. Teamentwicklung und –Pflege können im Verlauf eines IT-Projekts an Meilensteinen unter Nutzung eines Vorgehensmodells erfolgen, beispielsweise dem Rational Unified Process (RUP) und dem Spiral Modell [So06]. Besonders geeignet sind Agile Methoden, da eine gute Projektkultur Teil ihrer Philosophie ist [Be01, CLC04, Gl10]. Zu Meilensteinen zählen Projektstarts und Anforderungsanalysen, Team Reviews und Projektabschlüsse. Die gemeinsame Reflektion des Teams erfolgt dabei zu den Fortschritten in der Produkterstellung und in der Art der Zusammenarbeit. Dies sollte und kann als Teil der Qualitätssicherung gesehen und durchgeführt werden [Wi96].

2.6 Arbeitsumwelt

Die Arbeitsbedingungen müssen soweit als möglich auf die Bedürfnisse der Menschen zugeschnitten sein [Ul05, Cu08]. Aspekte sind hier: Arbeitszeiten, Arbeitsorte (beispielsweise die Möglichkeit, zuhause zu arbeiten), Ausstattung des Arbeitsplatzes, Architektur und Zustand des Gebäudes.

Wichtiger Bestandteil jeglicher Arbeit, auch in IT-Projekten, ist eine auf die Bedürfnisse der Aufgaben zugeschnittene Technologie. Zum Projektstart müssen die erforderliche Hard- und Software und Informationstechnologien definiert und vorhanden sein. Im Verlauf muss der Projektleiter den Bedarf an zusätzlichen Instrumenten und ihre zeitnahe Bereitstellung sicherstellen.

[1] In scheinbar homogenen Teams kommen die Mitglieder nur aus einer Berufsgruppe und aus einem Land.

2.7 Arbeitsorganisation

Die Erstellung eines Produkts oder die Durchführung einer Dienstleistung, sei es nun im Routinebetrieb oder in Projekten, folgt einem mehr oder weniger definierten Prozess. Mehr ist an dieser Stelle oft wirklich mehr: je klarer Arbeitsprozesse beschrieben sind, desto geringer wird die Wahrscheinlichkeit von Fehlern. In einem so kreativen Bereich wie dem Software-Engineering gilt es hier, die Balance zwischen hilfreicher Unterstützung durch Vorgehensmodelle und Einengung durch eine übersteigerte Bürokratie zu halten. Daher empfehlen zahlreiche Autoren, anerkannte Vorgehensmodelle und Standards auf die Anwendbarkeit im eigenen IT-Projekt zu prüfen [bspw. Wi96, So06]. Nach der Auswahl eines geeigneten Modells sollte das Projektteam die eigene Arbeitsprozesse beschreiben, beispielsweise als Software Development Procedures [Wi96] oder als Software Development Guidelines [ISW07] Das Team muss die Prozessbeschreibungen regelmäßig auf ihre Angemessenheit überprüfen und fortschreiben. Die Qualität der Kommunikation mit Auftraggeber, Nutzern, Abteilungen der eigenen oder anderer Organisationen ist ein wichtiger Faktor für den Projekterfolg. Im Folgenden wird dargestellt, wie dies in Bezug auf den Kunden gestaltet werden kann.

2.8 Arbeitsinhalte

Der Kunde muss nicht von Beginn des Projekts an einbezogen werden. Er muss schon *vor* dem Projekt einbezogen werden. Im Projektverlauf und zum Abschluss gilt es, die Balance zu halten zwischen guter Zusammenarbeit und zu großer Nähe. So warnt Wiegers ausdrücklich davor, den Kunden zu sehr in den Entwicklungsprozess zu integrieren. Aufgabe des Kunden ist es, Unklarheiten bei der Anforderungsanalyse auszuräumen, Feedback zu den Zwischenergebnissen zu geben und sich an der Evaluation zu beteiligen (Tests, User Feedback) [Wi96]. Projektdefinition und Projektauftrag können nur in Zusammenarbeit mit dem Kunden erfolgen. Teamzusammenstellung und Projektplanung sind eine interne Angelegenheit. Während der Durchführung des IT-Projekts ist der Kunde an der Anforderungsanalyse, dem Design und der Implementierung zu beteiligen. Hierzu bieten sich Team Reviews an [Wi02, WCG08]. Beim Abschluss eines Projekts empfiehlt es sich, das Produkt nicht nur vom Kunden abnehmen zu lassen, sondern auch mit ihm den Projektverlauf zu reflektieren – zusätzlich zur Reflektion im internen Abschlussgespräch („Debriefing“).

Gute Projektkultur ist Teil der Philosophie Agiler Methoden. Der Kunde steht im Mittelpunkt. Ziel ist die Erstellung eines arbeitsfähigen Produkts zum Nutzen des Kunden am Ende eines jeden – möglichst kurzen – Entwicklungszyklus. Kreativität, Motivation, Eigenständigkeit, Verantwortung und kontinuierliche Weiterbildung der Entwickler sind entscheidende Faktoren für den Erfolg des Projekts. Die Reviews, die in der agilen Methode SCRUM als Sprint Planning, daily Stand-up Meeting und Post Sprint Meeting durchgeführt werden, zielen darauf ab, die Bedürfnisse des Kunden und der Mitarbeiter zu (er)kennen und auf sie zeitnah zu reagieren [Be01, CLC04, Gl10].

3 SWEC in Aus- und Weiterbildung

Für die Ausbildung im Hochschulumfeld zum Thema Software-Engineering Culture (SWEC) bietet das Projekt-basierte Lernen (PBL) Studierenden die Möglichkeit sich das Gebiet an konkreten Fällen und Projekten nachhaltig zu erarbeiten. Ziel dieses Vorgehens ist es, Aufmerksamkeit und Offenheit für dieses nicht-technische Thema bei den Studierenden zu wecken, Inhalte zu vermitteln und den Studierenden die selbständige Erschließung dieses Gebietes zu ermöglich, um dadurch eine Basis für das lebenslange Lernen zu diesem Thema zu schaffen [RFWS00, OMM05].

3.1 Fall- und Projekt-basiertes Lernen

Die Erarbeitung von Lernzielen im Rahmen realistischer Szenarien führt zu einer besseren Motivation von Lernenden (Schüler, Studierende, Experten in der Weiterbildung) und zeigt im Vergleich zu herkömmlichen Methoden bessere Ergebnisse in kurz-, mittel- und langfristigen Lernergebnissen [FWSR00, OMM05]. Dies motivierte die Entwicklung und Einführung des Fall-, Problem- und Projekt-basierten Lernens in Medizin, Jura, Betriebswirtschaft, Informatik und vielen anderen Gebieten [PBJJ02, Bl04].

Das Lernen im Projekt fußt auf der Theorie des kognitiven Lernens: Das Lernen durch Einsicht und am Modell befähigt die Lernenden, ein Feld eigenständig zu erschließen, Lösungen zu entwickeln und wissenschaftliche Methoden anzuwenden. Außerdem stärken sie ihre soziale Kompetenz und bauen ihre Kenntnisse in Teamarbeit und Projektmanagement auf und aus [Sl96, FWSR00, Bl04, PBJJ02]. Der Lehrende fungiert dabei als „facilitator, mentor, guide“ [WS09].

Projekt-basiertes Lernen im Hochschulumfeld findet im Rahmen von Seminaren von ein bis zwei Semester Dauer statt [PBJJ02, Bl04]. Die Studierenden identifizieren zu Beginn des Semesters ein Thema zu einer Arbeit, die sie als Projekt in Kleingruppen durchführen werden. Alternativ kann der Lehrende auch Themen vorschlagen.

Ein weiteres Anwendungsgebiet ist die Integration von Studierenden in Forschungsprojekte, in deren Rahmen sie ihre Abschluss- oder Doktorarbeiten durchführen [De07, WS09].

Das Projekt-basierte Lernen wird hervorragend durch das Blended Learning unterstützt [PBJJ02, HP06, HWG10, We10]. Es setzt sich zusammen aus Präsenzzeit, Nutzen einer Webplattform zur Hinterlegung von Lernmaterialien und Arbeiten der Studierenden und zum online Feedback, sowie dem Selbststudium der Studierenden allein oder in Kleingruppen. In der Präsenzzeit stecken der Lehrende und die Studierenden die Rahmenbedingung ab. Wie in jeder Gruppe, die zum Team werden soll, gehört hierzu auch die Vereinbarung von Werten, Zielen und Regeln. Des Weiteren finden hier Impulsvorträge zu verschiedenen Themen statt, stellen die Studierenden den Stand ihrer Arbeit und auch den Abschluss vor, diskutieren Studierende und Lehrender die Arbeiten und offene Fragen und geben Hilfestellungen und Anregungen.

Die gemeinsame Reflektion zum Seminar, zur Zielsetzung, Zufriedenheit der Studierenden und des Lehrenden mit dem Verlauf und mögliche Anpassungen im Seminar bilden die formative Evaluation. Diese führt der Lehrende als Leitfaden-gestützte Kurzinterviews und Diskussionen an Meilensteinen des Seminarverlaufs durch. Meilensteine sind der Start des Seminars (Erwartungen, Befürchtungen, Wünsche), nach einem, zwei und drei Vierteln der Lehrveranstaltung und am Schluss. Hier sollte auch eine quantitative Evaluation mittels Fragebogen durchgeführt werden. Diese ist an vielen Hochschulen als zentrale Befragung etabliert [bspw. EP10].

3.2 Umsetzung in der universitären und Hochschulausbildung

Die im Abschnitt (2) beschriebenen Themen wurden in mehreren Seminaren für Studierende und Doktoranden der Informatik und der Medizin behandelt. Der Ursprung liegt in der Forschungsgruppe „Informationssysteme im Gesundheitswesen“ am Institut für Medizinische Informatik der RWTH Aachen. Diese multi-disziplinäre Forschungsgruppe bearbeitete Themen zu Informationssystemen und mobilen Endgeräten im Gesundheitswesen. Das auf fünf Jahre angelegte Hauptprojekt hatte die Entwicklung eines web-basierten Informationssystems über Krankenhäuser zum Gegenstand [ISW07]. In den Jahren 2002 bis 2007 führten wir folgende Seminare durch:

- „Methoden der projektbasierten, interdisziplinären Forschung und Entwicklung in der Medizinischen Informatik“, RWTH Aachen 2002-2007 [WS09],

- „Wissensmanagement“, Seminar im Schwerpunktfach Medizinische Informatik, RWTH Aachen 2004/2005,

- „Qualitative Forschungsmethoden in der Medizinischen Informatik“, RWTH Aachen 2004-2006 [WWS06].

In die weitere Entwicklung von Seminaren zum Thema SWEC flossen Erfahrungen aus Seminaren zu Moderation und Projektmanagement ein [BCH07]. Das Seminar „Human Factors in Informatics“ (BHT Berlin 2009/2010) schließlich war ein einsemestriges Seminar, in dem Studierende der Medieninformatik sich das Thema SWEC erarbeiteten [We10]. Obwohl der Begriff „Human Factors in Informatics“ vor allem das Gebiet der Software-Ergonomie umfasst [SR08] [Fußnote: vgl. bspw. TU Berlin - Studiengang Human Factors (M.Sc.) http://www.humanfactors.tu-berlin.de/], behandelte diese Lehrveranstaltung die sozialen Aspekte von IT-Projekten: Die Studierenden sollten grundlegende Prinzipien und Gesetzmäßigkeiten des Menschen als Ersteller und Anwender von Informationstechnik kennen lernen. Problembereiche und Lösungsansätze in Bezug auf Konflikte, die im Umfeld eines Informatikers auftreten können, sollten beherrscht werden, wobei die Studierenden auch mit Diversity (Vielfalt: bspw. Gender, Nationalität, Beruf, Alter) umgehen können.

3.3 Ergebnisse: Was hatten die Studierenden davon?

Die Lehrveranstaltungen wurden formativ und summativ mittels Fokusgruppen-Interviews evaluiert [zur Methode qualitativer Evaluationsmethoden siehe BD06]. Zu Beginn eines Semesters oder eines Blockes erhoben wir die Ausgangssituation, Erwartungen, Befürchtungen und Wünsche der Studierenden. Im Verlauf der Lehrveranstaltung (LV) untersuchten wir im Abstand von vier bis sechs Wochen die Zufriedenheit mit der LV, offene Fragen, Unklarheiten und Verbesserungsvorschläge. Wir konnten nach den Interviews unmittelbar auf die Fragen und Anregungen der Studierenden mit der anschließenden Diskussion reagieren und die LV kontinuierlich weiter entwickeln. Die abschließende summative Evaluation fand zum einen wiederum als Gruppeninterview [WWS06, WS09], und zum anderen im Fall der einsemestrigen Veranstaltung [We10] außerdem mittels eines Fragebogens statt.

In allen Seminaren zeigten sich die Studierenden und Doktoranden mit der Methode des Fall- und Projekt-basierten Lernens sehr zufrieden. Sie fühlten sich gut betreut und durch die Lehrenden und die Arbeit in der Gruppe in der Weiterentwicklung ihrer fachlichen, methodischen und sozialen Kompetenz gefördert [WWS06, WS09, We10]. In den Jahren 2002 bis 2007 traten 32 Studierende und 4 Auszubildende in die Forschungsgruppe „Informationssysteme im Gesundheitswesen“ ein [WS09]. 25 Studierende schlossen ihre Arbeiten erfolgreich ab: 16 Studienarbeiten in Medizinischer Informatik, 4 Diplomarbeiten in der Informatik und 3 (plus 2 mittlerweile ebenfalls abgeschlossenen) Dissertationen in der Medizin plus 4 Trainee-Projekte der Auszubildenden. Die drei weiteren hier vorgestellten Seminare schlossen alle Studierenden erfolgreich ab [WWS06, We10].

3.4 Übertragung in Weiterbildungsformate

In der Weiterbildung von IT-Experten in Seminaren zur SWEC müssen Impulsvorträge zwischen praktischen Übungen genügend Raum einnehmen, um einen gemeinsamen Stand des Wissens der Teilnehmer zu erreichen. Fall-basierte Übungen, die sich auf einzelne Aspekte der SWEC beziehen, können in Kleingruppen und im Forum durchgeführt werden, und ein Äquivalent zum PBL bilden [OMN05, WS09]. Die Dauer solcher Seminare sollte bei zwei bis drei Tagen liegen. Seminare, die einige Stunden oder einen Tag umfassen, bieten den Teilnehmern die Möglichkeit, sich mit ersten Grundlagen vertraut zu machen.

4 Fazit

Die Ausbildung von IT-Experten in den sozialen Grundlagen des Software-Engineering ist eine wichtige Grundlage für die erfolgreiche und zufrieden stellende Arbeit in IT-Projekten [Wi96, DL99]. Das Fall- und Projekt-basierte Lernen unter Verwendung des Blended Learning ist ein sehr gut geeigneter Weg, die Lernenden für das Gebiet zu interessieren und ihnen Grundlagen und Methoden zu vermitteln [FWSR00, OMM05, HP06, WS09].

Software-Engineering Culture muss auch Gegenstand der kontinuierlichen Weiterbildung sein. So können IT-Experten beispielsweise in zwei bis dreitägigen Seminaren ihr Wissen vertiefen, Erfahrungen und Ideen mit Kollegen reflektieren und mittels des Fall-basierten Lernens das Erlernte erproben und die praktische Anwendung zu üben. Der Ausbau solcher Seminare zu Weiterbildungen in mehreren Blöcken unter Nutzung des Blended Learning ist dann ein weiterer Schritt [PBJJ02, HP06, HWG10, We10].

Die Teilnehmerinnen und Teilnehmer können damit unter anderem erlernen: Auswahl und Anwendung von Maßnahmen an Meilensteinen (z.B. zur Teambildung beim Projektstart, zu Kommunikation mit Auftraggeber und Nutzern), Früherkennung von Risiken (z.B. Personalausfälle aus unterschiedlichen Gründen oder projektexterne Hindernisse innerhalb und außerhalb der eigenen Organisation) und Präventions- und Interventionsmöglichkeiten. Sie können dies auf eigene Arbeiten anwenden. Damit können sie zu einer Qualitätsverbesserung von Strukturen, Prozessen und Produkten in ihren Projekten und in ihrer Organisation beitragen.

Die Lehrenden in diesen Seminaren müssen mit Methoden des Fall- und Projekt-basierten Lernens und Lehrens vertraut sein und über Kenntnisse und Berufspraxis in Organisationsentwicklung und IT-Projekten verfügen [WS09].

Literaturverzeichnis

[BCH07] Baumann M, Claßen-Linke I, Herrler A, Ohnesorge-Radtke U, Spreckelsen C, Weßel C. Curriculum "Neue Medien, Kommunikation, Didaktik in der Medizin". Aachener Schriften zur Medizinischen Informatik 2007; 02. Aachen: RWTH Aachen, Institut für Medizinische Informatik 2007

[BD06] Bortz J, Döring N. Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler. 4., überarbeitete Auflage. Berlin, Springer 2006

[Be01] Beck K et al. Manifesto for Agile Software Development - The Mainfesto - Principles - History. http://agilemanifesto.org/ 2001 - 26/06/2010

[Bl04] Bleimann U. Atlantis University – A New Pedagogical Approach beyond E-Learning. In: Furmell S, Dowland P (Ed.) INC 2004 Conference Proceedings. Plymouth, INC 2004: 553-560

[CHM09] Curtis B, Hefley WE, Miller SA. People CMM: A Framework for Human Capital Management. 2nd Edition. Reading, MA, Addison-Wesley 2009

[CLC04] Cohen D, Lindvall M, Costa P. An Introduction to Agile Methods. Advances in Computers 2004; 62: 1-66

[Cu08] Cummings TG (Ed.) Handbook of Organization Development. Los Angeles, Sage Publications 2008

[De07] Demiris G. Interdisciplinary innovations in biomedical and health informatics graduate education. Methods Inf Med. 2007; 46: 63-66

[DL99] DeMarco T, Lister T. Peopleware: Productive Projects and Teams. 2nd edition. New York, Dorset House Publishing Company 1999

[EFQM09] European Fondation for Quality Management. EFQM Excellence Model 2010 version. Brussels 2009

[EP10] Electric Paper. EvaSys. Education Survey Automation Suite. Homepage. http://www.electricpaper.de/produkte/evasys-education.html - 26/06/2010

[FWSR00] Felder RM, Woods DR, Stice JE, Rugarcia A. The Future of Engineering Education. II. Teaching Methods That Work. Chem Engr Education 2000; 34(1): 26-39

[GI05] GI e.V. Bachelor- und Masterprogramme im Studienfach Informatik an Hochschulen (Dezember 2005). Empfehlung der Gesellschaft für Informatik e.V. Neuauflage der GI-Standards zur Akkreditierung von Informatik-Studiengängen aus dem Jahr 2000. Bonn, GI e.V. 2005 http://www.gi-ev.de/fileadmin/redaktion/empfehlungen/GI-Empfehlung_BaMa2005.pdf - 26/06/2010

[Gl10] Gloger B. SCRUM. Der Paradigmenwechsel im Projekt- und Produktmanagement - Eine Einführung. Informatik Spektrum 2010; 33: 195-200

[HP06] Heinze A, Procter C. Online Communication and Information Technology Education. Journal of Information Technology Education 2006; 5: 235-249

[Hu97] Humphrey WS. Managing Technical People. Reading, MA, Addison-Wesley 1997

[HWG10] Holden JT, Westphall PJL, Gamor KI. An Instructional Media Selection Guide for Distance Learning - Implications for Blended Learning. Implications for Virtual Worlds. Boston MA, United States Distance Learning Association USDLA 2010 http://www.usdla.org/USDLA_Ins_Media.pdf - 30/06/2010

[ISW07] Ißler I, Spreckelsen C, Weßel C. Implementing Software Development Guidelines in a Medical Informatics Research Project. Methods of Information in Medicine 2007; 46: 641-645

[Ke06] Kerzner H. Project Management: A Systems Approach to Planning, Scheduling, and Controlling. 10th edition. Hoboken NJ, Wiley 2009

[Kl00] Kling R. Learning About Information Technologies and Social Change: The Contribution of Social Informatics. The Information Society, 2000; 16: 217–232

[Le07] Levi D. Group Dynamics for Teams. Thousand Oaks, CA, Sage Publications 2007.

