

Gesellschaft für Informatik e.V. (GI)

publishes this series in order to make available to a broad public recent findings in informatics (i.e. computer science and information systems), to document conferences that are organized in cooperation with GI and to publish the annual GI Award dissertation.

Broken down into

- seminars
- proceedings
- dissertations
- thematics

current topics are dealt with from the vantage point of research and development, teaching and further training in theory and practice. The Editorial Committee uses an intensive review process in order to ensure high quality contributions.

The volumes are published in German or English.

Information: <http://www.gi.de/service/publikationen/lni/>

ISSN 1614-3213

ISBN 978-3-88579-445-5



Gesellschaft für Informatik (Hrsg.): Informatiktag 2012

GI-Edition

Lecture Notes in Informatics

Gesellschaft für Informatik (Hrsg.)

Informatiktag 2012

Fachwissenschaftlicher Informatik-Kongress

23. und 24. März 2012

B-IT Bonn-Aachen International Center
for Information Technology, Bonn





Gesellschaft für Informatik (Hrsg.)

Informatiktage 2012

Fachwissenschaftlicher Informatik-Kongress

23. und 24. März 2012

**B-IT Bonn-Aachen International Center for Information
Technology in Bonn**

Gesellschaft für Informatik e.V. (GI)

Lecture Notes in Informatics (LNI) - Seminars

Series of the Gesellschaft für Informatik (GI)

Volume S-11

ISSN1614-3213

ISBN 978-3-88579-445-5

Volume Editor

Gesellschaft für Informatik e.V.

Ahrstraße 45

53175 Bonn

E-Mail: gs@gi.de

Redaktion: Ludger Porada

E-Mail: ludger.porada@gi.de

Series Editorial Board

Heinrich C. Mayr, Alpen-Adria-Universität Klagenfurt, Austria

(Chairman, mayr@ifit.uni-klu.ac.at)

Dieter Fellner, Technische Universität Darmstadt, Germany

Ulrich Flegel, Hochschule für Technik, Stuttgart

Ulrich Frank, Universität Duisburg-Essen, Germany

Johann-Christoph Freytag, Humboldt-Universität zu Berlin, Germany

Michael Goedicke, Universität Duisburg-Essen, Germany

Ralf Hofestädt, Universität Bielefeld, Germany

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

Ingo Timm, Universität Trier

Karin Vosseberg, Hochschule Bremerhaven, Germany

Maria Wimmer, Universität Koblenz-Landau, Germany

Dissertations

Steffen Hölldobler, Technische Universität Dresden, Germany

Seminars

Reinhard Wilhelm, Universität des Saarlandes, Germany

Thematics

Andreas Oberweis, Karlsruher Institut für Technologie (KIT), Germany

© Gesellschaft für Informatik, Bonn 2012

printed by Köllen Druck+Verlag GmbH, Bonn

Wissenschaftliche Tagungsleitung

Alfred Zimmermann, Hochschule Reutlingen

Programmkomitee

Hermann Engesser, Informatik Spektrum

Ulrich Furbach, Universität Koblenz

Rul Gunzenhäuser, Universität Stuttgart

Walter Hower, Hochschule Albstadt-Sigmaringen

Otto Spaniol, RWTH Aachen

Gottfried Vossen, Universität Münster

Alfred Zimmermann, Hochschule Reutlingen

Workshops - Themen und Partner

Mobile Solutions in der industriellen Praxis

Capgemini

Social Media Security & Risk Management

DETECON

Enterprise Architecture Management - Grundlagen und Best-practices

iteratec Gesellschaft für iterative Softwaretechnologien mbH

Social Media – Nichts für Unternehmen?

msg systems ag

Moderne Software und eine Versicherung - (Wie) Passt das zusammen?

PPI Financial Systems

Übersicht über das Rahmenprogramm und Partner der Informatiktage 2012

Neben den Workshops wurde den Teilnehmerinnen und Teilnehmern ein attraktives Rahmenprogramm geboten. Wir bedanken uns bei unseren Partnern, ohne deren Unterstützung die Informatiktage nicht hätten stattfinden können.

Rahmenprogramm:

Wettbewerbspräsentationen des informatiCup 2012

Wissenschafts- und Absolventen-Workshop

Leitung: Prof. Dr. Alfred Zimmermann, Fachhochschule Reutlingen

Workshop-Spezial: Software-Ingenieure als Unternehmer

Leitung: Prof. Dr. Ernst Denert, sd&m-Gründer

Hauptvortrag:

Prof. Dr. Josef Rammig, Universität Paderborn

Cyber-Physical-Systems

Talk-Show des Computer Clubs 2

Postersessions

Workshops in drei Sessions

Moderation der Veranstaltung:

Prof. Dr. Alfred Zimmermann, Reutlingen University

Partner und Sponsoren der Informatiktage:

Partner:

B-IT Bonn-Aachen Center

Sponsoren:

Capgemini

DETECON

iteratec Gesellschaft für iterative Softwaretechnologien mbH

msg systems

PPI Financial Systems

Medienpartner:

IX

Computer Club 2

Staufenbiel Media GmbH

Vorwort zum Tagungsband 2012

Ich begrüße Sie herzlich zu den zwölften Informatiktagen, die in diesem Jahr unter dem Motto „Wirtschaft trifft Wissenschaft“ stehen.

Wirtschaft und Wissenschaft – das sind zwei Partner, die wir bei den Informatiktagen zusammenbringen, um aktuelle Trends zu diskutieren, sich gegenseitig zu befruchten und auch gemeinsam in die Zukunft zu blicken.

Fachleute sprechen bereits jetzt von der „vierten industriellen Revolution“, die sich mit Themen wie dem Internet der Dinge und Cyber-Physical-Systems anbahnt; und auch die Bundesregierung weiß um die Bedeutung dieser Revolution und hat im Rahmen ihrer High-Tech-Strategie ein ressortübergreifendes Zukunftsprojekt „Industrie 4.0“ Förderprogramm aufgelegt.

Wir beleuchten diese Themen sowohl in unserem Hauptvortrag als auch in der Talk-Runde des Computer Clubs 2 und erhoffen uns einen lebendigen Austausch zwischen Wissenschaft und Wirtschaft.

Zu den Informatiktagen selbst: Die Informatiktage der Gesellschaft für Informatik bieten alljährlich einigen der besten Studierenden Gelegenheit, ihre fachlichen Ideen und ihre Zukunftspläne mit Vertretern der Wirtschaft und erfahrenen Wissenschaftlern zu diskutieren. Hier werden Kontakte nicht nur für eine Karriere an der Hochschule, sondern auch zu den IT-Unternehmen, die im Rahmen der Informatiktage spannende Workshops anbieten, geknüpft. Ohne die Beteiligung von engagierten Unternehmen wären die Informatiktage nicht möglich. Ihnen gilt also mein besonderer Dank!

Allen Teilnehmerinnen und Teilnehmern wünsche ich eine interessante und erfolgreiche Veranstaltung.



Prof. Oliver Günther, Ph.D. (GI-Präsident)

Inhaltsverzeichnis

Grundlagen der Informatik

Christian Abele, Michael Schaidnager

Beschreibung eines Empfehlungsalgorithmus basierend auf Vektordifferenzmengen ohne Regelerstellung 15

Johannes Hofmann

Leistungsanalyse Evolutionärer Algorithmen auf GPUs im Vergleich zu Multikern-CPU's 19

Patrick Kabasci

Automated Program Verification Using SAT- and SMT-Encodings of Diophantine Constraints 23

Martin Lang

Erreichbarkeit mit Ressourcenschranken für Pushdownsysteme 27

Andreas Platau

Schwarmintelligenz für Handlungsreisende 31

Joachim Redies

An extension of the GiNaCRA library for the Cylindrical Algebraic Decomposition.. 35

Künstliche Intelligenz

Marcel Ackermann

Distant Supervised Relation Extraction 39

Oliver Dunkley, Bill Okal

Development of a Cheap Chess Robot: Planning and Perception 43

Dominik Grzelak

Automatisierte Erkennung von Influencern im Social Web 47

Felix Kerger

Implementing an AI for Mayagarden, a strategic 2 player game 51

Christian A. Müller

3D Objekt Kategorisierung in häuslichen Umgebungen..... 55

Jan-Thorsten Peter

Soft String-to-Dependency Hierarchical Machine Translation 59

Johannes Schwandke

Bestimmung von semantisch ähnlichen Worten mit Hilfe von Kookkurrenzen und Wortstämmen 63

Anke Tallig

SOLVE–Geschlechterspezifisches Problemlösungsverhalten 67

Softwaretechnik

Alexander Brömmer, Marius Fink, Till Schomborg

Simbac – Ein Analysetool zur Simulation von Bachelorsabschlussquoten 71

Christian Cikryt

Entwicklung eines Konfigurationsvalidierers 75

Mathias Gebbe

Migration eines web-basierten Projektmanagement- und Kollaborations-Portals 79

Helge Holzmann

Videounterstützte Ablaufverfolgung von Tests für Anwendungen mit grafischer Benutzeroberfläche 83

Sebastian Kleinschmager

Eine empirische Studie mit Java und Groovy über den Einfluss von statischen Typsystemen auf Entwicklereffizienz bei der Benutzung und Anpassung von Softwaresystemen 87

Christian Lück

Konzeption und Realisierung eines Frameworks für transparente RPCs in PHP 91

Benjamin Saul

Entwicklung und Untersuchung einer Softwarearchitektur für medizinische Anwendungen am Beispiel von Hüftgelenken 95

Markus Towara

Numerische Optimierung eines Ölansaugkanals mithilfe adjungierter Topologieoptimierung 99

Mensch-Computer-Interaktion

Alexander Gottschall

Durchführung einer Usability-Evaluation am Beispiel des Warenwirtschaftssystems "CounTable" 103

Malte Jannasch, Dennis Ludl, Claudia Vöhringer

Low-Cost NUI 107

Stefanie Müller, Joan Tomás Pape
Entwicklung eines iPhone Uni-Navigators 111

René Reiners
A Pattern Evolution Process – From Ideas to Patterns 115

Datenbanken und Informationssysteme

Nadine Hellweg
Informationsqualität in Unternehmenswikis – Eine internationale empirische Studie zu Kriterien der Informationsqualität 119

Moritz Kaufmann, Matthias Ringwald
Porting a Functional Query Language to the Parallel Data Flow Framework Stratosphere..... 123

Yvonne Mülle
Outlier Mining Based on Subspace Analysis in High Dimensional Databases 127

Andreas M. Wahl
alpha-OffSync: An Offline-Capable Synchronization Approach for Distributed Document-Oriented Process Management in Healthcare..... 131