[Me01] Meyers Lexikonredaktion (Hg.). Meyers großes Teaschenlexikon in 25 Bänden.Mannheim, Bibliographisches Institut 2001

[OMN05] O'Neill G, Moore S, McMullin B (eds). Emerging Issues in the Practice of University Learning and Teaching. Dublin, AISHE 2005. - http://www.aishe.org/readings/2005-1/collection.pdf - 26/06/2010

[PBJJ02] Pape B, Bleek WG, Jackewitz I, Janneck M. Requirements for Project-Based Learning – CommSy as an Exemplary Approach. In: Sprague RH, editor. Proceedings of the 35th Annual Hawaii International Conference on System Sciences. Los Alamitos; 2002

[RFWS00] Rugarcia A, Felder RM, Woods DR, Stice JE. The Future of Engineering Education. I. A Vision for a New Century. Chem Engr Education 2000; 34(1): 16-25

[Se06] Senge PM. The Fifth Discipline. The art and practice of the learning organization. 2nd edition. London, Doubleday 2006

[SEI10] Software Engineering Institute | Carnegie Mellon. Homepage. http://www.sei.cmu.edu/ - 26/06/2010

[Sl96] Slavin RE. Research on Cooperative Learning and Achievement: What We Know, What We Need to Know. Contemporary Educational Psychology 1996; 21: 43-69

[So06] Sommerville I. Software Engineering. 8th edition. Boston, Addison Wesley 2006

[SR08] Shackel B, Richardson SJ. Human Factors for Informatics Usability. Cambridge University Press 1991 (Hardcover); 2008 (Paperback)

[Ul05] Ulich E. Arbeitspsychologie. 6., überarbeitete und erweiterte Auflage. Stuttgart, Schäffer-Poeschel 2005

[WCG08] Weßel C, Christoph U, Geisler S. Efficient Quality Assurance in Process Models – An Example: The Combination of Team Reviews and Pass Arounds. In: Höhn R, Petrasch R, Linssen O (Ed.). Vorgehensmodelle und der Product Life-cycle - Projekt und Betrieb von IT-Lösungen. 15. Workshop der Fachgruppe WI-VM der Gesellschaft für Informatik e.V. (GI). Shaker, Aachen 2008: 176 - 189

[We10] Weßel C. Software Engineering Culture in Aus- und Weiterbildung – Ein Projekt-basiertes Lehr- und Trainingskonzept. In: Linssen O et al (Hg.) Integration von Vorgehensmodellen und Projektmanagement. 17. Workshop der Fachgruppe WI-VM der Gesellschaft für Informatik e.V. Aachen, Shaker 2010

[Wi96] Wiegers KE. Creating a Software Engineering Culture. New York (NY), Dorset House Publishing Company 1996

[Wi02] Wiegers K. Peer Reviews in Software. A Practical Guide. Boston, Addison Wesley 2002

[WS09] Weßel C, Spreckelsen C. Continued Multidisciplinary Project-Based Learning – Implementation in Health Informatics. Methods Inf Med. 2009; 48 (6): 558-563

[WWS06] Weßel C, Weymann F, Spreckelsen C. Streamlining Qualitative Research Methods for Medical Informatics - A Methodological Approach. In: Engelbrecht R, Hasman A (Ed.). European Notes in Medical Informatics. Ubiquity: Technologies for Better Health in Aging Societies. Proceedings of MIE2006. CD-ROM. ISSN 1861-3179. Munich, EFMI: ENMI European Notes in Medical Informatics Vol II No 2, 2006: MIE 2006 pp 371-376

Model-to-Code-Transformation mit der Model to Text Transformation Language

Roland Petrasch

Beuth Hochschule für Technik, Fachbereich VI
Luxemburger Str. 10, 13353 Berlin
petrasch@beuth-hochschule.de

Abstract: Die Model-to-Text Transformation Language ist eine Sprache der Object Management Group zur Transformation von formalen Modellen zu Text bzw. Code. Dieser Beitrag stellt beispielhaft die Code-Generierung auf der Basis eines UML-Klassenmodells vor, wobei einige der Sprach-Features in Verbindung mit OCL zum Einsatz kommen. Auch einige Hinweise zur praktischen Anwendung seien gegeben.

1 Model-to-Text-Transformation im Rahmen der MDA

Mit einer Model-to-Text Transformation Language lassen sich Modelle auf der Basis von Templates zu textuellen Artefakten transformieren, wobei der häufigste Fall wohl die Code-Generierung darstellt. Gemäß der MDA muss für das Ausgangsmodell ein MOF-konformes [MOF06] Metamodell vorliegen, z.B. die UML Superstructure [UML10], während dies für die Zielsprache bzw. die Zielarchitektur nicht der Fall sein muss. Die Metamodellelemente des Ausgangsmodells finden im Rahmen der Transformations-Templates im Sinne von Platzhaltern Verwendung, wobei die Zielsprachenkonstrukte direkt in die Templates „programmiert“ werden.

Während für die Model-to-Model-Transformation (M2M) Metamodelle für das Ausgangsmodell (Platform Independant Model, PIM) und das Zielmodell (Platform Specific Model, PSM) vorliegen, ist nur ein Ausgangsmetamodell für die Model-to-Text-Transformation (M2T) notwendig. Das PSM aus der M2M-Transformation wird wieder als PIM für die M2T-Transformation bezeichnet [PM06].

Die Object Management Group veröffentlichte 2008 die Version 1.0 der „MOF Model to Text Transformation Language (MOFM2T)“ [M2T08]. Insofern stehen mit der OVT [QVT08] und der MOFM2T nun für beide Transformationsarten (M2M, M2T) entsprechende Standards der MDA zur Verfügung. Im Folgenden sei aus praktischer Sicht eine kurze Einführung in die Code-Generierung mit der MOFM2T gegeben.

2 Model-to-Text-Transformation: Ein Beispiel

Für das Beispiel wird ein mit Topcased [TOP10] erstelltes UML-Modell zu Java-Code transformiert. Dafür wird die MOFM2T-Implementierung von Acceleo in der Version 3.0 genutzt [ACC10]. Die Basis ist Eclipse 3.6 (Helios) in Form der Modeling-Distribution [ECL10]. Es soll eine einfache Java-Applikation generiert werden, die im sog. Applikationsprojekt untergebracht wird. Im Generatorprojekt befinden sich hingegen die Templates zur M2T-Transformation, d.h. zur Code-Generierung.

Das Klassenmodell (PIM) für das Beispiel enthält vier Klassen (s. Bild 1): Die Klasse `Shoppping-Cart` besteht aus Artikeln, wobei die Aggregation zur Klasse `Article` mit einer Assoziationsklasse `ShoppingItem` für die Warenkorbpositionen ausgestattet ist. Weiterhin gehört der Warenkorb zu genau einem Kunden (Klasse `Customer`).

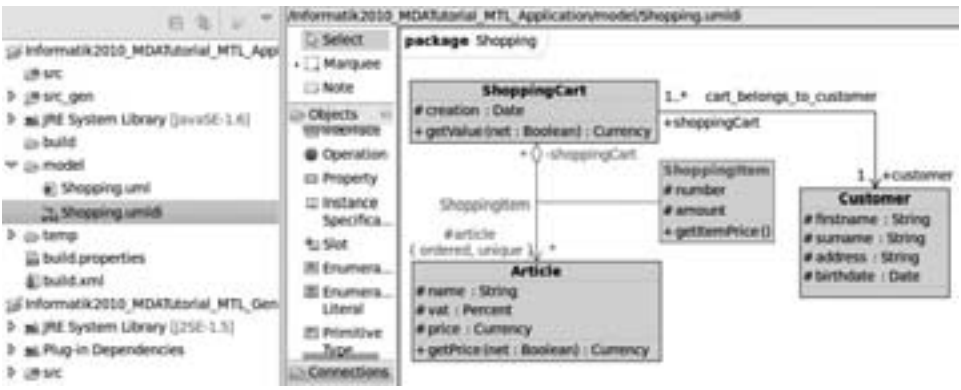


Bild 1: Applikationsprojekt mit UML-Klassenmodell als PIM

Im Generatorprojekt befinden sich die Acceleo Moduldateien mit den Templates. Bild 2 zeigt einen Teil des Templates für die Generierung von Java-Klassen, welches als Eingangsmodellelement eine UML-Klasse (Metaklasse `Class`) erwartet.



Bild 2: Generatorprojekt mit göffneter Template-Datei für Java-Klassen-Generierung

Ein Modul kann mehrere Templates enthalten. Tamplates in anderen Moduldateien lassen sich über Imports einbinden. Kommentare werden mit der comment-Direktive eingefügt, z.B. `[comment @main/]`, wobei die `@main`-Direktive den Einstiegspunkt beim Starten der Transformation markiert. Mit `[file …]` lässt sich eine Datei generieren, deren Inhalt durch nachfolgende Kommandos entsteht. Zunächst wird mit den folgenden Anweisungen die Klassendeklaration erzeugt:

```
[class.visibility/] class [class.name/] {
```

Gut erkennbar ist hier das Platzhalterprinzip des Templates: Überall dort, wo später bei der Code-Generierung Werte von Modellelementen des PIM eingesetzt werden sollen, stehen im Template die entsprechenden Metamodellelemente bzw. Ausdrücke, die diese Elemente verwenden, z.B. `[class.name/]`.Für die Behandlung der Properties sei zunächst auf das UML-Metamodell verwiesen [UML10, S. 48]: Dort sind die Metaklassen `Class` und `Property` durch eine Kompositionsbeziehung verbunden (s. Bild 3).



Bild 3: Ausschnitt aus dem Metamodell der UML 2.3: Metaklassen Class und Property (in Anlehnung an [UML10]

Im Template lassen sich die Attribute nun in Form einer Schleife behandeln. Das folgende Konstrukt zeigt den Schleifenkopf mit der Schleifenvariablen `prop` vom Typ `Property` (Metaklasse), die die Attribute (`ownedAttribute`) der Klasse `class` annimmt.

```
[for (prop:Property | class.ownedAttribute)]
```

Während einzelne Ausdrücke wie z.B. `[class.name/]` in eckigen Klammern stehen und durch das Zeichen „/“ terminiert werden, bilden Konstrukte wie die Schleife einen Block, der mit einer Anweisung abgeschlossen werden muss, z.B. `[/for]`. Template-Aufrufe wie z.B. `[generateUMLProperty2Java(prop)/]`, können Parameter enthalten (`prop`). Die Definition des aufgerufenen Templates sieht wie folgt aus:

```
[template public generateUMLProperty2Java(prop : Property)]
  [if (not prop.visibility.oclIsUndefined())]
   [prop.visibility.toString()/]
  [else]
   protected [comment default visibility/]
```

```
  [/if]
  [getAsJavaType(prop.type)/] [prop.name/];
[/template]
```

Die If-Anweisung behandelt den Fall, dass die Visibility des Properties undefiniert ist: In diesem Fall wird einfach `protected` eingesetzt, was den Generator etwas toleranter in Bezug auf unvollständige Eingangsmodelle werden lässt. Die Bedingung mit der Funktion `oclIsUndefined()` zeigt die Verwendbarkeit von OCL [OCL10] im Rahmen der Templates.

Eine Query hilft dabei, wiederkehrende Aufgaben auszulagern. Im Folgenden seien zwei Queries mit identischem Namen vorgestellt, die für einen Klassennamen einen entsprechenden Java-Dateinamen liefern:

```
[query public getJavaFileName(string : String) : String =
  string.concat('.java')/]


[query public getJavaFileName(class : Class) : String =
  getJavaFileName(class.name)/]
```

Die Syntax einer Query ist leicht zu verstehen: Die Signatur enthält die typischen Elemente: Sichtbarkeit, Bezeichner, Parameterliste und Rückgabetyp. Der Body wird durch das Zeichen „=“ eingeleitet. Das „`/]`“ markiert das Ende des Query-Definitionsblockes. Die beiden Query-Signaturen unterscheiden sich nur durch den Parameter. Der Aufrufer profitiert vom polymorphen Verhalten: Durch den Parametertyp wird automatisch die richtige Implementierung ermittelt:

```
[file (getJavaFileName(class), false)]
[file (getJavaFileName('Customer'), false)]
```

Während beim ersten Aufruf `getJavaFileName (class : Class)` Verwendung findet, führt die zweite Zeile zum Aufruf von `getJavaFileName (string : String)`. Eine Variable lässt sich durch die Anweisung `[let …]` definieren. Im folgenden Beispiel ist dies die Sichtbarkeit einer Operation, wobei der Wert der Variablen `vis` davon abhängt, ob im Modell die Visibility überhaupt definiert wurde. Ist dies nicht der Fall, wird „`public`“ vereinbart.

```
[template public generateUMLOperation2JavaMethod(op :
Operation)]
[let vis : String = if (op.visibility.oclIsUndefined())
      then 'public' else op.visibility.toString() endif]
  [vis/] [getAsJavaType(op.type)/] [op.name/]
           ([ParameterList(op,',')/]) {
[/let]
// [protected ('operation body implementation')]
  // ToDo: implement operation
```

```
// [/protected]
}
[/template]
```

Das o.g. Beispiel zeigt auch die Definition einer protected Region, die durch einen Block mit `[protected …]` und `[/protected]` begrenzt wird. Im Code erscheint dann der entsprechende Hinweis und die Aufforderung, die Operation zu implementieren, als Kommentar.

Assoziationen zwischen Klassen führen dazu, dass bei der Umsetzung in Java in einer oder beiden beteiligten Klassen entsprechende Member entstehen, die vom Typ der jeweils anderen Klassen sind. Im Folgenden sei ein Ausschnitt aus dem Template `generateAssociation` gezeigt, das genau diese Aufgabe erfüllen soll: Es soll für eine Assoziation einer Klasse eine MemberVariable und eine Getter- sowie Setter-Methode generieren.

```
[template public generateAssociation(class : Class, assoc :
Association)]
[let assocEnd : Property = assoc.ownedEnd->select
    (type<>class)->asSequence()->first()]
  [assocEnd.type.name/] [assocEnd.name/];
  [GenerateGetterSetter(assocEnd, assocEnd.visibility)/]
[/let]
[/template]
```

Der Aufruf dieses Templates erfolgt dann im Rahmen einer Schleife, die über alle Assoziationen einer Klasse läuft. Dies könnte wie folgt aussehen:

```
[for (assoc:Association | class.getAssociations())]
  [generateAssociation(class, assoc)/]
[/for]
```

Um die Zuweisung für die Variable `assocEnd` zu verstehen ist erneut ein Blick auf das UML-Metamodell zu empfehlen: Die Assoziation ist über eine Meta-Assoziation zu einer oder mehreren Properties verbunden (`navigableOwnedEnd`) (s. Bild 4).

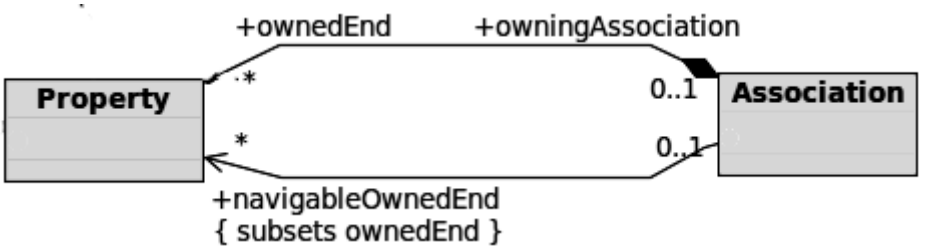


Bild 4: Ausschnitt aus dem Metamodell der UML 2.3: Metaklassen Property und Association (in Anlehnung an [UML10]

Kurz sei noch auf die Verwendung von OCL eingegangen. Wenn zu einer Klasse nur die konkreten Operationen ermittelt werden sollen, kann dies über eine Select-Funktion mit einer entsprechenden Bedingung erfolgen. Die Schleife für die Ausgabe der Namen aller konkreten Operationen einer Klasse `class` sieht wie folgt aus:

```
[for (op : Operation | class.ownedOperation->
      select(isAbstract=false))]
  [op.name/]
[/for]
```

3 Zusammenfassung

Die MOF Model-to-Text Transformation Language ist in Verbindung mit OCL ein mächtiges Werkzeug (nicht nur) zur Code-Generierung. Sie ist vom Ansatz her imperativ, typsicher und objektorientiert, daher unterstützt sie Techniken wie Polymorphie. Eine gute Erlernbarkeit ist auch durch die Syntax gegeben: Der überschaubare Befehlssatz umfasst die notwendigen Konstrukte und ist weitestgehend intuitiv verständlich.

Allerdings sollten die Templates im Sinne des Software-Engineerings mit den üblichen Qualitätsansprüchen konzipiert werden, um einen unwartbaren Wildwuchs zu verhindern. So gilt beispielsweise das Programming-by-Contract, d.h. bei jedem Template sollte klar sein, wie es zu verwenden ist, z.B. durch die Angabe der Pre-/Post-Conditions sowie der Invarianten. Auch sollte man auch hier die Entwicklungsprinzipien des S.O.L.I.D. Nicht vergessen, z.B. sollte ein Template oder eine Query im Sinne des Single Responsibility Principles nur genau für eine einzige Aufgabe verantwortlich sein.

Weiterhin besteht bei einigen Zielarchitekturen bzw. Zielsprachen die Gefahr, dass die Templates vom Umfang her unhandlich werden. Dem kann partiell durch die Beachtung der bereits genannten Techniken des Software-Engineerings entgegnet werden. Aber ab einer bestimmten Komplexität der Transformationen und Abstraktion der Ausgangsmodelle ist eine vorherige M2M-Transformation kaum zu vermeiden. Besonders wenn Designaspekte zu berücksichtigen sind, die manuelle Eingriffe erfordern. Derartige Entwurfsentscheidungen lassen sich praktisch kaum auf der Code-Ebene fällen – zumal im generierten Code bereits viele architekturorientierte Festlegungen enthalten sind. Frühzeitig sollte daher über M2M-Transformationen nachgedacht werden.

Die Diskussion um DSL (Domain Sepcific Languages) würde den Rahmen sprengen, es sei jedoch festgestellt, dass die UML als GPL (General Purpose Language) trotz ihrer Komplexität auf der Metamodellebene für viele Anwendungsgebiete geeignet erscheint. Was die Model-to-Text-Transformation mit MOFM2T angeht, ist die Verwendung einer eigenen MOF-konformen DSL jedoch genauso gut möglich.

Die hier verwendetet Acceleo-Implementierung in der Version 3.0 ist ausgereift und kommt als Eclipse-Plugin mit speziellem Editor mit Code Completion, Debug-Fähigkeit und Traceability daher. Für den Praxiseinsatz ist diese in Verbindung mit den anderen Eclipse-Projekten (EMF, UML etc.) gut geeignet. Der professionellen Anwendung von

MDA im Bereich Model-to-Text steht also nichts im Wege – vorausgesetzt das Personal ist ausreichend qualifiziert und eine Einführungsstrategie sowie ein passender Entwicklungsprozess für die modellgetriebene Software-Entwicklung sind vorhanden und akzeptiert.

Literaturverzeichnis

[ACC10] Obeo: Acceleo-Website, http://www.acceleo.org/pages/planet-acceleo/, Juli 2010

[ECL10] Eclipse Foundation: Eclipse Modeling Distribution (Helios Release), http://www.eclipse.org/downloads/packages/eclipse-modeling-tools-includes-incubating-components/heliosr, Juli 2010

[MDA03] Object Management Group: MDA Guide, Version 1.0.1, 2003

[MOF06] Object Management Group: MOF / Meta Object Facility Core, Version 2.0, 2006

[M2T08] Object Management Group: MOF Model to Text Transformation Language, Version 1.0, 2008

[OCL10] Object Management Group: Object Constraint Language, Version 2.2, 2010

[PM06] Petrasch R, Meimberg O: Model Driven Architecture. dpunkt Verlag, 2006

[QVT08] Object Management Group: MOF QVT - Version 1.0, 2008

[TOP10] Topcased Project, http://www.topcased.org/, Juli 2010

[UML10] Object Management Group: UML Superstructure and Infrastructure, Version 2.3, 2010

M2M-Transformation mit der QVT Relations Language

Siegfried Nolte

Beethovenstraße 57
22941 Bargteheide
siegfried@siegfried-nolte.de

Abstract: QVT ist ein Sprachkonzept der Object Management Group zur Transformation von formalen Modellen. In diesem Beitrag werden ausgehend von einem einfachen Fachklassenmodell die zentralen Konzepte und Techniken der Transformationssprache Relations Language vorgestellt, so dass abschließend eine vollständige Transformation eines UML-Modells der PIM-Ebene in eins der PSM-Ebene vorgenommen werden kann.

1 Einordnung der QVT in die MDA

Die *Model Driven Architecture* (MDA) der Object Management Group (OMG) beschreibt einen Ansatz der Anwendungsentwicklung auf der Grundlage von modellbasierten Techniken [MDA03]. Auf der einen Seite sind dies Sprachen und Verfahren zur Modellierung von Sachverhalten einer betrachteten Domäne, auf der anderen Seite handelt es sich um Konzepte zur Transformation der Modelle bis hin zur Erzeugung von Software-Bausteinen. Die Modellierung eines betrachteten Ausschnittes der realen Welt, der Domäne, erfolgt unter Verwendung von formalen, metamodellbasierten Modellierungssprachen, zum Beispiel der Business *Process Modeling Notation* (BPMN) oder der *Unified Modeling Language* (UML).

Ein Anwendungsentwicklungsprozess im Allgemeinen wie auch Modellierung im Besonderen ist iterativ inkrementell, der Entwicklungsprozesses findet über mehrere Abstraktionsebenen statt, wobei die Modelle immer weiter formalisiert und konkretisiert werden bis hin zu einer maschineninterpretierbaren Form, zum Beispiel einem Code-Artefakt in einer beliebigen Programmiersprache. Rechnergestützte Transformation von Modellen aus einer Entwicklungsebene in eine andere – aus der rechnerunabhängigen Modellierungsebene (CIM) über die plattformunabhängige Modellierungsschicht (PIM) und die plattformspezifische Schicht (PSM) zum Implementierungsmodell (IM) – kann dabei von großem Wert sein [PM06].

Der Vorschlag der OMG zur Transformation von Modellen ist die QVT, die *Query View Transformation*, eine Spezifikation von Sprachen, mit denen Modelle einer formalen Modellierungssprache in Modelle anderer formaler Modellierungssprachen überführt werden können [QVT08]. Mit Veröffentlichung der Version 1.0 im April 2008 ist die QVT ein von der OMG herausgegebener Standard. Die QVT bietet zwei Klassen von Sprachen zur Modelltransformation an, zum einen deskriptive Sprachen, von denen als

Beispiel hier die *Relations Language* (QVT-R) näher beleuchtet werden soll, und zum anderen imperative Sprachen, ein Beispiel dafür ist die *Operational Mappings* (QVT-O).

2 Modelle und Metamodelle

QVT-Sprachen arbeiten auf Modellen, welche mit Hilfe von formalen Modellierungssprachen, sogenannten MOF-Sprachen, entwickelt worden sind [Nol07]. Formale Modellierungssprachen sind Sprachen, denen eine wohldefinierte Syntax zugrunde liegt. Nach der MOF-Spezifikation der OMG [MOF06] wird die Definition von Modellierungssprachen in ihrem syntaktischen Teil unter Verwendung der UML – konkret, UML-Klassendiagramme – vorgenommen. Die formale Spezifikation einer Modellierungssprache in einem Modell wird als Metamodellierung bezeichnet, das Modell der Modellierungssprache ist ein Metamodell. Alle OMG-Modellierungssprachen, so auch die BPMN, UML und SysML, sind formale Modellierungssprachen in diesem Sinne. Das Ergebnis einer Transformation auf dieser Grundlage ist jeweils wieder ein formales Modell, das mit einer formalen Modellierungssprache dargestellt und im Rahmen des inkrementellen Entwicklungsprozesses weiter bearbeitet werden kann.

3 Transformation mit QVT-R

An einem einfachen UML-Modell werden einige grundlegende Konzepte und Transformationstechniken mit der deskriptiven QVT-Sprache QVT-R demonstriert [Nol09]. Mit der Wahl von UML als Modellierungssprache ist das Metamodell festgelegt. Die Fachklassen, die im Rahmen der Transformation behandelt werden sollen, sind unter Verwendung eines speziellen UML-Profiles mit dem Stereotyp `<<entity>>` markiert. Die Möglichkeit der Einarbeitung von domänenspezifischen Belangen in die Modellierungssprache UML ist eine sehr interessante Technik, die das UML-Metamodell mit UML-Profilen anbietet [UML10]. Als Zielmodell soll ebenfalls ein UML-Klassendigramm erstellt werden, welches bereits einige Designaspekte wie Paketierungen, Schnittstellen und Muster aufzeigt. Als Entwicklungsplattform wird ein aus freien Komponenten bestehendes Produkt eingesetzt [Nol10], welches auf der Basis der Eclipse Modeling Workbench entwickelt worden ist [EMP10]. Zur Unterstützung der Transformation soll mediniQVT [Med10] dienen.

3.1 Transformationen, Relationen und Domänen

QVT-R-Transformationen haben einen Namen und eine Liste von Parametern, in der die Modellkandidaten angegeben werden. Die Modelle sind Instanzen von Modelltypen, diese referenzieren Metamodelle, die im Kontext bekannt sein müssen [Nol07]. In dem Eclipse-basierten mediniQVT müssen die Metamodelle entweder als ein Eclipse-Plugin vorliegen oder alternativ in Form des universellen Austauschformates XMI [XMI07]. Die erforderlichen Metamodelle werden im Rahmen der Eclipse-Präferenzen „QVT Metamodels“ spezifiziert.

Das folgende Beispiel überführt ein UML-Quellmodell in ein UML-Zielmodell, wobei alle Packages eines Quellmodells mit ihrem Namen in Packages eines Zielmodells transformiert werden. Der Kern einer QVT-R-Transformation ist die Relation. Eine Relation ist die Spezifikation einer gültigen Beziehung zwischen einem Element eines Quellmodells und einem Element eines Zielmodells. In dem Beispiel werden alle `Packages` des Quellmodells `source` gemäß den vorgegebenen, in diesem Fall sehr einfachen, Bedingungen geprüft und diejenigen, die die definierten Bedingungen erfüllen, werden als Packages im Zielmodell `target` angelegt:

```
transformation GI1_Transformation (source:uml, target:uml)
{
   top relation Packages
   {
      pckgName : String;
      checkonly domain source srcPkg : Package
      { name = pckgName };
      enforce   domain target dstPkg : Package
      { name = pckgName };
   }
}
```

Die Prüfungen der Bedingungen und die Zuweisungen werden nach den Prinzipien des *Pattern Matching* vollzogen. Diesem Zweck dienen die Domänenspezifikationen. Jede Domäne beschreibt ein Pattern, welches sich auf die Attribute des angegebenen Modellelementes bezieht. Im Fall von **`checkonly`**-Domänen erfolgt eine Prüfung und ein Abgleich der Werte der Elemente eines Quellmodells, im Fall **`enforce`** werden die Attribute des Prüfkandidaten mit den gegebenen Werten überschrieben und damit ein Objekt im Zielmodell erzeugt. Beim **`checkonly`** ist eine Besonderheit zu beachten, die hier ausgenutzt wird: `pckgName` ist eine selbstdefinierte freie Variable, diese besitzt initial keinen Wert, sie ist also `oclUndefined`. Beim *Pattern Matching* bezogen auf die Domäne „**`checkonly domain`** `source`“ wird der Wert von `name` an die freie Variable `pckgName` übergeben, womit die Bedingung erfüllt wird. Mit dieser Technik ist ein Initialisieren von Zielelementen unter Verwendung von Quellelementen möglich.

3.2 Top-level- und Non-Top-Level-Relationen

In der QVT-R wird unterschieden zwischen *top-level*- und *non-top-level*-Relationen. *top-level*-Relationen – gekennzeichnet mit dem Schlüsselwort **`top`** – sind solche, die bei der Ausführung einer Transformation grundsätzlich aktiviert und abgearbeitet werden, *non-top-level*-Relationen sind solche, die explizit aufgerufen werden müssen. In einer Transformation muss es mindestens eine *top-level*-Relation geben. Die *top-level*-Relation `Packages` würde also beim Start der Transformation alle Packages eines Modells nach einem vorgegebenen Muster untersuchen und die entsprechende Aktion im Zielmodell ausführen, vorausgesetzt, es gibt `Packages` im Quellmodell. Also sollte für jedes Element eines Quellmodells, welches „top“ in einem Diagramm vorkommen kann – also

zum Beispiel `Package`, `Class`, `Interface`, `DataType` – eine *top-level*-Relation definiert sein.

Wenn es Beziehungen zwischen den Elementen im Modell gibt, die es nötig machen, dass ein Element im Rahmen einer Objekterzeugung im Zielmodell bereits existiert, dann ist das in den Pattern-Definitionen durch ein **`when`**-Prädikat entsprechend zu formulieren. In folgendem Beispiel sind sowohl die Relationen `Packages` wie auch `Classes` *top-level*. Es gibt jedoch Klassen, die Komponenten von Paketen sind, in dem Fall müssen die entsprechenden Pakete bereits bearbeitet und im Zielmodell angelegt sein, bevor die Bearbeitung der Klassen erfolgen kann. Dies wird durch den expliziten Aufruf der Relation `Packages` in dem **`when`**-Prädikat der Relation `Classes` gewährleistet.

```
transformation GI2_TopLevel ( source:uml, target:uml )
{
   top relation Packages
   {
      pckgName : String;
      checkonly domain source srcPkg : Package
      { name = pckgName };
      enforce   domain target dstPkg : Package
      { name = pckgName };
   }
   top relation Classes
   {
      className : String;
      checkonly domain source srcCls : Class
      {
         name  = className,
         owner = srcPkg : Package {}
      };
      enforce   domain target dstCls : Class
      {
         name  = className,
         owner = dstPkg : Package {}
      };
      when { Packages  ( srcPkg, dstPkg ); }
   }
}
```

Wenn allerdings in dem Modell eine hierarchische Ordnung zwischen den Elementen existiert, wie das zum Beispiel bei UML-Modellen der Fall ist (mal abgesehen davon, dass im UML-Metamodell das Attribut `owner` ohnehin nur lesbar und nicht veränderbar ist und somit in der **`enforce`**-Domäne nicht verwendet werden kann), dann ist es oft einfacher, die Transformation explizit auszulösen durch Aufrufe von *non-top-level*-Relationen in den **`where`**-Klauseln.

```
transformation GI2_NonTopLevel ( source:uml, target:uml )
{
   top relation Package
   {
      checkonly domain source srcPkg : Package {...};
      enforce   domain target dstPkg : Package {...};
      where { Classes  (srcPkg, dstPkg ); }
   }
   relation Classes
   {
      checkonly domain source srcP : Package {...};
      enforce   domain target dstP : Package {...};
   }
}
```

3.3 when- und where-Prädikate

In den obigen Beispielen sind zwei unterschiedliche Bedingungstypen eingeführt worden, **when**- und **where**-Klauseln. **when**-Prädikate dienen als Vorbedingung für die Erzeugungskomponente einer Relation, die letztendlich erfüllt sein muss, damit der generative Teil überhaupt bearbeitet werden kann. Das **where**-Prädikat bewirkt, dass immer dann, wenn ein Zugriff auf ein Objekt oder eine Variable in der **enforce**-Domäne erfolgt, die in der **where**-Klausel beschriebenen Aktionen ausgeführt werden.

Zur Erinnerung: *top-level*-Relationen werden mit dem Start einer Transformation aktiviert und ausgeführt. Ein explizites Auslösen einer *top-level*-Relation kann nur über ein **when**-Prädikat vorgenommen werden. Der Aufruf von *non-top-level*-Relationen erfolgt ausschließlich in **where**-Klauseln. Angenommen, es gibt ein UML-Profil `UMLEJB`, in dem der Stereotyp `<<entity>>` definiert und mit der Metaklasse `Class` assoziiert ist. In dem folgenden Beispiel werden dann ausschließlich die Klassen behandelt, die mit dem Stereotyp `<<entity>>` markiert sind. `getAppliedStereotype()` und `isStereotypeApplied()` sind Standardfunktionen, die den Zugriff auf Stereotypen erlauben.

```
relation Classes
{
   className, className1 : String;
   st: Stereotype;
   checkonly domain source srcP : Package
   { packagedElement = srcCls : Class { name=className }};
   enforce   domain target dstP : Package
   { packagedElement = dstCls : Class { name=className1}};
   when { st = getAppliedStereotype('UMLEJB::entity');
          srcCls.isStereotypeApplied(st); }
   where{ className1 = className + '_DerivedFromEntity'; }
}
```

In der **where**-Klausel wird erzwungen, dass der Name der Zielklasse um einen konstanten String ergänzt wird.

3.4 Primitive Domänen und Queries

Eine Relation kann beliebig viele **checkonly**- oder **enforce**-Domänen besitzen. Bei dem Aufruf von Relationen in **when**- oder **where**-Klauseln müssen dann entsprechend der Domänentypen und Reihenfolge Aufrufparameter übergeben werden. In dem obigen Beispiel besitzt die Classes-Relation zwei Domänen, jeweils eine **checkonly** und eine **enforce**. Dementsprechend müssen beim Aufruf dieser Relation zwei Objekte vom Typ Package als Parameter mitgegeben werden. Neben den Domänen mit einer Check- und Generierungsfunktion erlaubt QVT-R, dass weitere auch neutrale, sogenannte **primitive**, Domänen definiert werden. In dem folgenden Beispiel wird damit der gerufenen Relation eine zusätzliche String-Variable übergeben, die zur Erzeugung des Namens des Zielpaketes herangezogen wird.

```
relation ClassesWithPrefix
{
   cName, cName1  : String;
   checkonly domain source pckg : Package
   { packagedElement = srcCls : Class { name = cName1 } };
   enforce domain target schm : Package
   { packagedElement = dstCls : Class { name = cName } };
   primitive domain prefix : String;
   where { cName = if prefix = '' then cName1
                   else prefix + '_' + cName1 endif; }
}
```

Aufruf:

```
ClassesWithPrefix (srcPckg, dstPckg, 'HelloWorld');
```

Der einzige Zweck für primitive Domänen ist die Option, der gerufenen Relation weitere Parameter zu übergeben. Ein ähnliches Ergebnis kann man mit der Verwendung von Hilfsfunktionen (**query**) erreichen.

```
query PrefixToClassname ( prefix : String, cName : String )
: String
{
   if prefix = '' then cName
   else prefix + '_' + cName endif
}
```

Aufruf:

```
   cName = PrefixToClassname('HelloWorld', cName1);
```

4 Zusammenfassung und Ausblick

Die QVT-R ist eine interessante, verhältnismäßig einfache Sprache, die allerdings als deskriptive Sprache und aufgrund des konsequenten *Pattern Matching*-Ansatzes eher etwas gewöhnungsbedürftig ist. Dieser Beitrag konnte lediglich einen kleinen Einblick geben in die Transformation von Modellen mit Hilfe der QVT-R. Für tiefer gehendes Interesse sei auf die angeführte Literatur verwiesen, insbesondere [Nol09]. Mittlerweile ist die Werkzeugsituation stabil und weitgehend ausgereift, so dass man sagen kann, der MDA-Ansatz der Transformationen von Modellen nach Modellen funktioniert. So mag es durchaus interessant sein, die modellgetriebene Entwicklung von Anwendungssystemen auch in der Praxis um den Aspekt der Transformation von Modellen zu ergänzen, und damit die Phasenübergänge eines iterativen Entwicklungsprozesses gerade in den frühen Phasen zu unterstützen.

Literaturverzeichnis

[EMP10] Eclipse Modeling Project; http://www.eclipse.org/modeling/

[MDA03] MDA Guide, Version 1.0.1. www.omg.org/cgi-bin/doc?omg/03-06-01, 2003

[Med10] mediniQVT; http://projects.ikv.de/qvt

[MOF06] MOF 2.0/Meta Object Facility Core, Version 2.0. http://www.omg.org/spec/MOF/, 2006

[Nol10] Nolte, S.: freeMDA - Eine freie Plattform für eine integrierte modellgetriebene Anwendungsentwicklung. EclipseMagazin - Heft 01, 2010

[Nol09] Nolte, S.: QVT – Relations Language. Springer-Verlag, 2009

[Nol07] Nolte, S.: Modelle und Metamodelle im Eclipse Kontext. ObjektSpektrum, Heft 6, 2007

[OCL10] Object Constraint Language. http://www.omg.org/spec/OCL/2.2/, 2010

[PM06] Petrasch R, Meimberg O: Model Driven Architecture. dpunkt Verlag, 2006

[QVT08] MOF QVT - Version 1.0. http://www.omg.org/spec/QVT/1.0/, 2008

[UML10] UML Superstructure and Infrastructure. http://www.omg.org/spec/UML/2.3/, 2010

[XMI07] OMG, MOF 2.0/XMI Mapping, Version 2.1.1. http://www.omg.org/spec/XMI/2.1.1/PDF/index.htm, 2007

M2M-Transformation mit der QVT Operational Mappings

Siegfried Nolte

Beethovenstraße 57
22941 Bargteheide
siegfried@siegfried-nolte.de

Abstract: QVT ist ein Sprachkonzept der Object Management Group zur Transformation von formalen Modellen. In diesem Beitrag werden ausgehend von einem einfachen Fachklassenmodell die zentralen Konzepte und Techniken der Transformationssprache Operational Mappings vorgestellt, so dass abschließend eine vollständige Transformation eines UML-Modells der PIM-Ebene in eins der PSM-Ebene vorgenommen werden kann.

1 Einordnung der QVT-O in die MDA

Modellgetriebene Architekturentwicklung bedeutet das iterative Erstellen von Anwendungssystemen über ständig wechselnde Modellierungs- und Transformationsschritte [MDA03, PM06]. Mit Hilfe von formalen Modellierungssprachen, sogenannten MOF-Sprachen, werden Modelle erstellt, die in einem Transformationsschritt mit QVT-Sprachen in ein neues Modell überführt werden, welches wiederum mit Hilfe einer formalen Modellierungssprache dargestellt und bearbeitet werden kann. Auf diesem Weg wird als Ausgangsmodell einer Entwicklungsphase jeweils ein konsistentes Modell durch Transformation initial bereitgestellt.

MOF-Sprachen sind Modellierungssprachen, denen nach dem MOF-Ansatz der OMG [MOF06] ein formales Metamodell zugrunde liegt. Das Metamodell ist gewissermaßen eine Datenstruktur, die nach der MOF-Spezifikation in Form eines UML-Klassendiagramms definiert wird [Nol07]. Alle OMG-Modellierungssprachen – so auch die BPMN, UML, SysML – sind formale Modellierungssprachen, die in diesem Sinne mit UML-basierten Metamodellen spezifiziert worden sind. Eine weitere, etwas leichtgewichtigere Möglichkeit der Beschreibung von formalen Modellierungssprachen besteht aus der Einarbeitung von domänenspezifischen Belangen in die Modellierungssprache UML. Dies Mittel stellt das UML-Metamodell mit den UML-Profilen zur Verfügung [UML10]. Im Rahmen dieses Beitrags arbeiten wir exemplarisch mit einem UML-Profil, welches eine einfache Erweiterung des UML-Metamodells in Form des Stereotyps `<<entity>>` benutzt. Mit diesem Stereotyp, der mit der Metaklasse `Class` assoziiert ist, werden die Fachklassen eines UML-Modells markiert, die mit der Transformation in besonderer Weise behandelt werden sollen. Mit der Wahl des UML-Profils ist UML als Modellierungssprache und damit als Modelltyp für die Transformation festgelegt.

2 Transformation mit QVT-O

QVT *Operational Mappings* (QVT-O) ist eine Sprache zur Transformation von formalen Modellen, die der Klasse der imperativen Sprachen zuzuordnen ist [QVT08]. In QVT-O werden die Transformationen nicht, wie bei der *Relations Language*, deskriptiv über gültige Relationen zwischen den Elementen von Quell- und Zielmodellen beschrieben, sondern imperativ durch Mapping-Operationen. Zur Erläuterung der Sprachkonzepte der QVT-O und der Modelltransformation [Nol09] soll ein einfaches Fachklassendiagramm der PIM-Ebene dienen. Die Fachklassen, die im Rahmen der Transformation in besonderer Weise behandelt werden sollen, sind unter Verwendung eines einfachen UML-Profiles mit dem Stereotyp `<<entity>>` markiert. Als Zielmodell wird ebenfalls ein UML-Klassendigramm hergestellt.

2.1 Transformationen

Eine QVT-O-Transformation beschreibt die Überführung einer beliebigen Anzahl von Quellmodellen in eins oder mehrere Zielmodelle. Quellmodelle wie auch Zielmodelle müssen von bestimmten Modelltypen sein, die zu Beginn einer Transformation deklariert werden. Modelltypen referenzieren formale Metamodelle, die im Kontext bekannt sind. Im Fall des Eclipse-basierten Werkzeugs QVT-Operational [EMP10] werden Metamodelle als Eclipse-Plugins organisiert [Nol07].

```
modeltype UML uses 'http://www.eclipse.org/uml2/2.1.0/UML';
transformation GI1_Transform (in src :UML, out trgt :UML);
main(){ src.objects()[Package]->map transformPackage(); }
```

Die **main**-Methode ist die Mapping-Operation, die als Einstiegspunkt für die Transformation gilt. Es handelt sich dabei um eine Mapping-Operation, die die Modelle selbst, hier `src` und `trgt`, zum Gegenstand hat. Aus dem Grund ist es auch nicht erforderlich – und bei manchen QVT-O-Interpretern auch nicht möglich –, Eingangs- und Ausgangsargumente für die **main**-Methode anzugeben. Mit dem Zusatz **in** und **out** wird festgelegt, ob ein Modell Quelle oder Ziel ist. QVT-O basiert auf OCL [OCL10]. Einerseits ist QVT-O mit Hilfe der OCL entwickelt worden, andererseits steht die OCL in QVT-O vollständig zur Verfügung, soweit sie in dem jeweiligen Interpreter implementiert ist. Die Syntax der Anweisungen entspricht im Wesentlichen der OCL-Syntax. Die obige Anweisung „`objects()[Package]`“ ist ein sogenannter *Shorthand* für die QVT-Standardfunktion „`objectsOfType(Package)`“.

2.2 Objekterzeugung mit Mapping-Operationen und Inline Mapping

Die Transformation besteht aus einer Reihe von Mapping-Operationen, mit denen Schritt für Schritt die Überführung von Elementen des Quellmodells in Elemente des Zielmodells beschrieben wird:

```
mapping Package::transformPackage() : Package
when { self.name != oclIsUndefined(); }{ name:=self.name; }
```

In diesem Code-Abschnitt wird ein `Package` eines Quellmodells in ein `Package` eines Zielmodells überführt. Die in dem optionalen **when**-Prädikat angegebene Vorbedingung regelt, dass dies nur dann erfolgt, wenn das `Package` des Quellmodells als Objekt der Metaklasse „`NamedElement`“ auch einen definierten Wert besitzt. **self** ist eine QVT-Standardvariable, die den Gegenstand des Mappings selbst, hier das aktuelle `Package`, repräsentiert. In Mapping-Operationen stehen weitere Standardvariablen zur Verfügung, **result**, die das Zielobjekt einer Operation in seiner jeweiligen Ausprägung repräsentiert, und **this**, womit die Transformation selbst adressiert werden kann, um zum Beispiel Informationen über den Status abzufragen.