Technische Informatik

Jan Barczak, Simon Hardt, Jens Schmidt, Kai-Uwe Müller, Björn Brachthäuser, Matthias. Mielke,
Ein evolutionärer Algorithmus auf einem FPGA zur gleichmäßigen Netzauslastung 135

Benjamin Erb
Concurrent Programming for Scalable Web Architectures 139

Kevin Ernst
Realisierung eines VoIP-Testnetzwerkes mit einem Network Impairment Generator .. 143

Olaf Gaydoss, Stephan Schaible, Michael Massoth
Smart Energy Grids: Ein interdisziplinäres Forschungsvorhaben an der Hochschule Darmstadt 147

André Himmighofen
Hardware Concepts for Elliptic Curvebased Zero-Knowledge Protocols 151

Alexander Kohout
Constraint Modeling and Matching in a Dynamic Context Service Discovery and Binding Process 155

Rainer Müller

Implementierung eines interrupt-gesteuerten OSEK-Betriebssystemkerns auf einem ARM-Cortex-M3-Mikrocontroller..... 159

Marina Rodríguez

Mental Stress Detection Using Multimodal Sensing in a Wireless Body Area Network 163

Simon Seif

Implementierung und Evaluierung eines verteilten Algorithmus für die Kanalzuweisung in drahtlosen Maschennetzen..... 167

Informatik in den Lebenswissenschaften

Matthias Bremser

Continuous Diameter Measurement of Vascular Structures in Ultrasound Video Sequences 171

Hajk-Georg Drost

Development of a phylogenetic transcriptome atlas of Arabidopsis thaliana embryo development 175

Alexander Gabel

A phylostratigraphic analysis of the Arabidopsis thaliana genome 179

Graphische Datenverarbeitung

Sergey Cheremukhin

Vereinfachte 3D-Rekonstruktion aus Sequenzen unkalibrierter Bilder..... 183

André Kunz, Rakulan Subramaniam

Automatische Pflanzenhöhenermittlung mittels Kinect..... 187

David A.Plecher

Klassifizierung von Anzeigeprinzipien in AR-Anwendungen anhand von definierten Paradigmata 191

Wirtschaftsinformatik

Sebastian Bourscheidt, Thomas Breuer, Tobias Brunner, Ben Fetler, Georg Fogel

ESARC - Referenzmodell und Ontologie für Enterprise Architecture Management 195

Steffen Brandt, Torsten Wiens, Michael Massoth

Marktanalyse einer Authentication-as-a-Service-Lösung bei Telefonaten mittels des neuen Personalausweises 199

Tobias Brunner

Ein patternbasiertes Vorgehensmodell für den Enterprise Services Architecture Reference Cube (ESARC) 203

Michael Falkenthal, Dierk Jugel

Ontologie zur Repräsentation und Auswertung von Problemdiagnose- und Problemlösungswissen im Kontext eines semantikerweiterten Kennzahlendashboards 207

Markus Fortak

Effiziente Nutzung von Cross-Selling im Bankumfeld 211

Daniel Gaebler, Bastian Maucher

Pattern-Katalog zur Etablierung von Vertrauen in Cloud Computing Technologien aus Sicht der Nutzer 215

Sebastian Grunow, Matthias Lenk, Sascha Roth

KPIs für die strategische Business/IT Alignment-Quantifizierung 219

Richard Günther

Integration ereignisbasierter Middleware-Systeme in kontextbasierte Geschäftsprozesse 223

Daniel Roppelt

Analyse der Fertigungstiefe bei der Fachkonzeptmodellierung im Umfeld der Billiganwendung Abili-T 227

Daniel Schlich

Geschäftsprozessmodellierung – automatisches Erzeugen von Benutzerschnittstellen mit der BPMN 2.0 231

Benjamin Westphal

Service Governance Patterns auf Basis des Open Group Referenzmodells 235

Sicherheit – Schutz und Zuverlässigkeit

Steffen Brandt, Torsten Wiens, Michael Massoth

Gegenseitige Teilnehmerauthentifizierung in Telefonaten mit dem neuen Personalausweis 239

Dennis Guse, Benjamin Müller

Gesture-based User Authentication for Mobile Devices using Accelerometer and Gyroscope 243

Chris Stein

Entwicklung eines Authentifikationssystems mit Fingerfotos von Smartphone-Kameras 247

Beschreibung eines Empfehlungsalgorithmus basierend auf Vektordifferenzmengen ohne Regelerstellung

Michael Schaidnagel, Christian Abele

Hochschule Reutlingen

Fakultät Informatik

Michael.Schaidnagel@Student.Reutlingen-University.de

Christian.Abele@Student.Reutlingen-University.de

Art der Arbeit: Projektarbeit

Betreuer/in der Arbeit: Prof. Dr. rer nat. Friedrich Laux

Abstract: Die Ableitung von Kaufempfehlungen aus Warenkorbanalysen ist eine typische Data Mining Aufgabe [WI05]. Hierfür existieren Algorithmen, wie beispielsweise der Apriori Algorithmus [Yo10] [AIS93], bei dem auf Basis von Regeln weitere Produktempfehlungen generiert werden. Diese Regelerstellung und Anwendung verursacht bei steigender Komplexität Speicher- und Rechenaufwand. Das hier vorgestellte Verfahren verzichtet auf die aufwendige Regelerstellung und liefert dabei vergleichbare, bis bessere Ergebnisse.

1 Einleitung und Datengrundlage

Der hier beschriebene Algorithmus (VDP) wird auf die Aufgabenstellung des Data Mining Cup 2011 (<http://www.data-mining-cup.de/>) angewendet. Hierbei sollen auf Basis von Warenkörben vergangener Transaktionen (die sog. Trainingsdaten) drei weitere Produkte (Items) vorhergesagt werden, die zu aktuellen Warenkörben (die sog. Testdaten) hinzukommen. Sowohl die Trainingsdaten B als auch die Testdaten X wurden in eine MySQL-Datenbank überspielt. Dies ermöglichte flexible, interaktive Datenbankanalysen mit SQL und eine vereinfachte Programmierung bei der Parallelisierung des Algorithmus.

2 Mathematische Beschreibung und Pseudocode

Der VDP erstellt keine Regeln aus der Trainingsmenge, sondern ermittelt für jeweils einen (Teil-) Warenkorb $x := \{x_1, x_2, \dots, x_m\}$ (Prämisse, Präfix) eine vorgegebene Anzahl k [$k \geq 1$] weitere Produkte y_1, y_1, \dots, y_k (Konklusion, Postfix), die zu diesem gehören bzw. als nächstes in den Warenkorb gelegt werden. Die vorhergesagten Produkte y_k werden aus einer Menge von Warenkörben B ermittelt, die Ähnlichkeit mit dem Warenkorb x besitzen. Dabei erfolgt die Auswahl der y_k nach der Häufigkeit ihres Auftretens in B . Die Ähnlichkeit zweier Warenkörbe wird durch den Quotienten a der gemeinsamen Items, zur Größe des gemeinsamen Warenkorbes bestimmt.

2.1 Mathematische Beschreibung des VDP Algorithmus

X sei die zu prognostizierende Warenkorbmenge,
 B sei die Trainingsmenge der Warenkörbe,
 P sei die Menge der Produkte in B , N_0 natürliche Zahlen inkl. 0,
 H sei die Hash-Map, mit Produkte b_i und Häufigkeit $cnt(b_i)$,
 a sei der Ähnlichkeitsfaktor mit dem die Mindestübereinstimmung errechnet wird,
 und k sei die Anzahl der zu prognostizierenden Produkte pro Warenkorb $x \in X$.

$\forall x \in X$

$H := \{(b_i, cnt(b_i)) \mid b_i \in b \in B\}$

$\forall b := \{(b_1, b_2, \dots, b_n) \in B:$

Wenn $(a \frac{|x \cap b|}{|x \cup b|} \leq |x \cap b| < n)$ dann $\forall b_i \in (b \setminus x) : cnt(b_i) + 1$

Top-k-items: $Top_k(H, (cnt(y), >))$

2.2 Pseudocode

```

(1) For each item  $x$  of  $X$  do {
(2)   create and initialize Hashmap  $H := \{(b_i, count(b_i))\}$ 
      for all items  $b_i$  found in  $B$  set  $count(b_i) := 0$ ;
(3)   For  $i = 1$  to #trainBaskets do {
(4)      $b := B[i]$ ;
(5)     if  $a \frac{length(x \cap b)}{length(x \cup b)} \leq length(x \cap b)$  and  $length(x \cap b) < \#items\_b$  then {
(6)        $p := b \setminus x$ ;
(7)       for  $j = 1$  to #items_p do  $count(p_j)++$ ;
(8)     }
(9)   }
(10)  for  $m = 1$  to  $k$  do { // Top-k Rank for H
(11)    if  $maxCount(H) = 0$  then skip loop
(12)     $Y[m] := maxCount(H)$ ; // complexity =  $k \mid H \mid$ 
(13)    Remove  $Y[m]$  from  $H$ ;
(14)  }
(15)  return  $Y$ ; // max k items
(16) }
```

2.3 Verbale Beschreibung des VDP

Der VDP lädt sich im ersten Schritt die Trainingsdaten B und die Testdaten X aus der Datenbank in den Arbeitsspeicher. Nun nimmt sich der VDP den ersten Warenkorb aus den Testdaten und durchsucht alle Trainingsdaten nach ähnlichen Warenkörben. Die Ähnlichkeit zweier Warenkörbe wird durch die Multiplikation des Ähnlichkeitsfaktors a mit dem Quotienten aus der Schnittmenge der Warenkörbe geteilt durch die Vereinigungsmenge der Warenkörbe bestimmt. Wird ein ähnlicher Warenkorb in den Trainingsdaten gefunden, wird die Differenzmenge zwischen diesen Warenkörben gebildet. Die Produkte der Differenzmenge werden als Kandidaten in die Hash-Map übernommen bzw. falls diese schon vorhanden sind wird die Häufigkeit (count) erhöht. Nachdem alle Trainingsdatensätze auf diese Weise durchsucht worden sind, werden die top k Kandidaten mit dem höchsten Count aus der Hash-Map ausgelesen und als Vorhersage übernommen. Dieser Vorgang wird für alle Warenkörbe der Testdaten durchgeführt.