Bei der obigen Mapping-Operation handelt es sich um eine Kurzform. Ein vollständig spezifiziertes Mapping wird in frei gestaltbaren Blöcken organisiert, jeweils einem zur Initialisierung (**init**), zur Durchführung (**population**) und zur Nachbehandlung (**end**) der Operation. Dies kann bei komplexen Mappings durchaus Sinn machen. Die komplette Mapping-Anweisung mit allen optionalen syntaktischen Bestandteilen sieht folgendermaßen aus:

```
mapping Package::transformPackage() : Package
when  { self.name != oclIsUndefined(); }
{
   init       { var pckgName := self.name; }
   population { object result:Package { name:=pckgName; }}
   end        { /* Terminierende Anweisungen, Meldungen, */
                /* Statusabfragen */ }
}
```

Der Aufruf einer Mapping-Operation erfolgt mit einer **map**-Anweisung:

```
src.objects()[Package]->map transformPackage()
```

Auf jedes Objekt vom Typ `Package` eines Eingangsmodells `src` wird die Mapping-Operation `transformPackage` angewendet. Mit dem Mapping haben wir eine Technik der expliziten Objekterzeugung kennen gelernt. Eine weitere Möglichkeit, Objekte in Zielmodellen zu erzeugen, besteht darin, dies implizit mit Hilfe eines Inline-Mappings vorzunehmen. Die Technik des Inline-Mappings bietet sich insbesondere dann an, wenn die Objekte, die im Zielmodell neu angelegt werden sollen, im Quellmodell nicht in irgendeiner Weise existieren, zum Beispiel die `getter`- und `setter`-Methoden für den Zugriff auf private Attribute:

```
mapping Class::transformClasses() : Class
{
   name := self.name;
   ownedOperation+=self.ownedAttribute->object(p) Operation
   { name := 'get_' + p.name; };
```

```
   ownedOperation+=self.ownedAttribute->object(p) Operation
   { name := 'set_' + p.name; };
}
```

2.3 QVT-O und imperative Ausdrücke – eine Auswahl

Alle QVT-Sprachen basieren auf der Object Constraint Language [OCL10]. Dadurch können in beliebiger Weise QVT-Anweisungen mit OCL-Anweisungen kombiniert werden, was den QVT-Sprachen eine große Flexibilität verleiht. Eine der wertvollsten Anweisungen ist der **log**-Ausdruck, mit dem an beliebiger Stelle in einer Mapping-Operation eine Ausgabe auf die Konsole vorgenommen werden kann, um zum Beispiel aktuelle Ausprägungen von Variablen oder Objekten anzuzeigen.

```
log ( 'Do' ); log ( self.name );
log ( 'Inhalt der self-Variable ', self );
```

Ein bedingter Ausdruck ist ein Ausdruck, in dem die Anweisungsteile in Abhängigkeit von einer Bedingung ausgeführt werden. QVT-O erlaubt als bedingte Ausdrücke **if**- und **switch** -Konstrukte:

```
if (self == oclIsUndefined())
then log ( 'Self hat keinen definierten Wert' );
else log ( 'Self.name = ' + self.name );
endif;

switch
{
   case (self == oclIsUndefined())
        { log ( 'Self hat keinen definierten Wert' ); }
   else { log ( 'Self.name = ' + self.name ); }
};
```

Schleifen und iterative Anweisungen werden entweder mit dem OCL-Iterationsausdruck **iterate** oder mit einem **while**- oder **for**-Ausdruck implementiert:

```
var s1:Integer := 0, s2:Integer := 0, s123 := Set{1,2,3};
s1 := s123->iterate(idx: Integer; i: Integer = 0 | i+idx);
s123->forEach (idx) { s2 := s2 + idx };
log ('Summe s1 = ', s1); log ('Summe s2 = ', s2);
```

Für das Sammeln von Objekten kann die **collect**-Anweisung hinzugezogen werden:

```
ownedAttribute := self.ownedAttribute->
   collect( p | object Property { name := p.name; } );
```

Für diese Konstruktion stellt die QVT-O-Spezifikation wieder einen sogenannten *Shorthand* zur Verfügung, der allerdings noch nicht von allen QVT-O-Interpretern unterstützt wird:

```
ownedAttribute :=
self.ownedAttribute->object (p) Property {name := p.name;};
```

2.4 Queries

Ein weiteres Hilfsmittel der QVT-O sind Queries. Dies sind Hilfsfunktionen, mit denen sich QVT- und OCL-Anweisungen bündeln und umfangreiche Skripte besser strukturieren lassen:

```
query getTypeName( srcType : Type) : String
{
   return if ( srcType.name = 'UnlimitedNatural' )
          then 'Double' else srcType.name endif;
}
```

Der Aufruf einer Query ist analog zu dem Aufruf einer Mapping-Operation, wobei die **map**-Anweisung weggelassen wird:

```
    name := getTypeName(self);
```

Queries in Kombination mit OCL-Anweisungen bieten recht vielfältige Möglichkeiten, mit denen sich das Leben eines MDA-Architekten doch sehr erleichtern lässt. Mit der folgenden Query zum Beispiel wird mit Hilfe der Standardfunktion `getAppliedStereotype()` geprüft, ob eine Klasse mit dem Stereotyp `<<entity>>` des UML-Profils `UMLEJB` markiert ist. Wenn ja, kann die Klasse im Rahmen der Transformation angemessen behandelt werden:

```
query Class::isStereotypedBy ( str : String ) : Boolean
{
   var st : Stereotype := null;
   st := self.getAppliedStereotype( 'UMLEJB::' + str );
   return if ( self.isStereotypeApplied(st) )
          then true else false endif;
}

mapping  Class::transformClasses () : Class
when { self.isStereotypedBy('entity') } { ... }
```

2.5 Weiterführende Konzepte

Häufig lassen sich nicht alle Transformationen und Operationen in einer Phase erledigen, so dass QVT-O spezielle Sprachmittel und Funktionen anbietet, damit bestimmte Trans-

formationsaktionen auch in nachfolgenden Phase vorgenommen werden können: Object Resolution und Intermediate Data. Alle Objekte, die im Rahmen einer Transformation angefasst werden, werden in einem *Trace* abgelegt. Mit Objekt-Resolution wird ein Ansatz beschrieben, die Objekte unter Verwendung des *Traces* zu adressieren. Dabei wird zwischen allgemeiner Resolution und spezieller Resolution unterschieden. Allgemeine Resolution gilt grundsätzlich für alle Objekte der gesamten Transformation; spezielle Resolution adressiert gezielt die Objekte einer bestimmten Mapping-Operation.

Allgemeine Resolutionsfunktionen sind:

- **`resolve()`**, welche zu einem Objekt des Quellmodells eine Liste aller Zielobjekte eines vorgegebenen Typs liefert, die aus dem Quellmodell generiert worden sind.
- **`resolveone()`** liefert das letzte Zielobjekt des gegebenen Typs, welches sich im Trace befindet.
- **`invresolve()`** wird angewendet auf ein Zielobjekt und liefert eine Liste aller Quellobjekte eines gegebenen Typs, die zur Erzeugung des Zielmodells benutzt wurden.
- **`invresolveone()`** liefert das letzte Quellobjekt eines gegebenen Typs.

Die speziellen Resolutionsfunktionen **`resolveIn()`**, **`resolveoneIn()`**, **`invresolveIn()`**, **`invresolveoneIn()`** sind analog definiert; bei ihnen wird jeweils noch die Mapping-Operation mit Namen angegeben, die bei der Selektion des oder der jeweiligen Elemente zusätzlich in die Betrachtung einbezogen werden soll.

Intermediate Data definiert ein Konzept, dem Metamodell im Rahmen einer Transformation dynamisch weitere Elemente hinzuzufügen, die nur während der Ausführungszeit der Transformation vorhanden sind und den Mapping-Operationen und Hilfsfunktionen zur Verfügung stehen. Es kann sich dabei um zusätzliche Metaklassen (**`intermediate class`**) oder um Attribute von vorhandenen Metaklassen handeln (**`intermediate property`**).

```
intermediate class Entity { name:String; kind:String; };
intermediate property Package::entities:Sequence (Entity);
```

In obigem Beispiel wird dem Metamodell eine weitere Klasse `Entity` hinzugefügt. Die Metaklasse Package erhält ein zusätzliches Attribut `entities`, welches eine Sammlung von Instanzen der Klasse `Entity` aufnehmen kann. Mit diesen Ergänzungen ist man in der Lage, Klassen im Modell, die zum Beispiel mittels eines Stereotyps als `<<entity>>` markiert sind, zu sammeln und vorübergehend einem Paket zuzuordnen, damit sie später bei der Erzeugung des Zielmodells angemessen eingebracht werden können. Bei der Vertiefung dieser Konzepte kann auf [Nol09] zurück gegriffen werden.

3 Zusammenfassung und Ausblick

QVT-O ist ein Bestandteil der *Eclipse Modeling Workbench* [EMP10] und steht damit als mittlerweile recht stabiles Produkt einem großen Kreis von Entwicklern und IT-Architekten zur Verfügung. Die für diesen Beitrag verwendete MDA-Entwicklungsplatt-

form ist in [Nol10] detailliert beschrieben. Die erforderlichen Metamodelle werden als Eclipse-Plugins erwartet, was jedoch mit Hilfe des Eclipse Modeling Frameworks und der UML Modeling Tools verhältnismäßig leicht zu bewerkstelligen ist [Nol07]. Das Metamodell der UML, so wie es hier verwendet wird, ist in einer spezifikationsnahen Ausprägung stets Bestandteil des Modeling Projektes. Die Repräsentation der Zielmodelle in einer grafischen Form ist jedoch nach wie vor eine Herausforderung, der sich die Anbieter von Modellierungswerkzeugen noch stellen müssen. Ohne dies ist es fraglich, ob das Angebot der Modelltransformation angenommen wird, da man mit mittlerweile sehr leistungsfähigen Model-nach-Text-Transformatoren auch schon in frühen Phasen der Entwicklung sinnvoll und effektiv tragfähige Ergebnisse produzieren kann. Trotzdem bleibt zu hoffen, dass Modelltransformationen über kurz oder lang mehr in die Kultur einer systematischen modellgetriebenen Architekturentwicklung einfließen werden. Dies hätte zweifelsohne wertvolle Auswirkungen auf die Qualität und Stabilität der zu entwickelnden Softwaresysteme zur Folge.

Literaturverzeichnis

[EMP10] Eclipse Modeling Project; http://www.eclipse.org/modeling/

[MDA03] MDA Guide, Version 1.0.1. www.omg.org/cgi-bin/doc?omg/03-06-01, 2003

[MOF06] MOF 2.0/Meta Object Facility Core, Version 2.0. http://www.omg.org/spec/MOF/, 2006

[Nol10] Nolte, S.: freeMDA - Eine freie Plattform für eine integrierte modellgetriebene Anwendungsentwicklung. EclipseMagazin - Heft 01, 2010

[Nol09] Nolte, S.: QVT – Operational Mappings. Springer-Verlag, 2009

[Nol08] Nolte, S.: Modelle und Transformationen – Eine pragmatische Betrachtung der MDA QVT. ObjektSpektrum - Heft 5, 8.2008

[Nol07] Nolte, S.: Modelle und Metamodelle im Eclipse Kontext. ObjektSpektrum, Heft 6, 2007

[OCL10] Object Constraint Language. http://www.omg.org/spec/OCL/2.2/, 2010

[PM06] Petrasch R, Meimberg O: Model Driven Architecture. dpunkt Verlag, 2006

[QVT08] MOF QVT - Version 1.0. http://www.omg.org/spec/QVT/1.0/, , 2008

[UML10] UML Superstructure and Infrastructure. http://www.omg.org/spec/UML/2.3/, 2010

Konferenz
SKIL 2010
–
Studentenkonferenz Informatik Leipzig

Thomas Riechert, Sören Auer

Studentenkonferenz Informatik Leipzig

(SKIL 2010)

Thomas Riechert, Sören Auer
Universität Leipzig
Institut für Informatik

{auer|riechert}@informatik.uni-leipzig.de

In Anbetracht stagnierender oder sogar zurückgehender Studierendenzahlen ist eine umfassende Nachwuchsförderung in der Informatik von besonderer Bedeutung. Studentenkonferenzen sind eine Möglichkeit die Identifikation für das Studienfach Informatik und die Begeisterung für IT-Themen allgemein bei Studenten zu wecken. Bei einer Studentenkonferenz reichen Studierende kurze Artikel über Studien-, Abschlussarbeiten oder in der Freizeit absolvierte Informatik-relevante Projekte ein. Andere Studierende, Doktoranden, wissenschaftliche Mitarbeiter und Professoren bewerten und diskutieren die eingereichten Arbeiten. Interessante und gut ausgearbeitete Einreichungen werden zur Präsentation auf der Konferenz angenommen und die besten Arbeiten mit Preisen prämiert. Eine Studentenkonferenz unterscheidet sich damit kaum von einer anderen wissenschaftlichen Konferenz. Die Themenvielfalt kann allerdings durch die Breite der vertretenen Themen größer sein und die wissenschaftliche Innovation ist bei der Bewertung der Arbeiten nicht immer das primäre Kriterium. Eine Studentenkonferenz hilft das kreative Potential von Studierenden besser sichtbar zu machen, Studierende für die Informatik und die Forschung zu begeistern, den Austausch zwischen verschiedenen Disziplinen innerhalb der Informatik zu stärken und insbesondere das gegenseitige Verständnis von Lehrkräften und Studierenden zu fördern.

In diesem Jahr wurde am Institut für Informatik der Universität Leipzig zum ersten Mal die Studentenkonferenz Informatik Leipzig (SKIL 2010) organisiert. Initiiert und maßgeblich organisiert wurde die diesjährige SKIL von der Forschungsgruppe Agile Knowledge Engineering and SemanticWeb (AKSW)[1]. Die Idee der Studentenkonferenz wurde von allen Lehrstühlen und Mitarbeitern positiv aufgenommen. Bedingt durch die relativ kurzfristige Vorlaufzeit und die erstmalige Organisation besteht allerdings noch Potential die Unterstützung für die Konferenz in folgenden Jahren zu verstärken und zu verbreitern.

[1] http://aksw.org

Eingereicht wurden 12 Beiträge, die ein Themenspektrum von der Anwendung von Requirements Engineering Methodologien bis zur Entwicklung von Semantic Web Applikationen aufspannen. Ungefähr die Hälfte der Einreichungen sind Zusammenfassungen von Abschlussarbeiten, die andere Hälfte Ergebnisse aus Praktika und Seminaren. Jeweils zwei Mitglieder des Programmkomitees begutachteten jede Arbeit und 10 Einreichungen wurden zur Publikation und Vorstellung bei der Konferenz ausgewählt. Auf Basis der Gutachten und mit Unterstützung der betreuenden Hochschullehrer und Mitarbeiter sind schließlich die in diesem Kapitel abgedruckten Zusammenfassungen entstanden.

Die Konferenz fand am 28. September 2010 im Rahmen der GI Jahrestagung 2010 in Leipzig statt.

Thomas Riechert und Sören Auer
Vorsitzende des Programmkomitees SKIL 2010

Programmkomitee

Besonderer Dank gilt den folgenden Mitgliedern des Programmkomitees der SKIL 2010:

- Timofey Ermilov
- Philipp Frischmuth
- Norman Heino
- Sebastian Hellmann
- Jens Lehmann
- Michael Martin
- Mohamed Morsey
- Axel-C. Ngonga Ngomo
- Claus Stadler
- Sebastian Tramp
- Jörg Unbehauen

Evolutionary Graph Clustering using Graph and Cluster Mixtures

Jan Rausch

ec@jan-rausch.de

1 Introduction

Many networks and accordingly their representation in graphs are subject to structural changes during the course of their existence [CKT06]. Examples for such evolutionary networks include friendship networks in online communities, co-authorship networks in the scientific domain and collocation networks in computational linguistics. Studying the evolution of such networks can provide vital data for understanding the dynamics of communities. This is the task underlying evolutionary clustering.

The idea behind evolutionary graph clustering is to create a clustering which doesn't differ much from the previous one and still has a high quality at the current snapshot. The most common approach is to provide an algorithm which optimizes a function that measures the current snapshot quality (CS) and the temporal changes (CT):

$$cost = \alpha * CS + (1 - \alpha) * CT \tag{1}$$

An example of a framework based on this approach is given in [Lin09] with FacetNet.

The aim of this work is to show that it is possible to reduce the problem of evolutionary graph clustering to the data level. Basically that means we want to change the graphs used for the clustering instead of providing a new clustering algorithm. Furthermore we aim at providing a framework which is independent from the underlying clustering algorithm. The main advantage of such an approach is obvious: most of the static algorithms available today could be used for clustering evolutionary graph. In the following, we show how we can use a linear graph and cluster mixture approach to tackle the problem of evolutionary graph clustering.

2 Graph and Cluster Mixtures

We use the well-known representation of a graph $G = (V,E,\omega)$ consisting of a set of Nodes N , a set of Edges E and a weight function ω. A Clustering Cl belonging to a graph G is a set of Clusters C_i. A Cluster $C_i \subseteq N$ belonging to a graph G is a subset of nodes of G similar in some sense. To achieve our goals we first have to define some operations on graphs:

Addition of two graphs $G_1 + G_2$:

$$G_1 + G_2 = (V_1 \cup V_2, E_1 \cup E_2, w_+(v_i, v_j) = w_1(v_i, v_j) + w_2(v_i, v_j))$$

Multiplication of a graph G with a real value r:

$$G_* = r * G = (V, E, w_*(v_i, v_j) = r * w_1(v_i, v_j))$$

The intersection of two graphs $G_1 \cap G_2$

$$G_\cap = G_1 \cap G_2 = (V_1 \cup V_2, E_1 \cap E_2, w_\cap)$$

with

$$w_\cap(v_1, v_2) = \begin{cases} w_1(v_1, v_2) & \Leftrightarrow (v_1, v_2) \in E_1 \cap E_2 \\ 0 & else \end{cases}$$

Graph built upon a Clustering:

$$Graph(Cl) = (\bigcup C_i, \bigcup (C_i \times C_i) \setminus \{(v_j, v_j) | v_j \in C_i\}, w_{Cl})$$

$$w_{Cl}(v_i, v_j) = w_{Cl}(v_j, v_i) = \begin{cases} |\{C_i | v_i \in C_i \wedge v_j \in C_i\}| & iff\ v_i \neq v_j \\ 0 \end{cases}$$

The resulting graph of the operation *Graph*(*Cl*) is a graph consisting of all nodes of the clustering and edges between nodes sharing one cluster (besides edges from a node to itself). The weight of the edges is equal to the number of clusters both nodes shares.

A graph mixture is a graph computed out of G_t and the previous clusterings or previous graphs. Due to lack of space, we only present one mixture in this work:

$$h_N(G_t) = \alpha * G_t + (1 - \alpha) * (G_t \cap Graph(Cl_{t-1})). \quad (2)$$

The resulting graph of $h_N(G_t)$ is a graph consisting of the same nodes and edges as G_t but different edge weights. The weight of the edges between nodes sharing a cluster in Cl_{t-1} increases, the weight between other nodes decreases.

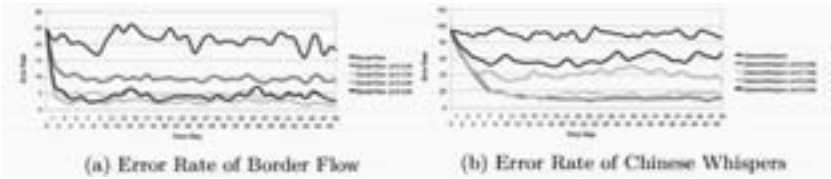


Figure 1: Error rate of the Four Community Experiment: *z=8* depending on α. The dark blue line is the error rate of the non-evolutionary clustering results. A lower Error Rate implies a better clustering result.

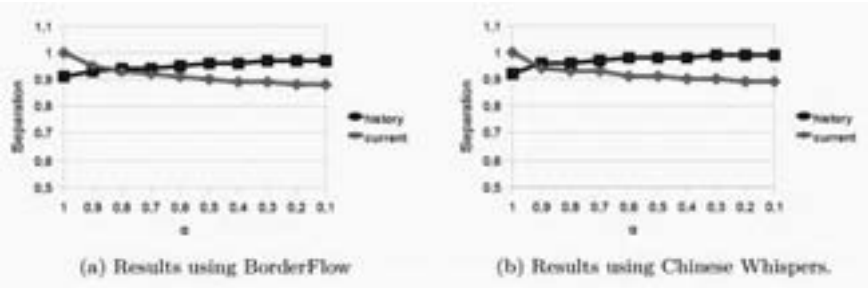


Figure 2: Separation values between current clustering and the previous one (history) respectively the clustering of a non-evolutionary version (current) using h_N.

3 Experiments

We use a synthetic dataset described in [Lin09] as "Synthetic Dataset # 2" and a DBLP dataset described in [BBBG09].

The synthetic dataset was used to show the capabilities of our approach to improve the clustering results of static graph clustering algorithms over time (see Figure 1). The dataset simulates a network consisting of four communities. Each community has 32 members (nodes). The mean number of edges of each node is 20. The parameter z sets the number of edges between a given node and nodes in a different community different from its own. In every time step, three of the members of each community change their community. The resulting data set is thus an evolutionary graph with scalable ratio of noise. Our first experiment shows that the clustering quality improves significantly if our approach is used. We obtained results comparable to that of algorithms such as described in [Lin09] which utilized an algorithm-based optimization. As the underlying clustering algorithms we used Border Flow [NS09] and Chinese Whispers [Bie06].

The DBLP dataset shows that our approach can increase and decrease the distance to the previous and current time step depending on (see Figure 2). With this dataset we want to show, that our approach is capable of producing a consistent clustering result. As the underlying clustering algorithm we used BorderFlow and Chinese Whispers. For the history we used the results of the previous time step and for the current clustering we used the clustering results of the non-evolutionary version. For different the clusterings were computed and compared to these clusterings. We used the separation metric as defined in [BvH06] to measure similarity. The outcome shows, that with decreasing the distance to history also decreases and the distance to the current results increases.

4 Conclusion

We showed that it is possible to perform evolutionary clustering using a combination of static clustering algorithms and graph mixtures. During our experiments we used a special mixture, h_N. We could show that we obtain results that are comparable to those of state-of-the-art algorithms designed especially for evolutionary clustering. Furthermore, the DBLP experiment showed that we can use a value to adjust the distance to history respectively the current snapshot. Of course this approach doesn't provide an all-in-one solution for every problem of evolutionary graph clustering. The main advantage of our approach is that it does not depend on the underlying algorithm and so one can use an algorithm already achieving good clustering results in static networks of the underlying area. In future work we have to analyze how this approach behaves with different mixtures, clustering algorithms and graph topologies.

References

[BBBG09] Michele Berlingerio, Francesco Bonchi, Björn Bringmann, and Aristides Gionis. Mining Graph Evolution Rules. In ECML PKDD '09, 2009.

[Bie06] C. Biemann. Chinese Whispers - an Efficient Graph Clustering Algorithm and its Application to Natural Language Processing Problems. In Proceedings of the HLT-NAACL-06, New York, USA, 2006.

[BvH06] Sylvain Brohée and Jacques van Helden. Evaluation of clustering algorithms for protein-protein interaction networks. BMC, 7:488, 2006.

[CKT06] Deepayan Chakrabarti, Ravi Kumar, and Andrew Tomkins. Evolutionary clustering. In KDD '06, pages 554–560, New York, 2006. ACM.

[Lin09] Yu-Ru Lin. Analyzing communities and their evolutions in dynamic social networks. ACM Trans. Knowl. Discov. Data, 3(2):1–31, 2009.

[NS09] Axel-Cyrille Ngonga Ngomo and Frank Schumacher. BorderFlow: A Local Graph Clustering Algorithm for Natural Language Processing. In CICLing, pages 547–558, 2009.

Abstracting Queries

Jonas Brekle

jonas.brekle@gmail.com

1 Introduction

The term “Semantic Web” was coined by Tim Berners-Lee and refers to the data model as well as methods and technologies that allows machines to understand the meaning - or “semantics” - of information. The W3C supervises the development of several standards that are related to the representation and processing of data that is semantically annotated. According to Tim Berners-Lee’s vision, the availability of machine-readable metadata shall enable software agents to access the Web in a deeper way. And thus be able to perform tasks automatically and locate related information on behalf of the user.

These technologies include the Resource Description Framework (RDF) as the underlying data model, a variety of data interchange formats (e.g. RDF/XML, N3, Turtle, N-Triples), powerful base vocabularies for data modeling such as RDF Schema (RDFS) and the Web Ontology Language (OWL). All of these are intended to provide a formal description of concepts, terms, and relationships of entities. This is what we call an ontology - a formalization of a conceptualization of knowledge.

My bachelor thesis [Bre10] (and hence this paper) deals with the OntoWiki project [ADR06], which is a open source semantic wiki and ontology engineering tool. It is one of the main projects of the “Agile Knowledge Engineering and Semantic Web” (AKSW) research group. OntoWiki provides the visual presentation of a knowledge base as an information map, with different views on instance data. It enables intuitive collaborative authoring of semantic content and publishing Linked Data. It is a modular PHP web application built on the *Zend* framework, using the *Erfurt* API, which is a multiple-backend database abstraction layer with integrated access control, versioning and caching features. It is currently under active development within the context of a European Union Research Project since 2006.

2 Abstracting Queries

A very common view on the data is a list, which displays all instances of a class. Due to various well-established and arbitrary user-defined schemes, this view is not powerful enough to keep up with the Semantic Data Web. Also users want to modify this view to their needs and filter for many criteria. This gave us the idea to extend this view to an API that enables further development: The list is generated using a SPARQL query[1] [PS07] that selects entities in the current RDF graph. This is done within a special class[2].

[1] http://www.w3.org/TR/rdf-sparql-query/
[2] OntoWiki_Model_Instances

Prior to my work queries were constructed by straight-forward string operations. An extension with filters and the extensive reuse with other components, that want to use the query as a black box and extend the query with additional patterns, renders naive string constructions improper. The solution proposed is an abstraction of the query language on the basis of its grammar, where the production rules are represented by classes but also has many convenience methods that spare developers the technical grammar details.

To give a short overview of the idea: Most important is the Query class; for example it holds methods to set limit, offset, projection modifiers and an object that represents the *where*-part. There is a *toString* method that recursively builds the query string. The *where* object is an instance of the *GraphPattern* class, which is a container for other objects and again has a *toString* method and so on. The complete documentation can be found in my thesis or in the online documentation[3] of OntoWiki. The details are not too interesting, so I will just give you some examples[4]:

```
$query = new Query();
$triple = new Triple(
       $s = new Var("s"),
       new Var("p"),
       new Var("o"));
$query ->addElement($triple)
       ->setLimit(20);
echo $query;
```

Produces the output:

```
SELECT *
WHERE {
?s ?p ?o
}
LIMIT 20
```

The idea is to have an object orientated approach of building queries that can operate on grammar level. This query for example selects all entities. Other components of OntoWiki can now get the query via the application framework, extend the query for their need; while regarding it as a black box and thus just adding a new view. For example some component wants to take the entities that have been selected previously and display them on a map by doing the following:

```
$triple2 = new Triple(
       $s,
       new IriRef("ex:geo"),
       new Var("location"));
$query ->addElement($triple2)
$results = $store ->sparqlQuery($query);
```

[3] http://docs.ontowiki.net/fw/
[4] The examples are slightly abbreviated for understand ability

The integration in our application is more complex of course: for example we have an additional abstraction layer for lists. So other components do not need to instantiate query objects, but rather call methods like *addShownPropery*, *addFilter* etc. Furthermore this abstract instances object is stored in the web server session to persist user navigation and realize application feeling.

There are for example the map component (that displays a geo-location view of the instances), a source component (that displays the N3 serialization), the versioning component (that shows the history of the selected instances), a community activity monitor and some modules that allow the user the modification of the list: a pager that changes limit and offset, the "shown properties" module that is able to add and remove properties of the instances to display, and finally a module to add custom filters like property-value restrictions (*value equals*, *bound*, *larger*, *smaller*, *between*, *contains*). It is also possible to invert the filter with a not checkbox and remove filters. Furthermore if you want to use a filter that does not match one of these schemes, you can type your own query, which then is parsed by a SPARQL parser [Bru10] and the result is displayed in the list - with the generic features of OntoWiki.

3 Implementation

What I want to present here is the general concept. The implementation introduced 71 PHP classes to map the 100 rules of the grammar; though this may seem much, it took 4354 lines of PHP code – so only an average of 61 lines per class. The classes itself are either a container for other objects (for example the *GroupGraph*-Pattern class simply hosts an array that contains objects like triples, filter expressions etc.). It then has methods to manage the elements. Or it is an element of a container (like for example an instance of the *Var* class) and has methods to modify or get the properties of the object (for example the name of the variable). But this distinction is not exclusive. There are two abstract helper classes to reuse this concept: *ContainerHelper* and *ElementHelper* and the first is inherited from the second.

The query classes extend one of these abstract classes and bring own functionality by defining some methods used to set specific properties of the object and then implement a *toString* method that builds a string representation of these properties.

The inheritance hierarchy of the query classes directly corresponds to the grammar. The types of arguments of methods and constructors, which are used for passing objects that should be contained in another query object, are restricted[5] according to the production rules of the grammar. For example the grammar defines the object (third part) of a triple to be a *GraphNode* and a *GraphNode* in turn (by various rules) as shown in figure 1.

The constructor of the Triple class restricts the third argument to the *GraphNode* PHP interface. Respectively the classes *Var*, *IriRef*, *RDFLiteral*, *NumericLiteral*, *BooleanLiteral*, *BlankNode*, *Nil*, *Collection* and *BlankNodePropertyList* implement this

[5] called type hinting as it is a new feature to the loosely typed language PHP

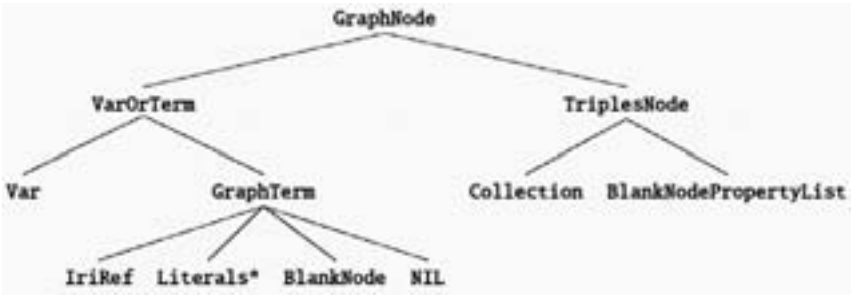


Figure 1: A hierarchy within the SPARQL grammar

interface. This concept is analogical applied to all grammar rules; this guarantees that only valid queries are constructed. All methods use the chaining mechanism: if a method would return void, *$this* (the current object) is returned instead (which is the case for setters etc.). This enables abbreviated method calls as in the first example.

4 Conclusion

The benefits of abstracting queries like described have been mentioned in section 1 and indeed: Queries, that are available in full detail as a tree of objects, where each object is of a specific type and thus has a defined semantic, are way more usable in situations in which a query needs to examined or changed. By encapsulating this functionality into classes we get the common advantages of object oriented programming and within our application we now make extensive use of it. Doing the same in other (type safe) programming languages and for other query languages (for example even SQL) seems a promising idea, though we have not looked into that. This approach of modifications of queries has shown to be both flexible and easy and in addition it enables further development and powerful workflows.

References

[Gla08] Werner Gladen. Performance Measurement: Controlling mit Kennzahlen. Gabler, 4. überarbeitete. Auflage, September 2008.

[Bre10] Brekle, J.: Spezifikation und Implementierung einer SPARQL Query API für PHP sowie Integration in OntoWiki (2010)

[ADR06] Auer, S., Dietzold, S., Riechert, T.: OntoWiki - a tool for social, semantic collaboration. In Proceedings of 5th International Semantic Web Conference 2006. Volume 4273 of Lecture Notes in Computer Science., Springer (2006) 736–749

[PS07] Prud'hommeaux, E., Seaborne, A.: Sparql query language for rdf (working draft). Technical report, W3C (2007)

[Bru10] Brunec, R.: Ein modularer und backend-unabhängiger SPARQL / SPARUL Parser für PHP (2010)

jQuery Plugin für eine erweiterte Autovervollständigung auf der Basis von SPARQL-Endpunkten

Marcus Nitzschke, Clemens Hoffmann, Thomas Schöne,
Konrad Baumheier, Marina Mitjagin

[mam08crr|mam07cmt|mam07bcd|mam07eox|mam06dva]
@studserv.uni-leipzig.de

Die automatische Vervollständigung von Sucheingaben ist eine häufig genutzte Funktion und kann in vielen Fällen das Finden von bestimmten Inhalten vereinfachen. Dennoch scheitern die meisten Implementierungen, wenn nur wenig Anhaltspunkte bei der Suche existieren und daraus resultierend eine große Menge an möglichen Ergebnissen vorhanden ist. Diese Lücke kann durch die Vorteile des Semantic Web mit seinen verknüpften Informationen geschlossen werden.
Das Ziel des Projektes aLiDa (autosuggested Linked Data) ist es, beliebige SPARQL-Endpunkte [PS08] zu durchsuchen und anhand von wenigen vorhandenen Informationen durch die RDF-Graphen [MM04] letztlich zum gesuchten Ziel zu navigieren. So soll eine erweiterte Autovervollständigung für eine Suche nach Ressourcen ermöglicht werden. Primär soll das Plugin im OntoWiki [1] zum Einsatz kommen, aber auch das Einbinden in andere Umgebungen ist problemlos möglich.

Das Projekt entstand während des Softwaretechnik-Praktikums im Sommersemester 2010 der Universität Leipzig in einer Gruppe von 5 Studenten.

Bei jQuery [2] handelt es sich zunächst um ein umfangreiches Framework für Javascript, mit dem es möglich ist moderne, dynamische Webanwendungen zu entwickeln oder aber normale Webseiten aufzuwerten. Die Stärken des Frameworks liegen vor allem in der Manipulation des Document Object Model (DOM) [ABC+98], also zum Beispiel einer HTML-Struktur. Des Weiteren lassen sich viele bekannte Javascript-Anwendungsfälle um einiges einfacher und intuitiver benutzen. Hier ist vorallem die Nutzung von Asynchronous Javascript and XML (AJaX) zum Versenden von HTTP-Requests zu nennen.

Derzeit umfasst das Plugin als Grundfunktionen das variable Aktivieren und Deaktivieren von SPARQL-Endpunkten, welche durchsucht werden sollen. Daneben gibt es eine erweiterte Darstellung der Suchergebnisse, die unter anderem den Unified Ressource Identifier (URI) [BL98] der Ressource beinhaltet.
Das Hauptaugenmerk des Plugins liegt auf den so genannten Facetten. Die Facetten umfassen alle Eigenschaften der Suchergebnisse. Das bedeutet, der Benutzer muss nicht die komplette Treffermenge manuell durchsuchen, sondern schränkt anhand der vorhan-

[1] OntoWiki: http://ontowiki.net/Projects/OntoWiki
[2] jQuery: http://jquery.com

denen Eigenschaften die Treffermenge zunehmend ein, bis er letztlich am gewünschten Suchergebnis angekommen ist. Dies ermöglicht komplexe und umfangreiche Suchen, die mit Standard-Suchmaschinen nur schwer umzusetzen sind.

Bei der Einbindung des Plugins wird die Möglichkeit geboten wichtige Parameter zu ändern und somit das Plugin an seine Bedürfnisse anzupassen. Darunter fallen Standardparameter, wie die Adressen der definierten SPARQL-Endpunkte, aber auch optionale Parameter, wie die Anzahl der angezeigten Ergebnisse, die Anzahl der vorgeschlagenen Facetten oder die Anzahl der einzugebenden Buchstaben bis die Suche beginnt.

Neben der Anpassung von Parametern gibt es außerdem die auch in jQuery oft genutzte Möglichkeit der Callback-Funktionen, mit denen das Plugin noch feiner angepasst werden kann und sogar komplett neue Funktionalitäten hinzugefügt werden können. Damit kann zum Beispiel die Standardsortierung der Ergebnistreffer nach Belieben angepasst werden.

Eine Einschränkung ist derzeit, dass aufgrund der Same Origin Policy [3] nur SPARQL-Endpunkte abgefragt werden können, die aus der selben Quelle wie der Einsatzort des Plugins stammen. Um das Problem zu umgehen, ist es denkbar in späteren Versionen JSONP [4] zu implementieren. Es ist aber ebenso möglich mit Hilfe eines lokalen Proxy die Anfragen an den entsprechenden Server weiterzuleiten.

Abschließend bleibt noch zu erwähnen, dass das Projekt bereits in mehreren OntoWiki-Instanzen evaluiert werden konnte.
Zukünftig wird die weitere Entwicklung als Projekt der Forschungsgruppe Agile Knowledge Engineering and Semantic Web (AKSW) fortgeführt. Weitere Informationen und Demos können daher der Projektseite [5] entnommen werden.

Literatur

[ABC+98] Vidur Apparao, Steve Byrne, Mike Champion, Scott Isaacs, Ian Jacobs, Arnaud Le Hors, Gavin Nicol, Robert Sutor, and Lauren Wood. Document Object Model (DOM) Level 1 Specification. W3C Recommendation, October 1998.

[BL98] Tim Berners-Lee. Cool URIs don't change, 1998.

[MM04] Frank Manola and Eric Miller. RDF Primer. World Wide Web Consortium, Recommendation REC-rdf-primer-20040210, February 2004.

[PS08] Eric Prud'Hommeaux and Andy Seaborne. SPARQL Query Language for RDF. World Wide Web Consortium, Recommendation REC-rdf-sparql-query-20080115, January 2008.

[3] The Art of Software Security Assessment. Same-Origin Policy Part 1. `http://taossa.com/index.php/2007/02/08/same-origin-policy/` (03.07.2010)

[4] Bob Ippolito. Remote JSON - JSONP. `http://bob.pythonmac.org/archives/2005/12/05/remote-json-jsonp/` (03.07.2010)

[5] Projektseite: `http://aksw.org/Projects/aLiDa`

Requirements Engineering für Berichtsapplikationen im Produktionsprozess des BMW Werk Leipzig mittels Technologien des agilen Informationsmanagement - Anwendung der SoftWiki-Methodologie zur Realisierung von Produktionscockpits

Enrico Vierig

enrico.vierig@gmx.de

Erfolgreiches Requirements Engineering (RE) ist eines der entscheidenden Einflussfaktoren bei der Entwicklung von Softwareprodukten. Wird die falsche Herangehensweise gewählt, kann eine Erkenntnis bezüglich falsch getroffener Entscheidungen und sich ändernden Anforderungen zu groben Zeitverschiebungen [Rup07] oder letztendlich sogar zum Scheitern des Projektes führen [sta04].

Der Artikel beschreibt die Anwendung der SoftWiki-Methodologie [LR09] am Beispiel einer Umsetzung eines Produktionscockpits im BMW Werk Leipzig. Dabei wird analog der Produktlinienentwicklung [PBL05], mit dem Ziel der Wiederverwendung von Softwareartefakten, vorgegangen. Bei der Anforderungserhebung soll die einheitliche Sichtweise zwischen allen beteiligten Stakeholdern verbessert werden. Hierbei kommt das SoftWiki-Tool als Kollaborationsplattform zum Einsatz. Mit diesem können Anforderungen erhoben und in einem semantischen und datenorientierten Wiki gepflegt werden. In der Phase des Requirements-Management werden durch das Entwicklungsteam Entscheidungen hinsichtlich verschiedener Dimensionen getroffen. Diese spiegeln sich beispielsweise in der Wahl des richtigen Tools zur Umsetzung eines Dashboards wider. Performance Dashboards zeigen komprimierte und aggregierte Informationen auf einer oder wenigen digitalen Webseiten [Gla08]. Dabei gibt es meist einstellbare Parameter bezüglich der Wahl einer Kennzahl, dem Detailgrad oder dem Trendregler, der eine Zukunftsprognose prozentual beeinflusst. Der Fachausdruck ”Produktionscockpit“ wird dabei oft als Synonym verwendet. Letztendlich erinnert ein Dashboard mit mehreren Tachometer-Diagrammen an ein Cockpit aus dem Automobil oder Flugzeug. Der übergeordnete Begriff „Berichtsapplikation“ bezieht sich auf jegliche Darstellungsform von Kennzahlen in Berichten mittels Technologien der elektronischen Datenverarbeitung.

Eine weitere Betrachtung zur Umsetzung des Projektes betrifft den Standardisierungsgrad. So muss entschieden werden, ob eine im Cockpit umzusetzende Kennzahl standardisiert werden muss und somit werksübergreifend zur Verfügung steht oder ob sie einen werksspezifischen Charakter hat. Sie kann auch bereichsindividuell sein und bietet dann weniger Informationsgehalt für andere Abteilungen. Für die Ermittlung von Kennziffern wird operativ auf Quelldatenbanken oder auf ein Datawarehouse zugegriffen. Jede dieser Dimensionen beschreibt ein eigenes Problemfeld, welches bei der Um-

setzung des visualisierten Kennzahlensystems beachtet werden muss. Die Lösungsmenge in jedem Bereich soll anhand der Anforderungen auf ein Minimum eingegrenzt oder idealer Weise auf eine richtige Entscheidung beschränkt werden.

Es werden zwei Ansätze verfolgt. Zum einen werden Entwickler und Fachexperten und damit ihre Erfahrungen direkt einbezogen. Zum anderen soll die Entscheidung automatisiert mittels Machine Learning Algorithmen, wie sie der DL-Learner [Leh09] umsetzt, gefunden werden. Da bei der Entwicklung des ersten Cockpits noch kein vollständiges Entscheidungsmodell vorliegt, soll die Klassifizierung initial durch Einsatz des DL-Learners automatisch geschehen.