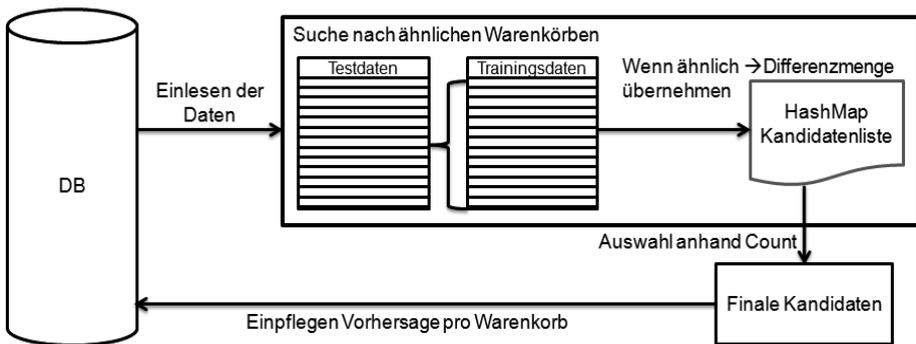


Abbildung 1: Graphische Darstellung VDP

3 Komplexität

Der Block (1)-(16) wird $|X|$ mal ausgeführt. Für die Erstellung und Initialisierung der Hashmap H sind $const |P|$ Instruktionen notwendig. Die Schleife (3)-(9) wird $|B|$ mal durchlaufen. Die Befehle in der Schleife können durch die Größe der Warenkörbe $|b \setminus x| \leq |b|$ abgeschätzt werden. Die Anzahl der Instruktionen in der m -Schleife (10)-(14) kann durch den Ausdruck $const_1 + const_2 * k * (|P|)$ abgeschätzt werden. D.h. der Aufwand für die Top- k items in B ist deutlich geringer als ein vollständiger Sort. Damit ergibt sich für die Zeitkomplexität $O(|X| * (|P| + |B| * |b \setminus x| + k * |P|)) \leq O(|X| * |P| * (1 + |B| + k)) \leq O(|X| * |P| * (|B| + k))$. Sind k und die Anzahl der Produkte begrenzt, d.h. unabhängig von der Anzahl der Trainingsmenge (Warenkörbe), dann berechnet sich die Zeitkomplexität nur noch aus der Anzahl zu prognostizierender Warenkörbe und der Größe der Trainingsmenge: $O(|X| * |B|)$.

4 Vorhersageergebnisse des VDP im Vergleich zum Apriori Algorithmus

Der eben beschriebene VDP Algorithmus wurde in Java implementiert und auf die Daten des DataMiningCups 2011 angewendet. Für jedes richtig vorhergesagte Produkt wurden Punkte vergeben. Er konnte für 13.36 % der Testdaten richtige Vorhersage treffen und benötigte hierfür lediglich 400 Minuten (ca. 6,6 Stunden). Daraus ergibt sich nach den Bewertungskriterien für richtig vorhergesagte Produkte des DataMiningCups eine Punktezahl von 42.209. Im Vergleich hierzu folgen die Zahlen aus der Anwendungen einer Apriori-Implementierung von [JFM11], da die Standardimplementierung zu keinem Ergebnis geführt hat. Mit der Apriori-Implementierung von [JFM11] war es möglich, 33.725 Punkte zu erreichen. Jedoch wurde hierfür inklusive Regelerstellung ca. 10 Stunden Rechenzeit benötigt.

5 Fazit

Der VDP-Algorithmus erreicht mit dem differenzmengenbasierten Ansatz ein leicht besseres Ergebnis als der Apriori. Der einfache und flexible Aufbau des Algorithmus ermöglicht die schnelle Veränderung der verschiedenen Eingangsparameter. Dadurch können leicht verschiedene Szenarien erstellt und ausgewertet werden. Der Laufzeitvorteil gegenüber den Standardverfahren kann leicht durch eine Parallelisierung des linearen Suchverlaufes erreicht werden. Hierfür muss lediglich die zu prognostizierende Warenkorbmenge X auf mehrere Threads aufgeteilt werden. Die oben genannte Durchlaufzeit wurde durch insgesamt 8 Threads erreicht. Demensprechend kann die Durchlaufzeit konstant durch hinzufügen weiterer Threads reduziert werden. Der größte Vorteil des VDP ist jedoch, dass die Anzahl unterschiedlicher Produkte, im Gegensatz zum Apriori, lediglich linearen Einfluss auf die Laufzeit hat. Beim Apriori potenziert sich bei einem hinreichend kleinen Support die Anzahl der möglichen Kandidaten. Da sich nur die Größe der Kandidaten Hash-Map konstant (Komplexität der Erstellung beträgt $const \cdot |P|$) vergrößert. Somit ist dieses Verfahren vor allem für große Datenmengen mit vielen unterschiedlicher Items und großen Warenkörben besonders geeignet.

Literaturverzeichnis

- [Yo10] Yong, L.: The Java Implementation for Apriori algorithm Based on Agile Design Principles. ICCSIT, Chengdu, 2000, S. 329-331.
- [AIS93] Agrawal, R., Imielinski, T., Swami, A.: Mining association rules between sets of items in large databases. In Proc. Of the ACM SIGMOD Conference of Management of Data, S. 207-216, Washington, D.C., 1993
- [JFM11] Falkenthal, M., Jugel, D, Maucher, B.: Nachbetrachtung und Analyse der Ereignisse der eingereichten Lösungen für den Data-Mining-Cups 2011, Referat, Hochschule-Reutlingen, Fakultät Informatik, Studiengang Wirtschaftsinformatik, 2011
- [WI05] Witten, I. H.; Frank, E.; Hall, M. A., Data mining: practical machine learning tools and techniques, Third Edition, Elsevier Inc., Burlington 2011

Leistungsanalyse Evolutionärer Algorithmen auf GPUs im Vergleich zu Multikern-CPUs

Johannes Hofmann*

*Lehrstuhl Rechnerarchitektur
Friedrich-Alexander-Universität Erlangen-Nürnberg*

Abstract: Evolutionäre Algorithmen (EAs) umfassen eine Familie von Metaheuristiken, welche oft dann zum Einsatz kommen, wenn ein Problem mittels exakter Methoden nicht effizient gelöst werden kann. Wegen ihres inhärenten Parallelismus scheinen EAs besonders für die Ausführung auf massiv paralleler Hardware, wie Grafikprozessoren (GPUs), geeignet zu sein. Diese Vermutung wurde durch eine Reihe von Experimenten, in denen Nvidias neue Fermi Architektur mit aktuellen Multikern-CPUs verglichen wurde, untersucht. Die Ergebnisse bestätigen die ursprüngliche Annahme: GPUs konnten einen Speedup von teilweise mehreren Größenordnungen aufweisen.

1 Einleitung

Die Familie der EAs umfasst neben den vier klassischen Varianten — Genetische Algorithmen [Hol75], Evolutionsstrategien [Rec73] Evolutionäre Programmierung [FOW66] und Genetische Programmierung [Koz92] — auch neue, weniger bekannte Vertreter wie z.B. Grammatikalische Evolution [O’N01] und Differenzielle Evolution [PS97].

Obwohl sich die verschiedenen Varianten getrennt voneinander entwickelt haben und erst später unter dem Oberbegriff Evolutionäre Algorithmen zusammengefasst wurden arbeiten sie alle nach dem selben, biologisch-inspirierten Prinzip der darwinistischen Evolution. Statt nur eine Lösung für ein Problem zu berechnen wird ein populationsorientierter Ansatz verfolgt: Eine Menge von *Individuen*, wovon jedes einer möglichen Lösung des Problems entspricht, bilden eine sogenannte *Generation*. Unter Zuhilfenahme einer an das zu lösende Problem angepassten *Fitness*-Funktion wird jedem Individuum ein Fitness-Wert zugeordnet, welcher über die Güte einer Lösung Auskunft gibt. Anhand des Fitness-Werts werden im Selektionsschritt die besten Individuen einer Population als Eltern ausgewählt. Im Anschluss werden aus diesen durch Kreuzung und Mutation Nachfahren erzeugt, welche entsprechend einer vorher festgelegten Ersetzungstrategie entweder alleine oder zusammen mit einigen Eltern die nächste Generation bilden. Dies wird so lange wiederholt bis ein spezifiziertes Abbruchkriterium — z.B. eine maximale Anzahl an Generationen wurde erreicht oder eine ausreichend gute Lösung wurde gefunden — erfüllt ist.

*Johannes.Hofmann@informatik.uni-erlangen.de

Gemeinsam mit anderen Metaheuristiken werden EAs häufig dann eingesetzt, wenn exakte Methoden zu lange zur Lösungsfindung benötigen; die angegangenen Probleme erweisen sich also als besonders rechenintensiv. Da selbst EAs zur Bearbeitung solch komplexer Probleme eine hohe Laufzeit aufweisen können, kann es deshalb vorteilhaft sein die benötigte Zeit durch Parallelisierung zu vermindern.

Als Architekturen kommen dabei neben den klassischen Vertretern wie Multikern-CPU's oder Cluster- und Gridnetzwerken — auf denen die Parallelisierung von EAs bereits gut untersucht ist — auch GPU's in Frage, da sie sich, angetrieben durch den Hunger der Spieleindustrie nach immer realistischeren Grafiken, in den vergangenen Jahren zu einem besonders leistungsfähigen und gleichzeitig kostengünstigen Vertreter der parallelen Architekturen entwickelt haben.

2 Problemstellungen

Um eine möglichst hohe Aussagekraft aus den Ergebnissen ableiten zu können wurden drei verschiedene Vertreter aus der Familie der EAs ausgewählt: Evolutionstrategien, Genetische Algorithmen und Grammatikalische Evolution.

Die erste Aufgabe besteht aus einem Optimierungsproblem. Benannt nach Karl Weierstrass erweist sich die Funktion

$$f(x) = \sum_{n=0}^{\infty} a^n \cos(b^n \pi x)$$

mit $0 < a < 1$, $b \in 2\mathbb{N}-1$ und $ab > 1+3/2\pi$ als besonders geeignetes Optimierungsproblem: Weil die unendliche Summe der Kosinus-Funktionen nur durch eine endliche Summe approximiert werden kann besitzt die Funktion einen Parameter durch den die Funktion entweder speicher- oder rechenlastig wird.

Das Problem des Handlungsreisenden (engl. *Traveling Salesman Problem*, kurz TSP) ist eines der ältesten und am besten untersuchten Beispiele aus der Menge der NP-vollständigen Probleme [ABCC06]. Wegen seiner Beständigkeit im akademischen wie auch im praktischen Bereich wird das TSP gerne als Benchmark für Optimierungsalgorithmen verwendet, weshalb es mit in die Performanceevaluierung aufgenommen wurde.