Im Folgenden wird dann auf das entworfene Entscheidungsmodell zurückgegriffen. Damit die Unterstützung bei der Erhebung von Anforderungen und der Lösungsfindung in den entsprechenden Dimensionen gewährleistet werden kann, müssen folgende Schritte erarbeitet werden. (1) Im ersten Schritt muss die Kriterienlandschaft für Berichtsapplikationen identifiziert werden. Dazu müssen Kriterien spezifiziert und in entsprechenden Dimensionen kategorisiert werden. Für die Erhebung erster Anforderungen muss darauf aufbauend (2) ein initialer Fragenkatalog für den Kunden entworfen werden. Unter Verwendung dessen kann in Zusammenarbeit mit dem Kunden die (3) Anforderungserhebung unter Hilfe des SoftWiki-Tools vorgenommen werden. Im nächsten Schritt (4) soll ein Modell zur Entscheidungsfindung entwickelt werden, welches durch Gewichtung der entsprechenden Kriterien implementiert wird.

Der Prozess der parallelen Aufnahme weiterer und sich eventuell ändernder Anforderungen sowie der Entwicklung der Berichtsapplikation wird dabei ebenfalls durch die kollaborative SoftWiki-Plattform unterstützt. Die zutreffenden Entscheidungen bei der Entwicklung des Cockpits werden anhand des hinterlegten Modells getroffen. Im Kontext der Produktlinienentwicklung können die so getroffenen und dokumentierten Entscheidungen wieder verwendet werden. Dazu werden diese analysiert und zu Entscheidungsmustern zusammengefasst.

Literaturverzeichnis

[Gla08] Werner Gladen. Performance Measurement: Controlling mit Kennzahlen. Gabler, 4. überarbeitete. Auflage, September 2008.

[Leh09] Jens Lehmann. DL-Learner: Learning Concepts in Description Logics. Journal of Machine Learning Research (JMLR), 10:2639–2642, 2009.

[LR09] Kim Lauenroth und Thomas Riechert. Der SoftWiki-Ansatz für verteiltes Requirements Engineering mit großen Stakeholdergruppen. In Sören Auer, Kim Lauenroth, Steffen Lohmann und Thomas Riechert, Hrsg., Agiles Requirements Engineering für Softwareprojekte mit einer großen Anzahl verteilter Stakeholder, Jgg. XVIII of Leipziger Beiträge zur Informatik. Leipziger Informatik-Verbund (LIV), 2009.

[PBL05] Klaus Pohl, Günter Bäckle und Frank J. Linden. Software Product Line Engineering, Jgg. XXVI. Springer, 2005.

[Rup07] Chris Rupp. Requirements-engineering und-management. Hanser, 2007.

[sta04] Standish Group International, Inc.: Chaos Report v 3.0. http://www.standishgroup.com/, 2004.

MM2RDF: A Framework for Extraction of Semantic Multimedia-Metadata

Daniel Gerber

daniel.gerber@me.com

It is not possible to imagine a person's daily routine without multimedia anymore. Multimedia accompanies us as we leave the front door in the morning and turn on the mp3 player, buy a tram-ticket on a smartphone, check for new mail with the help of a netbook at the park or simply watch a television series after a hard day's work. We do not only consume or generate an extensive amount of multimedia but also hoard everything we can, due to the increasing storage capabilities of new end user devices. This leads to a multimedia oversupply and an information fragmentation on users' desktops and furthermore to an aggravated retrieval of multimedia content.

In this paper we present a framework which deals with the abovementioned problem, by extracting metadata from files and transforming them to a semantically enriched and machine processable format. This framework is part of a larger task which is discussed in detail in a master thesis titled *"Development and use of strategies for semantic multimedia management"*.

As defined in [BBJ06] multimedia is in general "the use of different technical media, presentation forms and sensory channels for communication". In particular these presentation forms may be text, pictures, audio, video and animation and form the input of the framework. The framework itself is light-weighted, written in PHP, easily extensible and uses several libraries, e.g. Erfurt[1] as well as Zend[2] for backend integration and a set of Linux command line applications for specialized multimedia operations like image resizing or video editing. Additionally, MM2RDF is highly extensible in terms of the vocabulary used to generate the semantic metadata. As the Semantic Web already provides a number of widely used vocabularies, which are suitable for multimedia annotation, there is no need to create and publish another one. A complete discussion of all concepts and properties used by this framework might be out of the scope of this paper, nevertheless a listing of used vocabularies should be discussed for the sake of completeness. The NEPOMUK File Ontology[3] and Contact Ontology[4] as well as the Dublin Core Metadata Element Set[5] is used on a very basic not yet multimedia specific level to type files as file resources and save basic properties like the file's name. For musical information the Music Ontology[6] is

[1] `http://www.aksw.org/Projects/Erfurt`
[2] `http://framework.zend.com/`
[3] `http://www.semanticdesktop.org/ontologies/nfo/`
[4] `http://www.semanticdesktop.org/ontologies/nco/`
[5] `http://dublincore.org/documents/dces/`
[6] `http://musicontology.com/`

used. Pictures get annotated using the EXIF RDF Schema[7]. The evaluation of video related vocabularies has not yet been completed. Moreover, Adobe's PDF vocabulary XMP[8] is used to describe PDF documents and the FOAF vocabulary is used to describe authors and producers.

Functional principle MM2RDF basically has two different usage scenarios. On the one hand, a single file is the framework's input. This file might be local or on a remote server. If the latter is the case, the size of the file gets determined and if it's below a threshold, the file gets imported to the local filesystem. Afterwards the metadata generation process gets started by initializing the backend (Virtuoso and ZendDB are supported currently) and generating common filesystem information like filename, super-folder, file-rights, file-hash etcetera. Subsequently the file's MIME type is determined and checked for existence with a RESTful request at `http://mediatypes.appspot.com/`. The so derived URIs are used to instantiate a MIME type specific metadata generator, which for example reads the EXIF information stored in the JPEG header and generates RDF Metadata in the form shown in the RDF PHP Specification[9]. On the other hand, MM2RDF can be used in a directory context to generate a semantic model of a local filesystem with the underlying folder hierarchy. This idea has been shown before in [SH09] and [SP10], but those approaches are built upon Java and do not mainly focus on multimedia metadata.

At the moment there exist twelve different MIME type extensions for the most commonly used file types (e.g. application/pdf, video/x-msvideo, audio/mp4 or image/jpeg) in MM2RDF. This extension mechanism is very straightforward and follows only some simple naming conventions. In addition, it has been evaluated that those extensions are working within an acceptable timeframe from about 1,5s for processing images to up to 6s for processing videos. An extensive description of this mechanism and of the evaluation can be found at [Ger10].

References

[BBJ06] Günter Bentele, Hans-Bernd Brosius, and Otfried Jarren. Lexikon Kommunikations- und Medienwissenschaften, 2006.

[Ger10] Daniel Gerber. Development and use of strategies for semantic multimedia management, 2010.

[SH09] Bernhard Schandl and Bernhard Haslhofer. The Sile Model — A Semantic File System Infrastructure for the Desktop, 2009.

[SP10] Bernhard Schandl and Niko Popitsch. Lifting File Systems into the Linked Data Cloud with TripFS, 2010.

[7] `http://www.w3.org/2003/12/exif/`
[8] `http://www.adobe.com/products/xmp/`
[9] `http://n2.talis.com/wiki/RDF_PHP_Specification`

Patterns for Semi-automatic Evolution and Refactoring of RDF Knowledge Bases

Christoph Rieß

c.riess.dev@googlemail.com

1 Introduction

With the emerging semantic web, RDF/OWL knowledge bases of all sizes came into existence and use. While applications are evolving, ontology concepts should change with them, but ontology refactoring was left behind for missing tool support or overcharging complexity of operations. To overcome this gap between the poorly structured data and the thoroughly engineered applications, a semi-automatic way of evolution on knowledge bases is presented here. Our approach will use evolution patterns and a processing system for them to restructure and assure data quality on modern semantic web data sources. Because of the nature of this problem not being completely new, there was knowledge from schema evolution on present-day relational database systems and software code refactorings in object-oriented programming used in background, as well as user's experience with RDF graph evolution at their domains. This work is structured mainly into 5 sections, where in Section 2 the conception and correspondence to underlying semantic web technologies is explained. Section 3 gives an overview over how patterns for our approach were acquired with methods of statistics and classified into three main classes. In the next section, Section 4, the integration of such a system with the OntoWiki framework, with a real-world example, is presented. A short overview of related concepts, not only in the field of ontology evolution, is given then (Section 5). Finally in the last Section 6 a conclusion and outlook to possible future work is given.

2 Concept and Technology

To ensure effective evolution of knowledge bases, patterns, like defined in this work, are trying to correspond as close as possible with the RDF data model. Patterns are executed by a pattern engine on an underlying triple store. Generally a pattern is basic (atomic) or compound, in the compound case, it consists of one or more (basic or compound) patterns. A base pattern is a triple with a set of variables V , a SPARQL query template S, and multiple update patterns U. The compound pattern is a sequence of n base or other compound patterns. An evolution process for a basic pattern is split into multiple steps. First there must be all variables bound, in second step the SPARQL query template is filled with bindings for V . It is executed and the result from the query is stored. In the third step changes get generated by evaluation of all update patterns U. The process for compound patterns isn't differing much from the one for base patterns, except that all sub patterns are executed in a well-defined sequence.

3 Core Pattern Library

A comprehensive library of patterns was created by analyzing related concepts in software refactoring and database system domain. There were also interviews with domain experts of in-use ontologies conducted to collect use cases for the evolution of ontologies, from which the most common patterns were extracted. A three level classification was developed, in regard to operations on the level of description logic as highest level, the level of instances for maintaining data quality, or even the lowest level of single entities, containing URI, language or data type's changes. A vocabulary for pattern exchange and increased reusability is also[1] provided.

4 Implementation

OntoWiki[2] is a collaborative semantic data wiki. Since version 0.9 it has a flexible extension system, which perfectly meets the requirements for front- and backend integration of the presented evolution pattern engine. As central element the pattern processing engine is developed as an OntoWiki component. This software offers, besides user interfaces for pattern execution, pattern management functions, too. It's also integrated with the backend, which means function support for SPARQL querying the triple store and adding or deleting statements when patterns get executed. Loading and storing patterns in the respective RDF Format is implemented, too.

5 Related Work

A very short overview for existing work is listed here. A quite similar composition and description of patterns is given in [DA10]. The classification in levels is an extended idea from [AP09]. In [FMK+08] the general classification for ontology operations is evaluated for different papers. The differences and extended requirements for ontology against database schema evolution are discussed in [NK04]. Some patterns were transferred from the book [Fow04] for source code refactorings.

[1] Pattern serialization scheme available under http://ns.ontowiki.net/SysOnt/EvolutionPattern/namespace
[2] Project page: http://code.google.com/p/ontowiki

6 Conclusion and Future Work

With the theory and practice a functional alternative for bulk editing multiple statements at once, converting ontology schemes, or maintaining data quality on RDF knowledge bases is given by our pattern system. The core pattern library has a starting set for most common use-cases and could be easily extended. Real-world implementation integrated with an existing collaborative semantic wiki enables the user to evolve their ontologies in a descriptive way, a lot like the querying of knowledge bases with SPARQL, but extended with various operations. Further investigations which cover the field of pattern management extensions and automatic recognition and execution have to be done.

References

[AP09] Muhammad Javed 0002, Yalemisew M. Abgaz, and Claus Pahl. A Pattern-Based Framework of Change Operators for Ontology Evolution. In Robert Meersman, Pilar Herrero, and Tharam S. Dillon, editors, OTM Workshops, volume 5872 of Lecture Notes in Computer Science, pages 544–553. Springer, 2009.

[DA10] Rim Djedidi and Marie-Aude Aufaure. NTO-EVO an Ontology Evolution Approach Guided by Pattern Modeling and Quality Evaluation. In Sebastian Link and Henri Prade, editors, FoIKS, volume 5956 of Lecture Notes in Computer Science, pages 286-305. Springer, 2010.

[FMK+08] Giorgos Flouris, Dimitris Manakanatas, Haridimos Kondylakis, Dimitris Plexousakis, and Grigoris Antoniou. Ontology change: classification and survey. Knowledge Eng. Review, 23(2):117–152, 2008.

[Fow04] Martin Fowler. Refactoring - improving the design of existing code. Addison-Wesley, 13. print. edition, 2004.

[NK04] Natalya Fridman Noy and Michel C. A. Klein. Ontology Evolution: Not the Same as Schema Evolution. Knowl. Inf. Syst., 6(4):428–440, 2004.

Extracting Knowledge Bases from table-structured Web Resources applied to the semantic based Requirements Engineering Methodology SoftWiki

Rubén Navarro Piris

ruben.navarro.piris@gmail.com

A lot of information on the Web is provided as HTML formatted tables and CSV[1] files. Such tables contain semantic information that can be derived from the embedded environment of the table as well from the heading of each column. Often the problem of integrating and linking this information into semantic web applications occurs. One way to solve this is a transformation of these tables into OWL ontologies.

The requirements engineering tool SoftWiki [DAR06] is an example of such a semantic web application. The SoftWiki methodology [LR09], based on a set of tools including the SoftWiki tool, is used to interpret and manipulate data based on the SoftWiki vocabulary. A common task using the SoftWiki methodology is to import data. The developed OntoWiki extension imports the issue tracker of Google Code[2] projects which is provided as a CSV file and transforms it into an ontology.

The paper describes a methodology that is divided into three parts: (1) Analysis of the table structure, (2) Analysis of the vocabulary and (3) Definition of a N3[3] template. In a typical table the structure is always the same: the columns contain the attributes and the rows the entities. The first row is the one that determines the semantic of the other rows, which contain data that follow this model. Because of its simplicity it is very space efficient, but it's difficult to further analyze or use this information because of the lack of explicit semantic relations.

city	country	Inhabitants	universities
Leipzig	Germany	518,862	Uni Leipzig, HTWK
Barcelona	Spain	1,621,547	UPC, UB, UPF

Table 1: Example of data in a table

We have to take into consideration also that in some case an attribute can be multi-valued, for example the attribute universities of a city[4] (cf. table 1). The table is also a good example of the biggest problem that the semantic data methodology and technologies try to solve: although a person is able to understand the meaning of the

[1] Comma Separated Values
[2] Google Code hosts Open Source Projects and is available at http://code.google.com
[3] Notation 3 (http://www.w3.org/TeamSubmission/n3/), an easy readable RDF/XML syntax alternative.
[4] Source: http://en.wikipedia.org/wiki/Leipzig, http://en.wikipedia.org/wiki/Barcelona

attributes of a city thanks to the context and its own previous knowledge, a computer does not have this information and cannot therefore use it. To solve this problem data can be semantically structured and tagged with the help of semantic ontologies. One of the most common used and simple table formats is CSV and because of that it is the supported format for the data importation. In order to put the new data into the SoftWiki (or OntoWiki) system it has to be adapted to a OWL[5] model using the semantic information of the first row of a table.

As a lot of information that shares the same structure is going to be imported, the definition of a template with an abstract definition of the matching between a row and its corresponding OWL triples is the best way to go. In this template the matching is achieved by using variables with the names of the attributes of the CSV table (or tables) that we want to import. The system is flexible enough to receive tables with attributes that won't be imported or tables with less attributes than expected or with empty attributes but still import all the information possible.

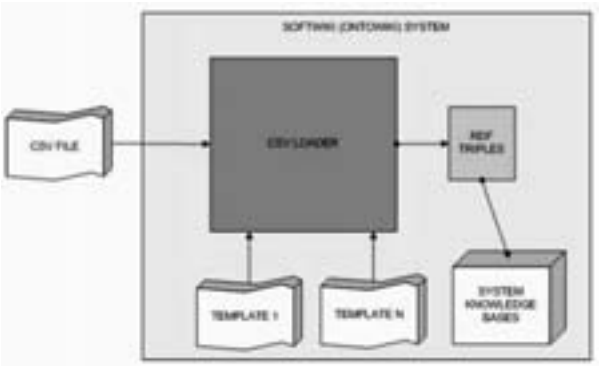


Figure 1: CSV Loader diagram

The importing process (cf. figure 1) would therefore be the following: (1) The CSV Loader application reads the CSV file, (2) According to the information sent by the user, a template is selected and read, (3) RDF triples are created by matching the template with every information row, (4) Finally the triples are stored into the knowledge bases system.

Although the project is focused in the importation of requirements from Google Code Issues to the SoftWiki platform, the application is built in order to be easily configurable. By these means if someone wants to add a template to import other type of data to OntoWiki the following steps are to be followed: (1) Define the template in a .ttlt file and include it in the extension folder, (2) Declare the template in the configuration file, (3) Define the parameters of the template in the configuration file: label, description, url with parameters if used. The application is open source and under development. Its latest version[6] is available online.

[5] Web Ontology Language: http://www.w3.org/TR/owl-features/
[6] http://softwiki.de/netzwerk/de/plattform/csv-load-projekt/

Related work

Convert To RDF[7] *"…is a tool for automatically converting delimited text data into RDF via a simple mapping mechanism."* [GGP+02] Interesting features of this tool are:

- Direct mapping of attributes (columns) to object with properties.
- Possibility of choosing the column that will identify the rdf object and the columns that will be mapped as object attributes, together with its rdf syntax.
- Use of a GUI[8] for defining the mapping. There is also a former version8 that uses predefined mapping files (with either a simple own defined syntax or a rdf syntax, based on a predefined ontology).

There are however some features not supported by ConvertToRDF but in CSVLoad:

- Defining multiple objects from a single row. A row does not necessarily contain information of only one object. Our ontology may, for example, include the City and Country objects. In this case, using the information of table 1, we would want to define one City object and one Country object.
- Support for multi-valued attributes. Using again the information of table 1, our ontology may also contain the University object. In this case the system has to be able of splitting the multi-valued attribute and process every attribute separately.
- Support for missing or empty attributes. CSVLoad works with predefined templates, but does not require that the defined attributes appear the exact way in the input CSV table. If an attribute does not exist in the table (or it is empty) it is automatically omitted of the resulting N3 triples. Also if there are attributes in the table, which are not defined in the template, they are omitted too.
- Data conversion. CSVLoad offers some data converting/processing functions. In this case just by adding the specific tag to an attribute the system will execute an specific routine/conversion to the tagged attribute.

The largest part of information on the Web is already stored in structured form, often as data contained in relational databases, but usually published by Web applications only as HTML mixing structure, layout and content [ADL+09]. The Triplify application[9], born with the idea of overcoming the chicken-and-egg dilemma (simultaneously lacking of semantic representations and semantics-conscious Web search facilities) that delays the expansion of the Semantic Web, permits the conversion of web information (extracted from a relational DB) into RDF, JSON and Linked Data. Interesting features of Triplify tool are:

[7] ConvertToRDF is available under: http://www.mindswap.org/ mhgrove/convert/
[8] Graphical User Interface
[9] http://triplify.org/About

- Easy to install and configure (with few SQL knowledge).
- Already pre-configured mappings to several popular Web applications.
- Focused on deploying the information of a Web into the Semantic Web, fact that provides several advantages: (1) Search engines can better evaluate the content and more easily find content, (2) Possibility of create customized data queries, for example, easy searching for a product with certain characteristics.

Triplify and CSVLoad are focusing on the conversion of table-based data to RDF. However, Triplify works directly on SQL while CSVLoad does it with CSV and template-based mapping, in which case the flexibility is higher because of the possibility of defining different templates for different types of tables. As CSVLoad ignores non defined or empty attributes it is easier to define more general purpose input configuration. Again, Triplify does not support data processing or converting.

Perspectives

Considering that every web application organizes and labels its information in different ways and is usually presented in HTML format, obtaining data from CSV files results insufficient. In this context a plug-in for SoftWiki is under development. This plug-in detects if a loaded requirement resource was imported from Google Code Issues and if positive it imports additional information from Google (community comments with author, date and attachments).

References

[ADL+09] Sören Auer, Sebastian Dietzold, Jens Lehmann, Sebastian Hellmann und David Aumueller. Triplify - Lightweight Linked Data Publication from Relational Databases. In Proceedings of the 17th International Conference on World Wide Web, WWW 2009, Madrid, Spain, April 20-24, 2009, Seiten 621–630, 2009.

[DAR06] Sebastian Dietzold, Sören Auer und Thomas Riechert. Kolloborative Wissensarbeit mit OntoWiki. In Proceedings of the INFORMATIK 2006 Workshop: Bildung von Sozialen Netzwerken in Anwendungen der Social Software, 2006.

[GGP+02] Jennifer Golbeck, Michael Grove, Bijan Parsia, Adtiya Kalyanpur und James Hendler. New Tools for the Semantic Web. In Knowledge Engineering and Knowledge Management: Ontologies and the Semantic Web, Seiten 23–38. 2002.

[LR09] Kim Lauenroth und Thomas Riechert. Der SoftWiki-Ansatz für verteiltes Requirements Engineering mit großen Stakeholdergruppen. In Sören Auer, Kim Lauenroth, Steffen Lohmann und Thomas Riechert, Hrsg., Agiles Requirements Engineering für Softwareprojekte mit einer großen Anzahl verteilter Stakeholder, Leipziger Informatik-Verbund (LIV), 2009.

Unterstützung von Innovationsprozessen mit IT-Werkzeugen – Der InnoRadar als Teil eines Innovationswerkzeugs

Florian Golemo

fgolemo@g-services.net

Abstract: Dieser Beitrag beschreibt ein IT-basiertes Werkzeug zur Unterstützung des im Rahmen des Forschungsprojektes „Systematische Erschließung von Bottom-Up-Innovationen“[1] entwickelten Innovationsprozesses. Der Artikel erläutert in knapper Form den wissenschaftlichen Hintergrund im Projekt und konzentriert sich auf die technische Umsetzung des unterstützenden Werkzeugs. Für die Umsetzung wurden auf Basis von JavaScript & PHP offene Web-Standards benutzt.

1 Motivation

Im Rahmen des Sys-Inno-Forschungsvorhabens entstand ein Vier-Phasen-Innovationsmodell mit dem Ziel die Innovationsfähigkeit insbesondere kleiner und mittlerer Unternehmen zu stärken. Besonderes Merkmal des entwickelten Innovationsmodells ist es, die Innovationskette rückwärts, das heißt „Bottom-Up“, ausgehend von Unternehmensbedürfnissen und alltäglichen Problemstellungen anstatt „Top-Down“ zu betrachten. Diese problemorientierte Betrachtungsweise konzentriert sich auf die Frage „was benötigt das Unternehmen wirklich“, statt von vornherein festzulegen, „wie das Unternehmen handeln sollte“. Die spezifizierte Vorgehensweise bei der Umsetzung von Bottom-Up-Innovationen wird in einem Vier-Phasen-Vorgehensmodell beschrieben. In diesem Modell werden die Abschnitte als Impulsphase, Analyse- und Spezifizierungsphase, Umsetzungsphase und Nachhaltigkeitsbewertung bezeichnet. In der Impulsphase wird die Art der Innovationsanbahnung beschrieben und es erfolgt ein erster Austausch über mögliche Ideen für Innovationen. Bei Neukontakten erfolgt darüber hinaus eine allgemeine Vorstellung der Kompetenzen. In der Analyse- und Spezifizierungsphase erfolgen die Definition der Form der Zusammenarbeit, eine inhaltliche Fixierung der betrachteten Innovation sowie die Definition von Zielen und die Wahl von Schritten zu deren Umsetzung.

[1] Weitere Informationen zu diesem Forschungsvorhaben unter www.sys-inno.de

In der Umsetzungsphase erfolgt die Realisierung der angestrebten Ziele und in der Nachhaltigkeitsbewertung eine Bewertung der betrachteten Bottom-Up-Innovation hinsichtlich der in der Analyse- und Spezifizierungsphase aufgenommenen Ziele. Gleichzeitig können positive Impulse für neue Innovationen oder möglicherweise notwendige Anpassungen vorgenommen werden.

Mit dem Sys-Inno-Tool wird ein webbasiertes Werkzeug zur Unterstützung dieses Prozesses entwickelt, welches eine durchgehende Dokumentation aller Arbeitsschritte des Innovationsmodells ermöglicht. Mit dem „InnoRadar“ wird dabei ein Hilfsmittel zur Visualisierung des Fortschritts sowie der Nachhaltigkeitsbewertung bereitgestellt. Dieses ermöglicht es, einzelne Innovationen entsprechend ihres Arbeitsstandes einander gegenüberzustellen. Der Branchenbezug ermöglicht es, Vorhaben mit ähnlicher Themenstellung zu gruppieren. Auf Basis der gesammelten Informationen über die durchgeführten Innovationsvorhaben kann eine Bewertung durchgeführt werden. Diese Bewertung erfolgt in den Bereichen Zielerreichung, Anwendung, Übertragbarkeit und Kooperation auf Basis von vorher definierten Kriterien. Den Erfolg einer Innovation oder den Fortschritt innerhalb des Innovationsprozesses zu messen und zu bewerten ist für eine systematische Herangehensweise unerlässlich.

2 Konzept

2.1 Konzept InnoRadar

Das Radar bildet die vier Innovationsphasen (wie oben genannt) als konzentrische Halbringe ab, auf denen die Projekte als Punkte eingezeichnet werden. Beginnend auf dem äußersten Ring bewegen sich diese Punkte mit dem Fortschritt des Projektes weiter nach innen. Wenn ein Projekt in der letzten Phase ankommt, erlangt dieses eine automatische Bewertung, welche ein weiterer Bestandteil des Radars ist.

Dieses Layout-Konzept bietet die Basis für zwei Visualisierungen: Den „Gesamtvergleich”, bei dem eine branchenübergreifende Sicht geboten wird und die „Einzelfallanalyse”. In der Ansicht „Gesamtvergleich” wird je ein Unternehmen als Eintrag (Punkt) in der Grafik dargestellt. Bei der „Einzelfallanalyse” kann ein Unternehmen die Hierarchie eines Innovationsprozesses verfolgen. Hierarchie bedeutet hier, dass dargestellt wird, in welchen Phasen, welche Optionen verfolgt wurden, welche dieser Optionen Ergebnisse hervorbrachten, etc.

In Abb. 2 wird das Konzept zum InnoRadar abgebildet, wobei hier die Branchen-Gruppierung angedeutet, die Beschriftung der Phasen direkt in den Halbkreis eingetragen wird und sämtliche Prozesse, zu denen Nachhaltigkeitsbewertungen vorliegen, permanent mit ihrem Bewertungsring versehen sind.

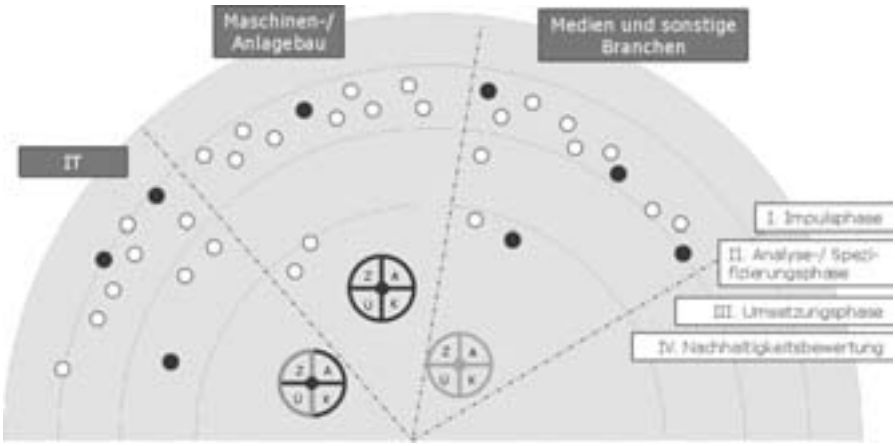


Abbildung 2: InnoRadar „Gesamtvergleich“ mit Nachhaltigkeitsbewertung.

2.2 Konzept Umsetzung

Grundüberlegung für die Umsetzung ist die zu wählende Technologie aus den Möglichkeiten: GD-Library, HTML5 canvas-Element, Adobe Flash, SVG. Dazu wurden Vor- und Nachteile der jeweiligen gegenübergestellt (siehe Abbildung 3).

GD-Library (in PHP)	
Vorteile	**Nachteile**
- liefert eine Grafik, die direkt heruntergeladen und weiterverwendet werden kann	- Grafiken sind statisch, also kann weder die Größe der Grafik automatisch an den Browser angepasst werden, noch können ohne Neuladen der Grafik Elemente hinzugefügt oder gelöscht werden
	- Grafiken sind pixel-, nicht vektorbasiert
HTML5 / canvas-Element (in HTML5/JavaScript) [W3C10]	
Vorteile	**Nachteile**
- im HTML-Standard enthalten / benötigt keine Plugins	- derzeit noch in wenigen Browsern unterstützt und jew. nur in den Nightly Builds
- benötigt keine zusätzliche Bibliothek zum Zeichnen, in das Canvas-Objekt kann direkt mit JavaScript gezeichnet werden	- geht nicht direkt in den DOM-Baum ein, einzelne Elemente können also nicht selektiv neu gezeichnet werden

Adobe Flash (in ActionScript / PHP)	
Vorteile	**Nachteile**
- optischer Mehrwert einfacher realisierbar	- proprietäres Browser-Plugin auf mobilen Geräten umstritten [Can10]
	- mitunter per default aus Sicherheitsgründen blockiert [DR10]
SVG[2] (in JavaScript)	
Vorteile	**Nachteile**
- nativ in allen gängigen Browsern bis auf IE [Sve10], jedoch zahlreiche Plugins vorh.	
- wird direkt im DOM gespeichert, Objekte können also interaktiv und selektiv gezeichnet / manipuliert werden	

Abbildung 3: Gegenüberstellung Kandidaten dynamische Grafik

Aufgrund des Umsetzungzieles fielen folgende Nachteile schwerer ins Gewicht: GD-Grafiken sind statisch, HTML5 kann bisher nur in wenigen Browsern dargestellt werden und Flash wird gelegentlich geblockt. Aus diesen Gründen fiel die Wahl auf die SVG-Technik, für deren Unterstützung eine JavaScript-Bibliothek benötigt wird. Die wesentlichen Unterschiede zum Konzept sind sieben distinkte Bereiche (Ebenen, auf denen Punkte eingetragen werden können) und damit die Vermeidung von wahlloser Anordnung. Des Weiteren wurde ein farblich abgestufter Hintergrund passend zum Farbschema des Sys-Inno-Tools gewählt, die Legende ausgelagert, und sämtliche Beschreibungen sowie die Nachhaltigkeitsbewertung werden erst beim Überfahren mit der Maus dargestellt. Elemente können selektiv verglichen werden, indem der Nutzer auswählen kann, welche Elemente dargestellt werden.

3 Umsetzung

Anhand dieses Konzeptes wurde eine Implementierung vorgenommen, die ein Zusammenspiel aus JavaScript und SVG (Scalable Vector Graphics) auf Client-und PHP auf Serverseite nutzt. Die Clientseite wird unterstützt von den Bibliotheken Raphael[3] (zur SVG-Darstellung) und jQuery[4] (für Effekte / Usability), während auf Serverseite eine eigene weitere PHP-Klasse zum Beantworten von AJAX-Anfragen programmiert wird.

[2] Nach SVG-Spezifizierung 1.1, http://www.w3.org/TR/2003/REC-SVG11-20030114/
[3] Eine Brücke zwischen JavaScript und SVG von raphaeljs.com
[4] Das JavaScript-Framework von jquery.com

Mit dieser Technik kann dem Nutzer ein hohes Maß an Bedienkomfort ermöglicht werden, während gleichzeitig ein Augenmerk auf die Aspekte Sicherheit, Performance und Kompatibilität gelegt wird. Außerdem kann so das Sys-Inno-Tool unabhängig von Geräten und Plattformen genutzt werden.

Bei der Umsetzung wird darauf geachtet, dass die Darstellung zu keinem Zeitpunkt überladen wirkt, weswegen Elemente nur auf einer von sieben Schichten dargestellt werden können (die vier Phasen und drei Zwischenschritte). Die Berechnung hierzu ist eines der Kernelemente des InnoRadar, weil hier die Elemente nicht nur jeweils auf einem Halbkreis positioniert werden müssen, sondern auch auf eine gleichmäßige Verteilung von Elementen entlang des jeweiligen Halbkreises geachtet werden muss. So wird beispielsweise ein einziges Element auf einem Ring bei 90° von insgesamt 180° gezeichnet, zwei Elemente bei 60° und 120°, etc. Wird jedoch nun eine weitere Branche angezeigt, wird der Halbkreis aufgeteilt und jede Branche erhält nur noch 90°, in denen wieder die Elemente analog verteilt werden (siehe dazu Abb. 4).

Das alles passiert dynamisch; beim Laden der Seite wird lediglich der Hintergrund gezeichnet und die AJAX-Daten-Anfrage an den Server ausgelöst. Die Daten werden serverseitig im JSON-Datenformat geschickt und beim Eintreffen am Client aus Sicherheitsgründen mit jQuery und nicht nativ in JS geparst, weil sonst ein eventueller Schadcode vom Server direkt mittels der “eval”-Funktion ausgeführt werden würde. Anschließend werden die Daten zur Auswahl bereitgestellt.

Wenn nun der Anwender einen zu untersuchenden Innovationsprozess aus der Liste ankreuzt, werden dessen Daten (Name, Fortschritt, Branche, etc.) in ein optimiertes Format gebracht und der Liste der darzustellenden Elemente zugeführt und diese Liste wird neu gezeichnet. So wird eine gute Performanz erreicht, durch geringe Speichernutzung und effiziente Asynchronität (auch beim Zeichnen).

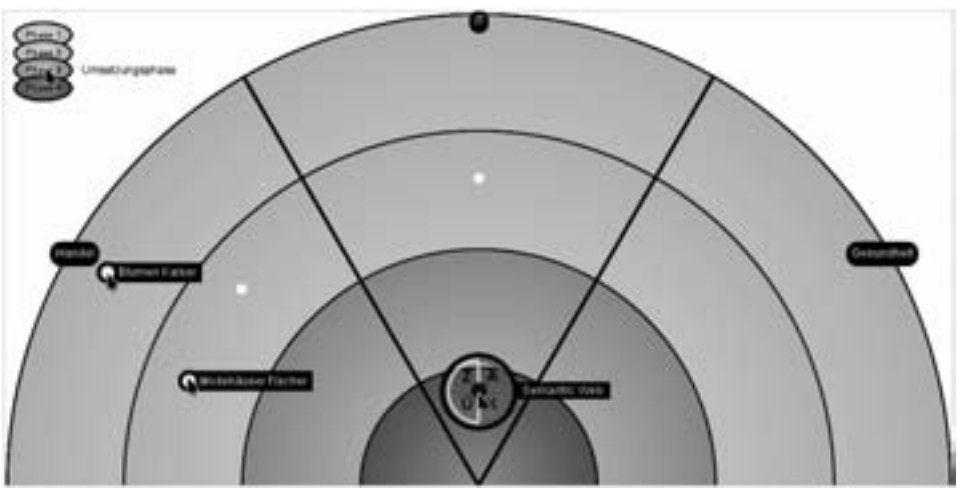


Abbildung 4: vorläufige Umsetzung des Radar mit Nachhaltigkeitsbewertung

Abbildung 4 zeigt den aktuellen Stand der Entwicklung, wobei hier mehrere Bildschirmfotos übereinander gelegt wurden, um darzustellen, wie Interaktivität eingebaut wurde. Innerhalb der Grafik wird die Beschreibung eines Eintrags bzw. die Nachhaltigkeitsbewertung erst angezeigt, wenn man mit der Maus darüber fährt oder ihn anklickt (hier bleibt bis zum erneuten Klick die Beschreibung sichtbar) und einmalig kurz, sobald das Element neu hinzugefügt wird.

4 Ergebnis

Das Projekt ist noch in der Entwicklung und wird intern und von den Projektpartnern getestet[5]. Öffentlicher Zugriff auf das Tool kann daher derzeit noch nicht geboten werden. Nach Abschluss der Entwicklung und weiteren Tests wird das Tool kostenfrei quelloffen zum Download angeboten. Das Ergebnis bisher ist eine intuitive Oberfläche, die dem Benutzer ermöglicht, die Fortschritte der eigenen Innovationsprojekte zu verfolgen, sich leicht gegenüber anderen Innovatoren im gleichen Segment zu positionieren und schlussendlich eine Übersicht zu haben, inwiefern eigene und verwandte Vorhaben erfolgreich waren. Die verwendeten Techniken ermöglichen einem breiten Spektrum von Nutzern den Zugriff, eine interaktive Bedienung ebenso wie auch eine leichtere Darstellung auf mobilen Endgeräten.

In Zukunft könnten solche Visualisierungen und sogar aufwändige Animationen komplett in HTML5 und JavaScript umgesetzt werden, ohne die Notwendigkeit externer Bibliotheken. Dies würde höhere Performanz und Stabilität, geringeren Entwicklungsaufwand und nahezu identische Darstellung auf allen Geräten bedeuten. Damit das passieren kann, muss der HTML5-Standard weiter ausgebaut und in jeden Browser implementiert werden und ein jeder sollte Wert darauf legen, seinen Browser aktuell zu halten.

Literaturverzeichnis

[Can10] www.canada.com: Steve Jobs addresses Adobe's Flash for mobile devices. http://www.canada.com/technology/Steve+Jobs+addresses+Adobe+Flash+mobile+device/2966575/story.html, abgerufen 02.06.2010

[DR10] Dormann, W.; Rafail, J.: Securing Your Web Browser. http://www.us-cert.gov/reading_room/securing_browser/, abgerufen 02.06.2010.

[Sve10] Svensson, P.: Creator of Web spots a flaw in Internet Explorer. http://www.msnbc.msn.com/id/26646919/, abgerufen 01.06.2010.

[W3C10] W3C: HTML Canvas 2D Context. http://www.w3.org/TR/2010/WD-2dcontext-20100304/, abgerufen 01.06.2010.

[Web10] www.webmasterpro.de: Webanalyse: Verbreitung von Webtechnologien und Plugins. http://www.webmasterpro.de/portal/webanalyse-technologien.html, abgerufen 25.05.2010.

[5] Weiterhin wurde das hierfür geschaffene Dokumentations-System des Sys-Inno-Tool schon in über 50 Fällen erfolgreich evaluiert.

Modellierung von touristischen Merkmalen am Anwendungsfall Vakantieland

Marvin Frommhold

marvin.frommhold@gmail.com

Applikationen, welche mannigfaltige Informationsstrukturen verarbeiten, benötigen ein flexibles Datenschema, sowie eine darauf angepasste Abfragesprache. Semantische Web Applikationen, welche derartige Informationsstrukturen verarbeiten, verwenden dazu RDF[1] und SPARQL[2]. Ein Beispiel dafür ist das niederländische Tourismusportal vakantieland.nl, das eine eigens dafür geschaffene eTourismus-Ontologie [Mar08] verwendet. Zu den in RDF gespeicherten Informationen zu touristischen Ressourcen (POI[3]) gehören unter anderen zwei Taxonomien. Eine Taxonomie bildet dabei eine Klassenhierarchie der Typen von POIs ab, die andere eine Hierarchie von touristischen Merkmalen (folgend kurz Merkmale genannt) zu den POIs. Da der Datenbestand aus über 20000 POIs besteht, sind effektive und ohne Vorkenntnis nutzbare Filterfunktionen nötig, um den Benutzern die Möglichkeit zu bieten, aus der Fülle an POIs die für ihn relevanten Instanzen herauszusuchen. Eine effektive Art der stufenweisen Einschränkung stellt dabei das facettierte Durchsuchen [ODD06] dar, welches bereits für die Typenhierarchie der POIs existiert. Jedoch ist aktuell noch keine Funktionalität vorhanden, POIs auch nach deren Merkmalen zu filtern. Hier liegt es auf der Hand, das facettierte Durchsuchen auch für die Merkmale umzusetzen. Die derzeit vorliegende Taxonomie für diese Merkmale ist aber aus Sicht der Ontologie-Entwicklung sowie aus inhaltlichen Aspekten fehlerhaft und weist Inkonsistenzen auf. In dieser Arbeit werden diese Probleme durch einen in [SMMS02] definierten Evolutionsprozess aufgedeckt und durch Anwendung bereits existierender sowie zu definierender Modellierungsmuster behoben. Im Anschluss an diesen Evolutionsschritt wird ein facettiertes Durchsuchen für die Merkmale von POIs in die Vakantieland-Applikation implementiert. Diese Implementierung soll als Evaluation dienen, um zu zeigen, dass aus Sicht des Ontologie-Designs eine generische und fehlerfreie Modellierung der Merkmalsangaben vorliegt.

Ein Ontologie-Evolutionsprozess besteht aus sechs Phasen. Die beiden wichtigsten Phasen in Bezug auf diese Arbeit sind zum einen die Änderungsermittlung sowie die Änderungsdurchführung. In der Phase der Änderungsermittlung wurde die bestehende Taxonomie analysiert und verschiedene Problemfelder identifiziert. Für jedes dieser Problemfelder wurden Muster definiert, wie diese jeweils korrekt zu modellieren sind. Ein Beispiel für eine falsche Modellierung stellte die Angabe der Zugehörigkeit eines POIs zu einer Hotelkette dar. Dieses Merkmal wurde bisher mit einem speziellen Prädikat für die verschiedenen Hotelketten realisiert: *<poiUri> ex:MercureHotel "1"*.

[1] Resource Description Framework - http://www.w3.org/RDF/
[2] http://www.w3.org/TR/rdf-sparql-query/
[3] POI steht für Point Of Interest

Diese Modellierung entspricht aber nicht dem Modellierungsparadigma von RDF. Da es sich bei einer Hotelkette um eine Ressource handelt, für die weitere Aussagen getroffen werden können, sollte dies auch so abgebildet werden. Das definierte Muster zum Modellieren solcher Merkmale sieht nun vor, eine Ressource vom Typ *HotelChain* durch das Prädikat *partOf* mit dem POI zu verbinden. Für das obige Beispiel würde das neue Tripel wie folgt aussehen: *<poiUri> ex:partOf ex:MercureHotel*.

Zusätzlich zu den Mustern wurde ein Fragen-Katalog erstellt, mit dessen Hilfe das zu modellierende Problem durch gezielte Fragestellung analysiert und ein dazu passendes Muster zugewiesen wird. Durch die Anwendung des Katalogs soll die Wartbarkeit der Ontologie verbessert werden, indem neue Merkmale durch Wiederverwendung bestehender Muster in die Datenbasis eingepflegt werden und somit die Konsistenz und einheitliche Speicherung auf Dauer gewährleistet wird. Die letzte Phase im Evolutionsprozess stellt die Änderungsdurchführung dar. Dies wurde im konkreten Fall durch die Implementierung eines Scripts realisiert, welches die eTourismus-Ontologie von dem Zustand vor dieser Arbeit in einen neuen Zustand transformiert, der gegenüber dem alten Zustand alle erforderlichen Änderungen enthält. Das Script nutzt dafür die Funktionalität des Moduls zur Evolution und der Überarbeitung von RDF Graphen [Rie10] der Applikation OntoWiki[4]. Die Änderungen werden in einzelnen Schritten durchgeführt, wodurch der Administrator in der Lage ist, den Änderungsprozess schrittweise nachzuvollziehen und etwaige Anpassungen vorzunehmen.