Als dritte Problemstellung wurde die grammatikalische Evolution Neuronale Netze (NNs) [Arb95] gewählt. NNs werden häufig zur Modellierung von Systemen verwendet, die sich aufgrund ihrer Komplexität der Beschreibung durch traditionelle Modelle entziehen. Die Qualität eines NN ist dabei abhängig von seinen Inputs und seiner Topologie, wobei es keine exakte Methode gibt um die "richtigen" Inputs oder die "richtige" Topologie zu finden; vielmehr werden in der Praxis Heuristiken verwendet bis man bei einer Kombination von Inputs und Topologie angelangt ist, die den Ansprüchen gerecht werden. EAs bieten eine Lösung für dieses Dilemma indem sie eine Population von NNs mit verschiedenen Inputs und Topologien verwenden und durch Kreuzung bestehender, "guter" NNs versuchen die beste Kombination von Inputs und Topologie zu finden.

Zum Vergleich der unterschiedlichen Plattformen wurde jeder der drei oben genannten Algorithmen zunächst als sequentielle Version für CPUs implementiert. Im Anschluss wurde diese unter Zuhilfenahme von OpenMP und Algorithmen aus der `libstdc++` Bibliothek `<parallel/algorithm>` parallelisiert. Aufbauend auf der parallelen Implementierung für CPUs wurde schließlich eine CUDA Version des Algorithmus für GPUs erstellt.

Als Hardware kamen CPU-seitig Intel Xeon X5650 und AMD Opteron 2435 Prozessoren zum Einsatz; jede der CPUs besitzt 6 Kerne. Als Vertreter für GPUs wurden eine Nvidia Tesla C1060 (G200-Chipsatz) und eine Nvidia GeForce GTX480 (Fermi-Chipsatz) gewählt; besonders hervorzuheben ist hierbei, dass die neue Fermi Architektur im Gegensatz zu ihrem Vorgänger (G200) als erste GPU eine vollständige Cache-Hierarchie inklusive L1 Cache besitzt.

3 Ergebnisse

Als Anhaltspunkt für den erwarteten Speedup dienten die Herstellerangaben der *single precision* GFLOPS: GeForce GTX 480 (1345 GFLOPS), Tesla C1060 (936 GFLOPS), Opteron 2435 (62 GFLOPS) und Xeon X5650 (64 GFLOPS). Der Quotient der Leistungen von GeForce GTX und der schnelleren CPU (Intel Xeon) lässt zunächst einen Speedup von etwa 21 erwarten.

Tatsächlich war die Fermi-Karte bei der Weierstrass-Optimierung 210 mal schneller als der parallele Algorithmus auf der Xeon CPU und sogar 320 mal schneller als auf dem AMD Prozessor; die Tesla GPU erreicht einen Speedup von immerhin 70 (Xeon) bzw. 107 (Opteron). Die Diskrepanz zwischen erwartetem und tatsächlichem Speedup lässt sich u.a. durch die fehlende Vektorisierbarkeit des Problems erklären: CPUs beziehen einen Großteil ihrer GFLOPS-Performance durch gleichzeitige Datenverarbeitung in SIMD-Registern (SSE, AVX). Lässt sich Vektorisierung für ein Problem nicht anwenden liegt ein Großteil der Leistung brach (bis zu 87.5% im Fall von AVX).

Für das zweite Problem wurde der Algorithmus nach Vorbild von Yengoku und Yoshihara [SY93] implementiert. Ihr Algorithmus verwendet neben Komponenten aus der Familie der GAs einen sogenannten *expert knowledge* Operator, welcher eine an das Problem angepasste Metaheuristik [Cro58] verwendet um vorhandene Lösungen zu verbessern. Die Leistung der GPUs fiel hierbei eher ernüchternd aus: Während die GTX 480 noch beinahe so schnell wie die AMD CPU war (Speedup 0.9) war sie deutlich langsamer als die Xeon CPU (Speedup 0.6); die Tesla konnte nur ein Fünftel bzw. ein Sechstel der Leistung der Intel bzw. AMD CPU erreichen. Die schlechte Performance ist durch das Verwenden des *expert knowledge* Operators zu erklären: Der Programmfluss der verwendeten *greedy 2opt* Heuristik ist zu divergent um effizient auf der Hardware der GPU ausgeführt zu werden.

Im dritten Algorithmus konnten die GPUs den Wettbewerb wieder deutlich für sich entscheiden. Zwar verlangte die effiziente Parallelisierung des *Backpropagation Algorithm* zum Trainieren der NNs eine aufwändige Abbildung der einzelnen Netzwerke auf die *Streaming Multiprozessoren* der GPU, jedoch konnte der dadurch anfallende, nicht vollständige

parallelisierbare Teil mittels paralleler Reduktion noch so beschleunigt werden, dass die GeFoCe einen Speedup von 61 gegenüber der AMD und 55 gegenüber der Intel CPU erreichen konnte; die Tesla war 20 bzw. 18 mal schneller als die AMD bzw. Intel CPU.

4 Fazit

Wie erwartet erweist sich die Parallelisierung von EAs durch ihren populationsbasierten Ansatz als äußerst effizient: Bei Kreuzung, Mutation, Evaluierung der Fitness lassen sich die Individuen auf die einzelnen Kerne der GPU abbilden. Auch Aufgaben wie die Sortierung der Individuen nach ihrer Fitness — z.B. benötigt für die Auswahl der neuen Generation — sind mit Hilfe der Thrust-Bibliothek [HB10] effizient umsetzbar.

Werden EAs durch spezielle an das Problem angepasste Operatoren erweitert ist jedoch zu prüfen, ob diese effizient auf GPUs abgebildet werden können, um gegebenenfalls durch alternative Heuristiken ersetzt zu werden.

Literatur

- [ABCC06] D. Applegate, R. Bixby, V. Chvátal und W. Cook. *The traveling salesman problem: a computational study*. Princeton University Press, Princeton, NJ, 2006.
- [Arb95] M. A. Arbib, Hrsg. *The Handbook of Brain Theory and Neural Networks*. MIT Press, Cambridge, MA, 1995.
- [Cro58] G. A. Croes. A Method for Solving Traveling-Salesman Problems. *Operations Research*, 6:791–812, 1958.
- [FOW66] L. J. Fogel, A. J. Owens und M. J. Walsh. *Artificial Intelligence through Simulated Evolution*. John Wiley & Sons, New York, 1966.
- [HB10] Jared Hoberock und Nathan Bell. Thrust: A Parallel Template Library, 2010. Version 1.3.0.
- [Hol75] J. H. Holland. *Adaption in Natural and Artificial Systems*. University of Michigan Press, 1975.
- [Koz92] John R. Koza. *Genetic Programming: On the Programming of Computers by Natural Selection*. MIT Press, Cambridge, MA, 1992.
- [O’N01] Michael O’Neill. *Automatic Programming in an Arbitrary Language: Evolving Programs with Grammatical Evolution*. Dissertation, University Of Limerick, Ireland, August 2001.
- [PS97] Kenneth Price und Rainer Storn. Differential Evolution: A simple evolution strategy for fast optimization. *Dr. Dobb’s Journal of Software Tools*, 22(4):18–??, April 1997.
- [Rec73] Ingo Rechenberg. *Evolutionsstrategie, Jgg. 15 of problemata*. Friedrich Frommann Verlag (Günther Holzboog KG), Stuttgart, 1973.

Automated Program Verification Using SAT- and SMT-Encodings of Diophantine Constraints

Patrick Kabasci
patrick.kabasci@rwth-aachen.de

Mentor: Prof. Dr. J. Giesl
Advisor: Dr. C. Fuhs

Group: Foundations of Informatics

1 Introduction

Termination analysis is an important aspect of formal verification, as a positive proof of a program's (universal) termination shows that no infinite loop is entered on any input.

Term rewriting [BN98] is a notion used e.g. in research of automated termination proving. It has very simple syntactic structure, yet is Turing-complete. Automatic tools to prove termination of a term rewriting system (*TRS*), e.g. [GST06, End06, HM07] compete in an annual *termination competition* [MZ07] on a database of example problems with different reasons for (non-)termination. Advances in automated TRS termination proving can be adapted to proving termination of programs in programming languages, either by certain transformations of these programs to term rewrite systems or by applying a version of the newly developed termination technique directly to programming languages where applicable. Termination problems of TRSs are usually considered as *dependency pair problems* [GTS05] as the graph-like structure of dependency pairs with relaxed monotonicity conditions improves the use of techniques to prove termination.

Several techniques to solve termination problems for TRSs include the search for an order with a mathematical notion of boolean combinations of *Diophantine constraints*, e.g. [FGM⁺07, FNO⁺08, FGM⁺08, EWZ08]. A (non-complete) method for solving these constraints is a transformation to the satisfiability problems of propositional logic (*SAT*) by using an approach called SAT modulo theories (*SMT*).

2 Contributions and research

An automated termination prover, APROVE, is developed at RWTH Aachen University's Lehr- und Forschungsgebiet Informatik 2 [GST06]. This thesis was done in the context of its development: In this thesis, we tackled the chain of going from termination problems to Diophantine problems and encoding those to SAT, a core component in APROVE, in the dimensions of developing more powerful techniques and theory, developing heuristics allowing for more efficient solving, and improving the development process for the implementation of such SAT encodings.

An important observation is that although the transformation steps themselves are modular, the suitability of encodings and transformations vary by the problem type. Both trans-



Figure 1: The considered three-step approach of solving termination problems via sound conversion to Diophantine problems which are then solved by a SAT-solver back-end. This approach is used for several techniques for proving termination of TRSs in APROVE.

formation steps mask structure and complexity of termination problems, but the masking does not remove the structure, it only hides it, i.e., hard termination problems tend to be converted to hard SAT instances.

2.1 Termination problems

A way to convert relative termination problems to a variant of dependency pair problems was shown

Relative termination [Ges90] is a formalism in which certain rules of a TRS are marked as *strict* and others as *weak*. It means that any reduction may only contain finitely many strict rules, but – besides that – may be infinite. Relative termination is useful for evaluating fairness and proving confluence of general term rewrite systems, and is therefore useful in automated theorem proving.

As the dependency pair method was not applicable to relative TRSs unless strong conditions were fulfilled, one needed to use strictly monotonic reduction orders to prove relative termination. Orders based on Diophantine constraints are severely weakened by forbidding coefficients to be zero, which is a consequence of strict monotonicity. Finding a way to apply the dependency pair method for relative termination problems has been posted to the list of open research problems in term rewriting [DT10] by Giesl and Zantema.

We found a way to reduce a relative termination problem of greatly extended shape (only the relative rules still need to be right-linear) to a DP problem where certain nodes in the dependency graph are marked as weak, and after application of DP framework [GTS05] processors, components of the graph which only contain weak nodes can be disregarded. However, a residual relative termination problem, which consists of the weak rules and a subset of the strong rules which are in a certain way closely connected to the weak rules, still needs to be solved the classic way. This result enables us to solve relative termination problems while partly dropping monotonicity requirements. We fulfilled our aim by proving relative termination of two particular problems in the Termination Problem Data Base [tpd], `gcd.trs` and `gcd.list.trs`, relative termination of which could not be proven automatically before. The proof of these systems is interesting as they both calculate a commonly-used function, and are derived from standard termination problems calculating this function which can easily be proved terminating, however termination of the relative systems could not be proven automatically without our technique.

2.2 From termination problem to Diophantine problems

Limitational orders provide a method to solve TRSs with infinitely many rules, and reduce the size of coefficients needed by polynomial orderings

Polynomial orderings are the most basic types of orderings usable to prove termination of

TRS which can be found by solving Diophantine constraints. We conducted an analysis showing that not only are most reference problems which are solvable using a polynomial ordering also solvable using small coefficient ranges, but furthermore that increasing coefficient range requires a much larger amount of searching time, with empirically exponential bit-complexity. This is to be expected due to NP-completeness of SAT. We introduced an approach allowing certain coefficients to be arbitrarily large while keeping overall coefficient range small. This approach did not result in improvements in practice, as the overhead constraints introduced made solving this problem as hard as solving an increased-range polynomial order search in the first place. However, this result is interesting in two theoretical ways: First it allows to solve infinite systems not solvable using classic polynomial orderings, and second it combines the notions of arctic and classic polynomial interpretations with only very slight modifications in the order itself.

2.3 Solving Diophantine problems

Satisfiability of non-linear integer arithmetic restricted to coefficients of powers of two or zero can be solved soundly and completely by reduction to linear integer arithmetic

Satisfiability of Diophantine constraints, or non-linear integer arithmetic, is undecidable over the naturals [Mat73]. Our encoding theoretically allows us to completely solve Diophantine constraints over an infinite subset of the naturals. Current state-of-the-art SMT solvers prove too weak to solve the resulting constraints. As SMT solvers become stronger in solving problems with both a difficult theory part and a difficult propositional skeleton, this approach may become feasible.

2.4 Transforming Diophantine problems to SAT

As a heuristic, shifts instead of multiplications can be used to efficiently model non-linear constraints

Deriving from the completeness result above, we abstracted to use powers of two as a heuristic. As rational polynomial and rational matrix orderings are commonly used classes of orderings for TRSs and DP termination problems, and such problems empirically tend to only need coefficients of powers of two or zero [Luc07], restricting interpretations to these coefficients and using modeling of shift operations rather than multiplications was found to increase efficiency. The swifter failing and success of certain processors can be leveraged in APROVE's configuration to allow the application of other techniques. As the termination competition has a maximum time limit, this efficiency can be critical in just solving instances or just timing out. In later applications of proving termination of actual programs, problems are expected to be larger but simpler. This leads to an increased need of such heuristics. Shifts instead of multiplications allow us to gain 15% efficiency on the instances solved over a full run of APROVE with default configuration.

A comprehensive analysis of encodings for additions has shown Ripple-Carry-Adder trees to be an efficient modeling of additions for non-linear problems

While sorters and bucket adders as addition encodings in SAT have been proposed as best practices for linear applications such as Pseudo-boolean constraint solving [Sör10], we have shown that the benefits of these circuits cannot be leveraged in the problems solved in APROVE, and that benefits of comparatively simple RCA trees outweigh them in this case, as other optimizations [FGM⁺07] can be used which are good in the case of non-

linear applications.

2.5 Solving SAT problems

Debugging SAT modeling has been made easier by a graphical debugging tool

We developed SATView, a graphical tool which allows debugging both satisfiable and unsatisfiable formulae using techniques applied in VLSI debugging. It allows automatic refinement to find reasons of unsatisfiability, and is tightly integrated in APROVE. We intend to further extend SATView to be able to handle formulae resulting from arbitrary applications, so that it can be useful to the general research community.

References

- [BN98] F. Baader and T. Nipkow. *Term rewriting and all that*. Cambridge University Press, 1998.
- [DT10] N. Dershowitz and R. Treinen. The RTA list of open problems, 2010. <http://rtaloop.mancoosi.univ-paris-diderot.fr/>.
- [End06] J. Endrullis. Jambox, Automated Termination Proofs For String and Term Rewriting, 2006.
- [EWZ08] J. Endrullis, J. Waldmann, and H. Zantema. Matrix interpretations for proving termination of term rewriting. *Journal of Automated Reasoning*, 40(2):195–220, 2008.
- [FGM⁺07] C. Fuhs, J. Giesl, A. Middeldorp, P. Schneider-Kamp, R. Thiemann, and H. Zankl. SAT solving for termination analysis with polynomial interpretations. In *Theory and Applications of Satisfiability Testing*, pages 340–354. Springer, 2007.
- [FGM⁺08] C. Fuhs, J. Giesl, A. Middeldorp, P. Schneider-Kamp, R. Thiemann, and H. Zankl. Maximal termination. In *Rewriting Techniques and Applications*, pages 110–125. Springer, 2008.
- [FNO⁺08] C. Fuhs, R. Navarro-Marset, C. Otto, J. Giesl, S. Lucas, and P. Schneider-Kamp. Search techniques for rational polynomial orders. In *Intelligent Computer Mathematics*, pages 109–124. Springer, 2008.
- [Ges90] A. Geser. *Relative termination*. PhD thesis, Universität Passau, 1990.
- [GST06] J. Giesl, P. Schneider-Kamp, and R. Thiemann. AProVE 1.2: Automatic termination proofs in the dependency pair framework. *Automated Reasoning*, pages 281–286, 2006.
- [GTS05] J. Giesl, R. Thiemann, and P. Schneider-Kamp. The dependency pair framework: Combining techniques for automated termination proofs. In *Logic for Programming, Artificial Intelligence, and Reasoning*, pages 301–331. Springer, 2005.
- [HM07] N. Hirokawa and A. Middeldorp. Tyrolean termination tool: Techniques and features. *Information and Computation*, 205(4):474–511, 2007.
- [Luc07] S. Lucas. Practical use of polynomials over the reals in proofs of termination. In *Principles and Practice of Declarative Programming*, pages 39–50. ACM, 2007.
- [Mat73] YUV Matiasевич. On Recursive Unsolvability Of Hilbert’s Tenth Problem. In *Logic, methodology and philosophy of science: Proceedings*, pages 89–110. North-Holland, 1973.
- [MZ07] C. Marché and H. Zantema. The termination competition. *Term Rewriting and Applications*, pages 303–313, 2007.
- [Sör10] N. Sörensson. MiniSat 2.2 and MiniSat++ 1.1. Technical report, Sörensson Research and Development, 2010.
- [tpd] The Termination Problem Data Base (TPDB). <http://termination-portal.org/wiki/TPDB>.

Erreichbarkeit mit Ressourcenschranken für Pushdownsysteme

Martin Lang

RWTH Aachen - Lehrstuhl für Informatik 7

Transitionssysteme, die durch den Konfigurationsgraph von Kellerautomaten induziert werden, haben sich zu einem bedeutenden Werkzeug in der automatischen Programmverifikation entwickelt. Einerseits bieten sie die Ausdruckstärke, um rekursive Programme zu modellieren. Andererseits sind gute algorithmische Methoden zur effizienten Lösung zentraler Probleme wie Punkt-zu-Punkt-Erreichbarkeit vorhanden. Die Theorie dieser Transitionssysteme, die im Weiteren als *Pushdownsysteme* bezeichnet werden, ist gut verstanden. Ein Beispiel, das die praktische Anwendbarkeit der Ergebnisse belegt, ist der Java-Modelchecker jMoped.

Parallel zur Entwicklung komplexerer Automatenmodelle wie Kellerautomaten wurden quantitative Varianten endlicher Automaten entworfen. Diese Automaten definieren keine formalen Sprachen im klassischen Sinn, sondern induzieren eine Abbildung von Wörtern in eine andere Menge, wie beispielsweise die natürlichen Zahlen. Neben dem Ansatz der gewichteten Automaten, der 1961 durch M.-P. Schützenberger in [Sch61] eingeführt wurde, existieren verschiedene speziellere Automatenmodelle wie das Modell der Distanzautomaten. Diese wurden 1982 von K. Hashiguchi (siehe [Has82]) eingeführt, um das Sternhöhenproblem für reguläre Sprachen zu lösen. Vor kurzem verallgemeinerte die Arbeit [Col09] von T. Colcombet die bisherigen Erweiterungen von Distanzautomaten zur Theorie der *regulären Kostenfunktionen*.

In der vorliegenden Arbeit werden die reichhaltigen Ergebnisse über Pushdownsysteme und quantitative Automatenmodelle genutzt, um eine Theorie von Pushdownsystemen mit Ressourcenverbrauch zu entwickeln. Diese Theorie soll die Beantwortung von Verifikationsfragen über rekursive Systeme/Programme mit Ressourcenverbrauch ermöglichen. Zur Formalisierung rekursiver Systeme mit Ressourcenverbrauch wird ein kombiniertes Modell aus Pushdownsystemen und Ressourcenzählern eingeführt. Ausgehend von der Frage der Realisierbarkeit werden Methoden zur algorithmischen Analyse des Modells entwickelt. Hierbei steht die Fragestellung nach der Erreichbarkeit mit beschränkten Ressourcen im Vordergrund und bildet den Ausgangspunkt für die Entwicklung einer allgemeinen Theorie von Strukturen mit Ressourcenverbrauch sowie einer quantitativen Logik zur Spezifikation von Beschränktheitsbedingungen.

Bereits vor einiger Zeit wurden andere Modelle zur formalen Beschreibung von Systemen mit Ressourcenverbrauch vorgeschlagen und untersucht, die als Basis für die Ressourcenspezifikation Ideen aus der Theorie der gewichteten Automaten nutzen. In [BFL⁺08,

CdAHS03] wird das Modell von *Energie-Spielen* vorgestellt und untersucht. Vor kurzem wurden auf Basis dieser Arbeiten auch Energie-Spiele mit zusätzlicher Paritätsbedingung untersucht [CD10]. Obwohl im Rahmen der Erforschung dieser Art von Spielen ebenfalls Beschränktheitsfragestellungen untersucht wurden, besteht kein einfacher Weg Ergebnisse mit dem hier vorgestellten Modell auszutauschen. Dies liegt zum einen in dem unterschiedlichen Modell des Ressourcenverbrauchs begründet und zum anderen darin, dass im Kontext von Energie-Spielen lediglich Systeme mit endlichem Zustandsraum betrachtet werden. Die im Rahmen dieser Arbeit entwickelten Ergebnisse beziehen sich auf Systeme mit potentiell unendlichem Zustandsraum.