Nach der Durchführung des Evolutionsprozesses wurde das facettierte Durchsuchen für die Merkmalstaxonomie in die Vakantieland-Applikation integriert. Ähnlich der bereits zuvor bestehenden Funktionalität zur Filterung von POIs nach deren Typ, zum Beispiel die Beschränkung der Suche auf Hotels, ist es nun möglich, die Trefferliste durch Angabe von bestimmten Merkmalen weiter einzugrenzen. Im Bezug auf das eben genannte Beispiel kann der Benutzer nun diese Trefferliste noch weiter einschränken, indem er Merkmale, die nur für Hotels gelten, auswählen kann. Beispielsweise lassen sich so nur die Hotels anzeigen, welche ein Vier-Bett-Zimmer mit TV und Dusche besitzen.

Literatur

[Mar08] Michael Martin. Charakterisierung semantischer Web-Applikationen und Vorstellung eines Anwendungsfalles, 2008.

[ODD06] Eyal Oren, Renaud Delbru und Stefan Decker. Extending faceted navigation for RDF data. In 5th International Semantic Web Conference, Athens, GA, USA, November 5-9, 2006.

[Rie10] Christoph Rieß. Evolution und Refactoring von RDF-Graphen basierend auf Pattern, 2010.

[SMMS02] Ljiljana Stojanovic, Alexander Maedche, Boris Motik und Nenad Stojanovic. User-Driven Ontology Evolution Management. In EKAW '02: Proceedings of the 13th International Conference on Knowledge Engineering and Knowledge Management., Seiten 285–300. Springer-Verlag, 2002.

[4] http://ontowiki.googlecode.com

Entwicklung eines mobilen Social Semantic Web Clients

Natanael Arndt

arndtn@gmx.de

Soziale Netzwerke spielen eine immer größer werdende Rolle in unserem täglichen Leben und die Einrichtung und Pflege unserer Accounts und Daten in diesen Netzwerken wird immer aufwändiger. Die meisten Sozialen Netzwerke sind in sich geschlossen und Nutzer können ihre Daten häufig nicht von einem Netzwerk zu einem anderen Netzwerk kopieren. Das uns bekannte, von Tim Berners-Lee beschriebene, World Wide Web (WWW) kennt diese Art künstlicher Abtrennung einzelner Teilnetze nicht:

> *The idea of a boundless information world in which all items have a reference by which they can be retrieved.* [BLCL+94]

Das Social Semantic Web ist die Idee eines dezentralisierten und Service-unabhängigen Sozialen Netzwerkes, welches auf demWWW und Technologien des Semantic Webs aufbaut. Das Semantic Web ist eine Erweiterung des WWW. Es bietet die Möglichkeit, mit Hilfe des Resource Description Frameworks (RDF, [KC04]), Ressourcen, d. h. Web-Dokumente aber auch nicht-virtuelle Ressourcen wie Personen und Gruppen, miteinander in Beziehung zu setzen. Das Friend of a Friend (FOAF) Projekt entwickelt ein RDF Vokabular [BM04], welches es ermöglicht auf einheitlicheWeise eine Person durch deren Namen, JabberID, Weblog, u. ä. zu beschreiben und deren persönliche Beziehungen untereinander darzustellen. Jeder Person wird dabei eine WebID[1] zugeordnet, welche die Person eindeutig identifiziert. Werden diese Web IDs als Linked Data [BL09] publiziert, entsteht daraus ein dezentralisiertes Social Semantic Web.

Bisher existiert jedoch noch keine Möglichkeit in diesem Sozialen Netzwerk mit mobilen Endgeräten zu navigieren und daran teilzunehmen. Die Idee des mobilen Social Semantic Web Clients ist es daher, Teilnehmern dieses Sozialen Netzwerkes eine Möglichkeit zu geben, mit mobilen Geräten in diesem Netzwerk zu navigieren und es entsprechend zu manipulieren (z. B. neue Bekanntschaften dem eigenen Profil hinzuzufügen). Zu diesem Zweck wurde eine Infrastruktur aus Diensten entworfen und beispielhaft als Content-Provider für die Android-Plattform[2] implementiert.

Der Mobile Semantic Web-Dienst (msw) agiert dabei als Client innerhalb des Semantic Webs, welches er abfragt (z. B. indem er eine Ressource via Linked Data bezieht). Darüber hinaus stellt er seine Daten als RDF-Wissensbasis über eine einheitliche Schnittstelle auf der Plattform zur Verfügung. Zur Verwaltung der Wissensbasis auf der Android-Plattform wird Androjena[3] eingesetzt, welches eine Portierung des Jena

[1] http://esw.w3.org/WebID
[2] http://www.android.com/
[3] http://code.google.com/p/androjena/

SemanticWeb Frameworks [McB02] ist. Auf die Schnittstellen des Mobile Semantic Web-Dienstes können beliebige Anwendungen zugreifen und dabei Ressourcen abfragen und z. B. angeben ob diese dauerhaft oder nur temporär gespeichert werden sollen. Der Content-Provider speichert die Daten aus dem Semantic Web als RDF-Tripel in einem Tripelstore.

Aufbauend auf dem Mobile Semantic Web-Dienst wurde ein weiterer Dienst entwickelt, welcher Aufgaben von Sozialen Netzwerken auf den bereitgestellten Daten durchführt (z. B. die Telefonnummer eines Bekannten erfragt). Dieser Mobile Social Semantic Web-Dienst (mssw) stellt wiederum eine Schnittstelle für Anwendungen zur Verfügung. So können diese Anwendungen das private Profil des Benutzers importieren und darin referenzierte WebIDs anderer Personen als Liste oder einzeln abrufen. Weiterhin besteht die Möglichkeit, über die Kontakte des eigenen Profils hinaus auf Daten der Freunde von Freunden zuzugreifen. Zu diesem Zweck können beliebige Ressourcen aus dem Social Semantic Web abgerufen werden.

Eine Evaluation der entwickelten Dienste wurde durch deren Verwendung in einer Android-Benutzeroberfläche durchgeführt. Diese Anwendung bietet die Möglichkeit, das eigene Profil mit einer Liste der Kontakte zu betrachten, über die Kontaktliste und eine freie Sucheingabe weitere Profile anderer Personen abzurufen, sowie neue Kontakte dem eigenen Profil hinzuzufügen. Darüber hinaus werden die direkten Kontakte des Benutzers in das Adressbuch des mobilen Systems integriert. Dadurch ergibt sich für den Benutzer des Sozialen Netzwerkes der Vorteil, dass er die ihm zur Verfügung gestellten Daten nicht in das System kopieren muss. Stattdessen werden die Daten seiner Bekannten aus dem Social Semantic Web heraus mit seinem Mobiltelefon synchronisiert. Dadurch ist jeder Teilnehmer des Sozialen Netzwerkes nicht nur kontinuierlich auf dem neusten Stand mit seinen Daten, er hat diese auch genau dort wo er sie braucht, in seinem mobilen Endgerät.

Literatur

[BL09] Tim Berners-Lee. Linked Data. Design Issues 18 Jun 2009, W3C, 2009. http://www.w3.org/DesignIssues/LinkedData.html.

[BLCL+94] Tim Berners-Lee, Robert Cailliau, Ari Luotonen, Henrik Frystyk Nielsen, and Arthur Secret. The World-Wide Web. Commun. ACM, 37(8):76–82, 1994.

[BM04] Dan Brickley and Libby Miller. FOAF Vocabulary Specification. Namespace Document 2 Sept 2004, FOAF Project, 2004. http://xmlns.com/foaf/0.1/.

[KC04] Graham Klyne and Jeremy J. Carroll. Resource Description Framework (RDF): Concepts and Abstract Syntax. W3C Recommendation 10 Feb 2004, W3C, 2004. http://www.w3.org/TR/2004/REC-rdf-concepts-20040210/.

[McB02] Brian McBride. Jena: A Semantic Web Toolkit. IEEE Internet Computing, 6:55–59, 2002.

Autorenverzeichnis

Abecker, Andreas, 71, 73
Agarwal, Sudhir, 103
Akinalp, Coskun, 601
Algeier, Vadim, 850
Appelrath, H.-Jürgen, 404
Arndt, Natanael, 1004
Arnold, Oliver, 397
Artinger, Eva, 156, 181
Aschenbruck, Nils, 806
Auer, Sören, 433, 973

Bachmann, Erik, 660
Bamman, David, 542
Battistello, Giulia, 796
Bauckhage, Christian, 441
Bauer, Thomas, 357
Baum, Marcus, 904
Baumheier, Konrad, 983
Becker, Tobias, 168
Becker, Sven, 552
Beier, Georg, 397
Beine, Gerrit, 397
Belli, Fevzi, 319
Belter, Roman, 97
Bente, Stefan, 221
Benzina, Amal, 181
Beyazıt, Mutlu, 327
Birkhäuser, Benedikt, 143
Blaser, Rainer, 739
Bloehdorn, Stephan, 71
Bolles, André, 404
Bombosch, Uwe, 221
Boltes, Maik, 162
Bornes Varol, Marie-Christine, 559
Borowski, Marion, 552
Breitländer, Markus, 26
Breitner, Michael H., 615
Brekle, Jonas, 979
Brinkmann, Kevin, 812
Broeder, Dan, 530
Brückner, Markus, 175
Büchler, Marco, 523, 524, 572
Buckl, Sabine, 279
Budnik, Christof J., 319
Bunzel, Stefan, 503
Burlaga, Roland, 61

Chen, Willy, 115
Cichos, Harald, 345
Conrath, Markus, 497
Cordes, Jürgen, 503
Coskun, Gökhan, 79
Coskun, Tayfur, 156, 181
Crane, Gregory, 542
Cremers, Daniel, 874
Crestani, Marcus, 921

Dambreville, Frederic, 880
Danielsson, Lars, 913
Danowski, Kamen, 137
Dästner, Kaeye, 862
Daun, Martina, 892
Deistler, Timon Sebastian, 327
Delord, Bertrand, 503
Demuth, Michaela, 868
Dezert, Jean, 886
Diedrich, Christian, 333
Dietze, Stefan, 453
Diewald, Nils, 586
Donner, Anton, 192
Duan, Linshu, 339

Eckert, Claudia, 746
Eckert, Julian, 109, 299
Eggert, Sandy, 389

Ehlers, Frank, 824
Engel, Ulrich, 800
Engelland, Holger, 38
Erfurth, Christian, 135
Eren, Evren, 26
Eschbach, Robert, 357

Feldmann, Michael, 904
Fey, Dietmar, 126
Finsterbusch, Andrej, 711
Flieder, Karl, 383
Fränken, Dietrich, 904
Frielingsdorf, Bernd, 503
Frommhold, Marvin, 1002
Funk, Burkhardt, 647
Fürst, Simon, 503
Füßl, Wilhelm, 464

Geesen, Dennis, 404
Gehrke, Nick, 424
Gemlik, Arkadius, 615
Gerber, Daniel, 987
Geske, Ulrich, 691, 711
Giesen, Nils, 681
Gilberg, Alain, 503
Gleim, Rüdiger, 586
Gnasa, Melanie, 552
Gökçe, Nida, 327
Goeken, Matthias, 239, 305
Golemo, Florian, 996
Golubski, Wolfgang, 397
Goltz, Hans-Joachim, 693
Gómez, Jorge Marx, 681
Gorriz, Michael, 3
Govaers, Felix, 806
Gradinger, Reiner, 739
Grawunder, Marco, 404
Greiner-Mai, Thomas, 192
Grimm, Stephan, 71
Gronoff, Jean-Daniel, 559
Grossmann, Jürgen, 491
Güldali, Baris, 369
Gustafsson, Fredrik, 831

Haak, Liane, 681
Haase, Peter, 726
Habberstad, Hans, 831
Hafemann, Ingelore, 566
Hahn, Axel, 639
Hamp, Quirin, 150
Hanebeck, Uwe D., 904
Hashagen, Ulf, 468
Haubold, Tobias, 397
Häusler, Stefan, 639
Hecker, Thomas, 491
Hedenetz, Bernd, 497
Heidary, Khosrau, 503
Heinz, Joachim, 273
Heise, David, 633
Helbig, Johannes, 3
Hellige, Hans Dieter, 461, 466
Hess, Claudia, 115
Hesselmann, Margret, 221
Heute, Ulrich, 818
Heyer, Gerhard, 523, 524, 572
Hinneburg, Alexander, 758
Hinnerichs, Andreas, 491
Hinrichs, Erhard, 530
Hippe, Klaus, 720
Hirano, Hiroyuki, 503
Hoffmann, Clemens, 983
Höfer, Alexander, 339
Hofestädt, Ralf, 719, 720
Hofstedt, Petra, 699
Hogrebe, Frank, 418
Holl, Stefan, 162
Hollink, Laura, 447
Hollmann, Axel, 319
Hommel, Wolfgang, 265
Hörst, Julian, 844
Hoßfeld, Friedel, 127
Hurka, Jörg, 812

Hussmann, Heinrich, 339

Jacobi, Jonas, 404
Jaensch, Martin, 497
Janning, Ruth, 85
Janowski, Sebastian, 720
Jarke, Matthias, 10
Johannsen, Wolfgang, 311
Joswig, Jürgen, 19

Kacel, Rick, 503
Kahler, Helge, 557
Kania, Renate, 752
Karampiperis, Pythagoras, 143
Kaune, Regina, 838
Keil, Patrick, 201
Kirschke-Biller, Frank, 503
Klauer, Bernd, 123
Kleiner, Frank, 73
Klinker, Gudrun, 156, 181
Klopstock, Thomas, 746
Kluge, Rolf, 97
Knittl, Silvia, 265
Koch, Rainer, 168
Koch, Wolfgang, 796, 850, 874, 904
Kogel, Karl-Heinz, 874
Köhler, Bastian, 862
Kohlmayer, Florian M., 746
König-Ries, Brigitta, 135
Konle, Wolfgang, 785
Koppe, Roland, 639
Kormeier, Benjamin, 720
Krause, Jan, 333
Kratzsch, Claudia, 485
Krcmar, Helmut, 229, 259
Krehahn, Peter, 673
Kremers, Horst, 174
Kreuz, Ingo, 509
Krosta-Hartl, Pamela, 61
Kühn, Christian, 503
Kuhn, Klaus A., 719, 739, 746
Kuhrmann, Marco, 201
Kundisch, Dennis, 621
Kunkel, Bernd, 503
Kütz, Martin, 613

Lackner, Hartmut, 321
Lajtkep, Andre, 503
Lamla, Gregor, 739
Lautenschläger, Ronald R., 746
Lewitz, Olaf, 209
Lindgren, David, 831
Linssen, Oliver, 201
Löbe, Matthias, 734
Loke, Ben, 143
Loos, Andreas, 479
Looso, Stefanie, 285
Lucke, Ulrike, 91
Ludwig, André, 97
Luszak-Rösch, Markus, 79

Mähler, M, 187
Mahr, Bernd, 130
Martens, Alke, 91
Marterer, Robin, 143
Martignoni, Robert, 627
Mathäß, Tobias, 726
Matthes, Florian, 279
Mehler, Alexander, 586
Meier, Christian, 621
Meitinger, Thomas, 746
Mertens, Michael, 767
Michels, Carolin, 103
Mihaylova, Lyudmila, 907
Milicevic, Danijel, 305
Mittag, Kristin, 758
Mitjagin, Marina, 983
Mönch, Lars, 599
Mössinger, Jürgen, 503
Müller, Bernd, 410
Müller, Reinhard, 868
Müller, Wolfgang, 719, 752

Müller-Birn, Claudia, 433
Müller-Glaser, Klaus D., 521

Navarro Piris, Rubén, 992
Nestler, Simon, 156, 181
Neumann, Carsten, 491
Neumann, Markus, 615
Nicklas, Daniela, 404
Niemann, Michael, 299
Niemeyer, Peter, 647
Niewiejska, Julia, 806
Niggemann, Stephan, 649
Niittylä, Laura, 143
Nishikawa, Kenji, 503
Nitzschke, Marcus, 983
Nolte, Siegfried, 957, 964
Nörenberg, Ralf, 515
Nüttgens, Markus, 424

Oispuu, Marc, 844
Opitz, Felix, 862
Orthuber, Wolfgang, 453
Ortola, Marie Sol, 559

Painsky, Amichai, 773
Pannetier, Benjamin, 886
Paschke, Adrian, 79
Pauli, Manfred, 259
Pellengahr, Alexander, 231
Peris, Martina, 424
Petrasch, Roland, 950
Petzold, Hartmut, 471
Pichlmaier, Peter, 156
Pinte, Florian, 375
Pottebaum, Jens, 168
Prasser, Fabian, 739
Prokisch, Holger, 746

Racz, Nicolas, 245
Radusch, Ilja, 491
Ramacher, René, 599
Rausch, Jan, 975
Rechl, Hans, 739
Rechner, Horst, 491
Reich, Johannes, 597, 603
Reindl, Leonhard, 150
Reissing, Ralf, 515
Repp, Nicolas, 277
Ribe-Baumann, Liz, 175
Rick, Martin, 898
Riechert, Thomas, 973
Riedl, Reinhard, 253
Rieß, Christoph, 989
Rimkus, Robert, 503
Rojas, Isabel, 752
Romahn, Andreas, 215
Ronneberger, Torsten, 485
Roth, Sascha, 279
Rothe, Mario, 79, 711
Runte, Wolfgang, 610
Rusnjak, Andreas, 604

Saglietti, Francesca, 319, 375
Scheible, Jan, 509
Schermann, Michael, 229, 259
Schikora, Adam, 874
Schikora, Marek, 874
Schlangen, Isabel, 892
Schlieder, Christoph, 580
Schmerler, Stefan, 503
Schmidt, Benjamin, 32
Schnackenburg, André, 201
Schneider, Jörn, 477
Schneider, Steffen, 168
Schnelte, Matthias, 369
Schöne, Thomas, 983
Schranz, Astrid, 418
Schreckenberg, Michael, 125
Schröder-Preikschat, Wolfgang, 363
Schrödl, Holger, 203
Schulte, Stefan, 109, 277
Schulz, Arne, 818
Schulz, Christopher, 279

Schulz, Marcel, 604
Schumann, Sandy, 187
Schürr, Andy,345
Schweda, Christian M., 279
Schwenniger, Jochen, 143
Schwering, Benjamin, 231
Seget, Kathrin, 818
Seifert, Gero, 221
Seising, Rudolf, 463
Seufert, Andreas, 245
Seyfried, Armin, 162
Siebenhaar, Melanie, 109
Sommer, Martin, 291
Sorg, Michael, 43
Speck, Andreas, 597, 604
Sperber, Michael, 921
Sprenger, Jon, 615
Staab, Steffen, 433
Stalmann, Kai, 552
Stäubert, Sebastian, 734
Steffens, Ulrike, 277
Steinmetz, Ralf, 109, 299
Stimmer, Jörg, 627
Strecker, Stefan, 613
Streit, Roy L., 792
Strohschneider, Stefan, 187
Sum, Richard, 479
Svensson, Daniel, 913
Syldatke, Thomas, 115

Thiel, Sebastian, 351
Thomä, Reiner, 850
Thum, Daniela, 485
Thurau, Christian, 441
Titze, Andreas, 503
Tolksdorf, Robert, 79, 435
Töpel, Thoralf, 720
Troegner, Dagi, 705
Türker, Can, 719

Ulmke, Martin, 767, 913
Unger, Herwig, 601

Van Assem, Mark, 447
Van Kempen, Beate, 418
Vierig, Enrico, 985
Vollmar, Roland, 474
Vorley, Dave, 907

Waldschmidt, Klaus, 473
Walser, Konrad, 253
Waltinger, Ulli, 586
Watson, Alexander, 907
Weber, Jörg, 515
Weißleder, Stephan, 321
Weippl, Edgar, 245
Wendland Karsten, 273
Werner, Mario, 49
Weßel, Christa, 927, 938
Wieland, Sabine, 55
Wieneke, Monika, 856
Wind, Stefan, 203
Winter, Mario, 319
Wirmer, Lucas, 424
Wischermann, Dirk, 363
Wittenburg, Peter, 530
Wittig, Ulrike, 752
Wohlgemuth, Volker, 673
Wolf, Armin, 691, 711
Wolf, Frank, 273
Wolf, Patrick, 239
Wrobel, Stefan, 552
Wucholt, F, 187
Wullinger, Peter, 580
Wurst, Sebastian H. R. 739, 746

Yildirim-Krannig, Yeliz, 187

Zastrow, Thomas, 530
Ziegenbein, Dirk, 477
Ziep, Tobias, 673
Zitterell, Dirk, 351
Zuse, Horst, 124, 470

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 Language 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 Busi-ness 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 (Ergänzungsband).

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 – Erfahrungen 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

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 Anwendungsentwick-lung 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^2CS 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-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

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