1 Ressourcen-Pushdownsysteme

Ressourcen werden im Rahmen dieser Arbeit in Form von ganzzahligen Zählern modelliert. Diese Zähler starten bei Null und können inkrementiert sowie auf Null zurückgesetzt werden. Hierbei entspricht das Inkrementieren des Zählers dem Verbrauch einer Ressource und das Zurücksetzen dem (vollständigen) Wiederauffüllen. Dieses Zählerschema orientiert sich am Modell der *B-Automaten*. Leicht vereinfacht dargestellt sind B-Automaten nichtdeterministische endliche Automaten mit einer endlichen Menge von Zählern, die die oben beschriebenen Operationen unterstützen. Alle Transitionen des Automaten sind mit Operationen für die einzelnen Zähler annotiert. Die Zähler können dabei unverändert bleiben oder eine der oben beschriebenen Aktionen ausführen. In einem Lauf werden die Zähler auf Basis der annotierten Aktionen simuliert. Der Wert eines Laufes wird durch den höchsten im Lauf auftretenden Zählerwert bestimmt. Ein B-Automat definiert eine Funktion von den Wörtern in die natürlichen Zahlen und unendlich, indem jedes Wort auf den Wert des geringsten akzeptierenden Laufes abgebildet wird. Falls kein akzeptierender Lauf existiert, wird das Wort auf unendlich abgebildet.

Zur Modellierung von rekursiven Programmen mit Ressourcen wird dieses Zählerschema auf Pushdownsysteme übertragen. Ein *Ressourcen-Pushdownsystem* besteht somit aus einem endlichen Kellularalphabet, jeweils einer endlichen Menge von Zuständen und Zählern sowie Übergangsregeln, die mit Operationen für die einzelnen Zähler annotiert sind. Eine Konfiguration eines Ressourcen-Pushdownsystems besteht aus dem aktuellen Zustand sowie dem Kellerinhalt. Der Lauf eines rekursiven Programms wird mit einer Folge von Konfigurationen des Systems identifiziert. Analog zur Definition für B-Automaten werden die Zähler in der Konfigurationsfolge simuliert. Die benötigten Ressourcen für den Lauf werden durch den höchsten auftretenden Zählerwert festgelegt.

Motiviert durch die Frage der Realisierbarkeit eines technischen Systems werden Beschränktheitsfragen für Systeme mit Ressourcenverbrauch untersucht. Im Kontext eines mobilen Gerätes für vorbestimmte Anwendungen ist beispielsweise die Fragestellung interessant, ob das Gerät jede geplante Anwendung durchführen kann ohne zwischendurch aufgeladen werden zu müssen. Dies ist die Frage danach, ob der Energieverbrauch bei Start einer beliebigen Anwendung beschränkt ist. Im Rahmen dieser Arbeit wird diese spezielle

Fragestellung verallgemeinert und in Form des *Erreichbarkeitsproblems mit beschränkten Ressourcen* formalisiert. Seien A und B zwei Mengen von Konfigurationen. Die Menge B heißt *mit beschränkten Ressourcen von A aus erreichbar*, wenn eine Schranke $k \in \mathbb{N}$ existiert, so dass es für jede Konfiguration $c \in A$ eine Konfigurationsfolge gibt, die zu einer Konfiguration von B führt und weniger als k Ressourcen benötigt.

Mit der Saturierungsmethode existiert ein effizienter Ansatz zur Berechnung von Erreichbarkeit auf Pushdownsystemen. Sie geht auf Ideen von R. Book, F. Otto sowie J. Sakarovitch zurück [BS86, BO93] und basiert auf einem induktiven Ansatz. Die Methode startet mit einem Automaten, der als Sprache genau die Menge der Zielkonfigurationen erkennt. Diesem Automaten werden auf Basis der Transitionsregeln des Pushdownsystems sukzessive neue Transitionen hinzugefügt, die die Transitionen des Pushdownsystems simulieren, bis sich keine Veränderung des Automaten mehr ergibt. Der resultierende Automat erkennt je nach Variante die Menge aller Vorgänger- oder Nachfolgerkonfigurationen des Ursprungsautomaten. In dieser Arbeit wird auf Basis der in [LHDT87] vorgestellten Variante ein Saturierungsalgorithmus für das Punkt-zu-Punkt-Erreichbarkeitsproblem entwickelt, der die Ressourcennutzung in Ressourcen-Pushdownsystemen mitverfolgt. Das Ergebnis des Algorithmus ist ein synchroner Transducer mit der Semantik von B-Automaten, der ein Konfigurationspaar liest und die minimalen Ressourcen ausgibt, die zum Erreichen der einen Konfiguration von der anderen notwendig sind. Die entwickelte Methode ist darüber hinaus auf andere Arten von Transitionsannotation übertragbar.

Über normale Erreichbarkeit hinaus wird im Rahmen der Arbeit auch alternierende Erreichbarkeit auf Ressourcen-Pushdownsystemen in Form von Zwei-Spieler-Erreichbarkeitsspielen untersucht. Analog zum Erreichbarkeitsproblem mit beschränkten Ressourcen wird für diese Spiele untersucht, ob das Spiel mit beschränktem Ressourceneinsatz gewonnen werden kann. Hier konnte auf Basis einer Saturierungsmethode von T. Cachet [Cac02] gezeigt werden, dass dieses Problem im Falle einer regulären Zielmenge für Pushdownsysteme mit nur einem Zähler ohne Zurücksetzen entscheidbar ist.

2 Ressourcenstrukturen und Logik

Zur allgemeineren Untersuchung von Systemen mit Ressourcenverbrauch wird in der Arbeit das Konzept von Ressourcenstrukturen sowie die quantitative Logik FO+RR (First-Order+Resource Relations) entwickelt. Eine Ressourcenstruktur ist eine relationale Struktur, deren Relationen Ressourcenverbrauch im System abbilden. Dazu ordnet eine Relation jedem Tupel von Elementen die Anzahl der benötigten Ressourcen zu, um in dieser Relation zu sein. Ein Beispiel für diese Art von Relation ist die Erreichbarkeitsrelation auf Ressourcen-Pushdownsystemen, die jedem Paar von Konfigurationen die minimalen Ressourcen zuordnet, um von der einen die andere Konfiguration zu erreichen.

Die Logik FO+RR ist syntaktisch wie die Logik erster Stufe ohne Negation aufgebaut. Sie wird über Ressourcenstrukturen ausgewertet und ordnet jeder Formel eine natürliche

Zahl oder unendlich zu. Die Semantik der Logik basiert auf der Intuition, dass jeder Formel die Anzahl von Ressourcen zugeordnet wird, die nötig ist, um die Formel zu erfüllen. Die Logik ermöglicht die Spezifikation von Beschränktheitsanforderungen auf Systemeigenschaften, die sich in Logik erster Stufe ausdrücken lassen. Beispielsweise kann das Erreichbarkeitsproblem mit beschränkten Ressourcen in FO+RR ausgedrückt werden.

Zur effektiven Anwendung in der formalen Verifikation ist die effektive Berechenbarkeit der Semantik eine notwendige Voraussetzung. Im Rahmen der Arbeit wird eine große Klasse von Ressourcenstrukturen mit entscheidbarer FO+RR-Semantik identifiziert. Motiviert durch das Resultat, dass die Theorie erster Stufe automatischer Strukturen immer entscheidbar ist (siehe [KN95]), werden ressourcenautomatische Strukturen definiert. Analog zu normalen automatischen Strukturen müssen diese über eine Repräsentation des Universums in Form einer regulären Sprache über einem endlichen Alphabet verfügen. Ferner müssen die Relationen der Struktur durch einen synchronen Transducer mit B-Automaten-Semantik darstellbar sein. Unter Benutzung der von T. Colcombet nachgewiesenen Abchlusseigenschaften von regulären Kostenfunktionen wird in der Arbeit die effektive Berechenbarkeit von FO+RR auf ressourcenautomatischen Strukturen bewiesen.

Literatur

- [BFL⁺08] P. Bouyer, U. Fahrenberg, K. Larsen, N. Markey und J. Srba. Infinite Runs in Weighted Timed Automata with Energy Constraints. Jgg. 5215 of *LNCS*. Springer, 2008.
- [BO93] R. Book und F. Otto. *String-rewriting systems*. Springer, New-York, 1993.
- [BS86] M. Benois und J. Sakarovitch. On the complexity of some extended word problems defined by cancellation rules. *Information Processing Letters*, 23(5):281–287, 1986.
- [Cac02] T. Cachat. Symbolic strategy synthesis for games on pushdown graphs. *Automata, Languages and Programming*, Seiten 785–785, 2002.
- [CD10] K. Chatterjee und L. Doyen. Energy Parity Games. In *Automata, Languages and Programming*, Jgg. 6199 of *LNCS*, Seiten 599–610. Springer, 2010.
- [CdAHS03] A. Chakrabarti, L. de Alfaro, T. Henzinger und M. Stoelinga. Resource Interfaces. In *Embedded Software*, Jgg. 2855 of *LNCS*, Seiten 117–133. Springer, 2003.
- [Col09] T. Colcombet. Regular cost functions over words. *Manuscript available online*, 2009.
- [Has82] K. Hashiguchi. Limitedness theorem on finite automata with distance functions. *Journal of computer and system sciences*, 24(2):233–244, 1982.
- [KN95] B. Khossainov und A. Nerode. Automatic presentations of structures. In *Logic and Comp. Complexity*, Jgg. 960 of *LNCS*, Seiten 367–392. Springer, 1995.
- [Lan11] M. Lang. Resource-bounded Reachability on Pushdown Systems. Masterarbeit, 2011.
- [LHDT87] P. Lescanne, T. Heuillard, M. Dauchet und S. Tison. Decidability of the confluence of ground term rewriting systems. Bericht RR-0675, INRIA, 1987.
- [Sch61] M.P. Schützenberger. On the Definition of a Family of Automata. *Information and Control*, 4(2-3):245–270, 1961.

Schwarmintelligenz für Handlungsreisende

Andreas Platau

Kontakt: platauan@hs-albsig.de
Hochschule Albstadt-Sigmaringen
Fakultät Business and Computer Science

Art der Arbeit: Bachelor-Arbeit

Betreuer der Arbeit: Prof. Dr. Walter Hower

Abstract: Das Zusammenspiel einzelner Individuen ist das Hauptmerkmal eines Schwarms. Die daraus resultierende *Schwarmintelligenz* spielt eine entscheidende Rolle zur Lösung von kombinatorischen Optimierungsproblemen. Das Verhalten von Ameisen zeigt, dass diese die Fähigkeit besitzen, kürzeste Wege zwischen Nest und Futterquelle zu finden. Diese Arbeit soll aufzeigen, dass Ameisen-Heuristiken das „Problem des Handlungsreisenden“ erfolgreich angehen können. Die theoretischen Hintergründe werden anhand einer Ameisensimulation in einem Versuch praktisch gezeigt.

1 Einleitung

Das Beobachten von Ameisen hat in der Vergangenheit gezeigt, dass Ameisen die Fähigkeit besitzen, kürzeste Wege zwischen Nest und Futterquelle zu finden. Ameisen nutzen bei der Kommunikation chemische Duftstoffe, so genannte *Pheromone*. Dieser Aspekt spielt eine entscheidende Rolle zur Lösung von kombinatorischen Optimierungsproblemen. Speziell lässt sich dieses Verhalten beim Lösen des bekannten „Problem des Handlungsreisenden“, engl. Traveling Salesman Problem (TSP), aufzeigen. Das Verhalten der Ameisen wird in Heuristiken nachgebildet und ist dadurch in der Lage, graph-basierte Optimierungsprobleme zu lösen. Viele Ameisen-Heuristiken basieren auf dem ersten Ameisenalgorithmus *Ant System*. Weiterentwicklungen daraus sind Ameisenalgorithmen wie *MAX – MIN Ant System* und *Ant Colony System*.

Das TSP befasst sich mit dem Finden einer optimalen Rundreise durch eine gegebene Anzahl Städte unter Berücksichtigung der Kostenstruktur. Formal betrachtet ist das Problem ein vollständiger ungerichteter gewichteter Graph $G = (N, E, C)$, wobei N (engl. „nodes“, Knoten) die Menge der Städte, E (engl. „edges“, Kanten) die Menge der Verbindungen zwischen den Städten und C (engl. „costs“) die dazugehörigen Kosten repräsentieren. Bei einer optimalen Rundreise ist die Kosten-Summe minimal. Das Permutations-Problem ist exponentiell (siehe bspw. [How10]); der Suchraum hält theoretisch $(n-1)!$ Möglichkeiten bereit.

2 Systeme

Das *Ant System* (AS) besteht aus zwei Komponenten: einer Entscheidungsregel und der Pheromon-Update-Regel. Mit jedem Schritt, ausgehend von einer Stadt i , wird die nächste Stadt j anhand folgender Entscheidungsregel p_{ij}^k bestimmt:

$$p_{ij}^k = \frac{[\tau_{ij}]^\alpha [\eta_{ij}]^\beta}{\sum_{l \in \mathcal{N}_i^k} [\tau_{il}]^\alpha [\eta_{il}]^\beta}, \text{ falls } j \in \mathcal{N}_i^k \quad (1)$$

Dabei stehen τ_{ij} für das Pheromonlevel und η_{ij} für die Attraktivität (Distanz-Inverse) eines Pfades zwischen Stadt i und j . Die Parameter α und β beeinflussen diese lokalen Informationen und bestimmen somit das Verhalten der *künstlichen* Ameisen. Die Entscheidungsregel setzt die Wahl einer bestimmten Stadt j ins Verhältnis zu allen direkten Nachbarstädten \mathcal{N}_i^k von Ameise k in der Stadt i .

Nach jeder abgeschlossenen Tour werden die Pfade durch eine Pheromon-Update-Regel (Verdunstung ρ) aktualisiert:

$$\tau_{ij} \leftarrow (1 - \rho)\tau_{ij} \quad (2)$$

Danach wird die Pheromon-Konzentration erhöht:

$$\tau_{ij} \leftarrow \tau_{ij} + \sum_{k=1}^m \Delta\tau_{ij}^k \quad (3)$$

Pfade, die von vielen Ameisen genutzt werden, erhalten mehr Pheromone und locken in weiteren Iterationen mehr Ameisen an.

Eine Erweiterung des AS ist das *MAX-MIN Ant System* (MMAS). Dieses System nutzt hauptsächlich die derzeit *besten* Touren und legt ein Intervall $[\tau_{min}, \tau_{max}]$ für die Pheromonwerte fest. Die Entscheidungsregel und die Pheromon-Verdunstung sind wie beim AS, jedoch ist die Pheromon-Erhöhung eine andere:

$$\tau_{ij} \leftarrow \tau_{ij} + \Delta\tau_{ij}^{best} \quad (4)$$

Das *Ant Colony System* (ACS) unterscheidet sich vom AS dadurch, dass es eine aggressivere Entscheidungsregel nutzt und nur die beste Ameise (= beste Tour, daher *greedy*) eine globale Pheromon-Update-Regel anwenden darf. Ebenso wird eine *lokale* Pheromon-Update-Regel genutzt (gilt nach jedem Schritt für jede Ameise):

$$\tau_{ij} \leftarrow (1 - \xi)\tau_{ij} + \xi\tau_0 \quad (5)$$

Die lokale Update-Regel bewirkt, dass der kürzlich besuchte Pfad τ_{ij} in der Pheromon-Intensität reduziert wird. Dadurch wird dieser Pfad weniger begehrenswert für andere Ameisen und zwingt diese, andere Pfade zu erkunden (um evtl. bessere Touren zu erhalten). ξ steht für den lokalen Verdunstungsfaktor.

3 Experimente

Der Simulator TSPAntSim¹ erlaubt eine umfangreiche Parametrisierung der Ameisen-Heuristiken. Folgender Versuchsaufbau wurde für einen Vergleich gewählt:

- Tour: berlin52.tsp ($n := 52$ [Städte], Optimum: 7542) aus TSPLIB²
- 10 Durchläufe mit jeder Parametrisierung —
davon wurde der beste Wert als Vergleichswert herangezogen
- Jeweils 100 und 300 Iterationen für jede Einstellung
- Jede Heuristik erhält den Pheromon-Einfluss-Wert $\alpha := 1$ ([DS04], S. 71, Box 3.1)

Folgende Tabelle veranschaulicht die Versuchsergebnisse (farbige Werte markieren die „besten/optimalen“ Ergebnisse):

Ant System												
α	1	1	1	1	1	1	1	1	1	1	1	1
β	2	2	2	2	3	3	3	3	5	5	5	5
ρ	0.1	0.1	0.5	0.5	0.1	0.1	0.5	0.5	0.1	0.1	0.5	0.5
Ameisen	n	n	n	n	n	n	n	n	n	n	n	n
Iterationen	100	300	100	300	100	300	100	300	100	300	100	300
best. Wert	7678	7548	7622	7543	7570	7547	7658	7548	7674	7547	7657	7674
MMAS												
α	1	1	1	1	1	1	1	1	1	1	1	1
β	2	2	2	2	3	3	3	3	5	5	5	5
ρ	0.1	0.1	0.5	0.5	0.1	0.1	0.5	0.5	0.1	0.1	0.5	0.5
Ameisen	n	n	n	n	n	n	n	n	n	n	n	n
Iterationen	100	300	100	300	100	300	100	300	100	300	100	300
best. Wert	7662	7657	7737	7547	7679	7657	7679	7662	7662	7657	7679	7547
ACS												
β	2	2	2	2	3	3	3	3	5	5	5	5
ρ	0.1	0.1	0.5	0.5	0.1	0.1	0.5	0.5	0.1	0.1	0.5	0.5
ξ	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
Ameisen	10	10	10	10	10	10	10	10	10	10	10	10
Iterationen	100	300	100	300	100	300	100	300	100	300	100	300
best. Wert	8206	7912	7947	8066	8074	8206	8011	8181	7913	7934	8156	8186
ACS-Fortsetzung												
beta	2	2	2	2	3	3	3	3	5	5	5	5
rho	0.1	0.1	0.5	0.5	0.1	0.1	0.5	0.5	0.1	0.1	0.5	0.5
ξ	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
Ameisen	10	10	10	10	10	10	10	10	10	10	10	10
Iterationen	100	300	100	300	100	300	100	300	100	300	100	300
best. Wert	7843	7741	7652	7547	7636	7749	7542	7716	7569	7547	7542	7542

Abbildung 1: Versuchsergebnisse mit TSPAntSim

¹<http://yzgrafik.ege.edu.tr/projects/TSPAntSim/TSPAntSim.html>

²<http://comopt.ifi.uni-heidelberg.de/software/TSPLIB95/>

4 Bewertung

Um eine Vergleichbarkeit der Ergebnisse zu erzielen, wurde für alle Ameisen-Heuristiken der Parameter β schrittweise verändert. Ebenso wurde der Parameter ρ in zwei Werten getestet und es gibt jeweils ein Ergebnis für 100 und 300 Iterationen. Beim *Ant Colony System* (ACS) gibt es noch die Unterscheidung im Parameter ξ (Verdunstungsfaktor) für die lokale Pheromon-Update-Regel (0.1, 0.5).

In den Versuchsergebnissen erzielen AS und ACS die besten Ergebnisse. ACS hat darüber hinaus das beste Laufzeitverhalten. Sowohl 100 als auch 300 Iterationen einer Parametrisierung waren beim ACS in wenigen Sekunden berechnet. ACS erzielt mit 10 Ameisen bereits eine gute Performance, wobei AS eine Ameisenanzahl gemäß Städteanzahl benötigt (vgl. ([DS04], S. 71). ACS erzielt in drei Einstellungen den optimalen Wert von 7542 — dieses Ergebnis hat keine der anderen beiden Heuristiken erreicht. AS erzielt in der Summe die meisten „fast-optimalen“ Ergebnisse. MMAS hingegen hatte zweimal 7547 als Lösung, der drittbeste Wert lag um ca. 100 Punkte darüber (7657).

Die schrittweise Erhöhung des Parameters β von 2 bis 5 führt nicht zwingend zu Verbesserungen beim AS. Sowohl bei $\beta = 2$ als auch bei $\beta = 5$ erzielte AS annähernd optimale Ergebnisse. MMAS sollte für gute Ergebnisse einen Verdunstungsfaktor von $\rho = 0.02$ haben ([DS04], S. 90–91). Der Simulator TSPAntSim kann als ρ -Minimalwert nur 0.1 annehmen. Dieser Aspekt lässt darauf schließen, dass MMAS seine Stärken nicht ganz ausspielen kann und nur bei langen Laufzeiten und großen Touren gute Ergebnisse liefert. Dies ist auch dadurch begründet, dass MMAS anfangs eine sehr hohe Erkundungsphase, danach eine Ausnutzungsphase (die Erfahrung aus der Erkundung) hat — somit werden erst durch zunehmende Iterationen gute Ergebnisse geliefert.

ACS liefert die besten Ergebnisse mit dem Parameter $\xi = 0.5$ (3 mal optimal). Der Parameter β spielt für das ACS eine eher untergeordnete Rolle. Für $\beta = 2$ –5 erzielte es sehr gute, teilweise sogar optimale, Werte. Der Grund dafür scheint zu sein, dass ACS eine sehr aggressive Entscheidungsregel (anders als AS) einsetzt. Die *lokale* Pheromon-Update-Regel bewirkt, dass vermehrt Erkundungen der einzelnen Ameisen gemacht werden und dadurch schliesslich bessere Touren zustande kommen.

Das Experiment hat gezeigt, dass ACS nicht grundlos zu den besten Ameisen-Heuristiken zählt und für das TSP sehr gute Ergebnisse liefert. Jedoch gehört MMAS ebenso zu den besten Heuristiken ([DS04], S. 92), hat hier im Vergleich aber keine herausragenden Ergebnisse erzielt. Der Grund ist, dass das Verhalten der Ameisen-Heuristiken stark von der Parametrisierung abhängt ([DS04], S. 90).

Literatur

- [DS04] Marco Dorigo und Thomas Stützle. *Ant Colony Optimization*. The MIT Press, Cambridge Massachusetts, USA, 2004.
- [How10] Walter Hower. *Diskrete Mathematik — Grundlage der Informatik*. Oldenbourg Wissenschaftsverlag, München, 2010.

An extension of the GiNaCRA library for the Cylindrical Algebraic Decomposition

Joachim Redies
RWTH Aachen University

joachim.redies@rwth-aachen.de

Abstract: In order to decide whether conjunctions of multivariate polynomial equations are satisfiable, we implement an algorithm to compute cylindrical algebraic decompositions (CADs). To provide satisfying solutions, we use interval representations of polynomial roots. In contrast to other tools, our CAD implementation is suitable for a sat modulo theories embedding.

1 Introduction

The first-order theory of real algebra, i.e., the theory of multivariate polynomials with rational coefficients is decidable [Tar48].

Let $\mathbb{Q}[X_1, \dots, X_k]$ denote the ring of polynomials with rational coefficients in k variables.

In this paper, we sketch an algorithm which provides solutions for $\bigwedge_{i=1}^n P_i \diamond_i 0$ with $P_i \in \mathbb{Q}[X_1, \dots, X_k]$ and $\diamond_i \in \{<, \leq, =, \neq, \geq, >\}$ for $1 \leq i \leq n$.

In contrast to the original CAD algorithm [BPR06], we provide a modified version to increase the bounds of applicability (c.f. Section 4).

2 Interval Representations

Fractions are not expressive enough to represent exact polynomial roots because e.g. $\sqrt{2}$ is not representable as a fraction. However, even radical expressions are not sufficient. There are polynomial roots which are not representable with radical expressions. We introduce complete representation of polynomial roots:

Definition. A pair $(P,]a, b[)$ with $P \in \mathbb{Q}[X]$, $]a, b[$ is an open interval with $a < b$ and $a, b \in \mathbb{Q}$ is called *interval representation* [Mis93] if the interval $]a, b[$ contains exactly one root of P .

For an arbitrary small $\varepsilon > 0$ we can refine each interval representation with a bisection algorithm such that $|a - b| < \varepsilon$. For univariate polynomial equations an algorithm based

on Sturm sequences and Cauchy bounds [BPR06] provides the roots in form of interval representations. There is an algorithm to check if two different interval representations of polynomial roots represent equal numbers. This is used to decide whether common roots for a set of polynomials exist.

3 Cylindrical Algebraic Decomposition (CAD) Algorithm

The CAD algorithm consists of two phases: In the projection phase the number of variables is incrementally reduced with an elimination technique until a set of univariate polynomials is obtained. In the lifting phase the solutions computed with univariate polynomials are incrementally substituted.

Given a finite set of multivariate polynomials $\mathcal{P} \subset \mathbb{Q}[X_1, \dots, X_k]$ we project the polynomials successively to a set of univariate polynomials. This projection is called *elimination* and the elimination operator is denoted by $\text{Elim}_{X_k} : 2^{\mathbb{Q}[X_1, \dots, X_k]} \rightarrow 2^{\mathbb{Q}[X_1, \dots, X_{k-1}]}$. A set of polynomial constraints with polynomials $\mathcal{P} \subset \mathbb{Q}[X_1, \dots, X_k]$ has a common solution $(x_1, \dots, x_{k-1}, x_k) \in \mathbb{R}^k$ if and only if there is a solution $(x_1, \dots, x_{k-1}) \in \mathbb{R}^{k-1}$ for any set of polynomial constraints with polynomials in $\text{Elim}_{X_k}(\mathcal{P})$.

Example. Let be given the set $\mathcal{P} := \{X^2 + Y^2 - 1\} \subset \mathbb{Q}[X, Y]$. $\text{Elim}_Y(\mathcal{P})$ is defined as $\{X^2 - 1\}$. With the Sturm sequence/Cauchy bound algorithm we compute that there are two roots given by the interval representations $(X^2 - 1,] - 2, 0[)$ and $(X^2 - 1,]0, 2[)$.

Moreover, we add representations of the non-root connected subsets of \mathbb{R} in between roots; left of the roots and right of the roots. Therefore, we choose the three numbers $-3, 0$ and 3 . Both the interval representations and the representations of the non-root subsets are called *sample points* of $X^2 - 1$.

The definition of the elimination operator ensures for each polynomial constraint $X^2 + Y^2 - 1 \diamond 0$ with $\diamond \in \{<, =, >\}$ that if a solution exists then one of the solutions has an X -component equal to a sample point of $X^2 - 1$. By implication this means, it is sufficient to test solutions with X -components for a finite number of sample points [BPR06].

To compute Y -components of a potential solution, we substitute the X variable in the original polynomial $X^2 + Y^2 - 1$ with each sample point of $X^2 - 1$. This substitution process is called *lifting*. Exemplary, we substitute X with 0 and obtain $Y^2 - 1$. The equation $Y^2 - 1$ has exactly the same sample points as $X^2 - 1$. If we combine for example a sample point 0 of $X^2 - 1$ with a sample point $(Y^2 - 1,] - 2, 0[)$ of $Y^2 - 1$ then we obtain a 2-dimensional sample vector. The sign of the polynomial $X^2 + Y^2 - 1$ evaluated at this sample vector is 0 . Hence, we deduced that the constraint $X^2 + Y^2 - 1 = 0$ is satisfiable and that $\left(\begin{array}{c} 0 \\ (Y^2 - 1,] - 2, 0[) \end{array} \right)$ is a solution.

In general, we lift all sample points and calculate the new Y -sample points for each substituted polynomial. With the complete set of sample vectors, we are able to decide whether the polynomial constraints $X^2 + Y^2 - 1 \diamond 0$ are satisfiable for each $\diamond \in \{<, =, >\}$.

The above presented construction can be generalized to an algorithm for multivariate poly-

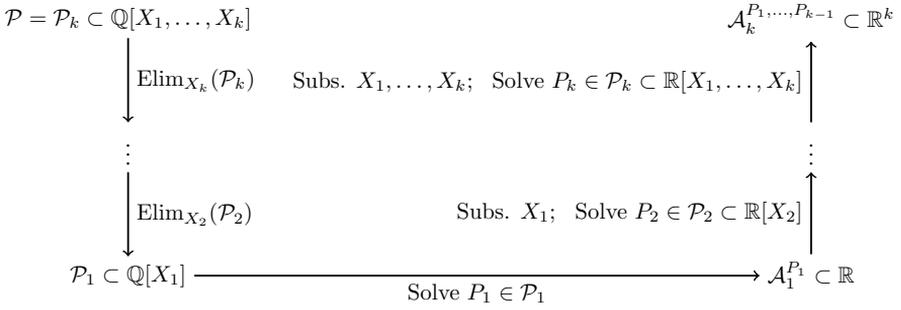


Figure 1: Projection phase and lifting phase.

nomial sets. We denote the polynomial elimination sets by $\mathcal{P}_1, \dots, \mathcal{P}_{k-1}$ where the original set \mathcal{P} is also referred to as \mathcal{P}_k . The set of i -dimensional sample vectors for a certain sequence of substituted polynomials P_1, \dots, P_i is denoted by $\mathcal{A}_i^{P_1, \dots, P_i}$ (c.f. Figure 1).

In general, for a finite set of polynomials \mathcal{P} in k variables with rational coefficients the union of all sample vectors represents a *Cylindrical Algebraic Decomposition*(CAD) of \mathbb{R}^k which is *adapted* to \mathcal{P} .

The number of polynomials in the output set is bound by $\mathcal{O}((|\mathcal{P}| \cdot d)^{3^{k-1}})$ where d is the maximal degree in \mathcal{P} and k is the number of variables. Thus, this is a polynomial bound in d and $|\mathcal{P}|$ but a doubly exponential bound in k . To decrease the number of computations, it is possible not to lift all sample points in a dimension-wise fashion but to choose single samples and to lift them to dimension k in a depth-first manner. If the computed sample vectors satisfies the constraints, the computation is terminated with lesser computation steps. However, this does not decrease the computation time if the constraint is unsatisfiable.

4 Implementation and Conclusion

The CAD algorithm is implemented within computer algebra systems like Reduce [DS96] or Mathematica [Str] and closed-source stand-alone implementations like QEPCAD [Bro03]. We provide an open-source C++ library called GiNaCRA [LÁ11] with an implementation of the CAD algorithm for interval representations. In contrast to other tools, our implementation

- uses the interval representation to provide solutions in case of satisfiability,
- uses an adapted projection technique to support the insertion of new polynomial constraints to a CAD without recomputation of the whole elimination set,
- is suited for embedding in SAT Modulo Theorie (SMT) solvers, making use of

highly tuned SAT solvers and enabling the combination of different decision procedures.

Due to its doubly exponential complexity, the CAD method can handle constraints with just a few variables. A possible work-around is to combine the CAD algorithm with the incomplete but more efficient virtual substitution method in SMT solving [CA11]. Our goal is to increase this applicability bound by combining the advantages of different approaches in the SMT solving context. E.g., it is possible to reduce the number of variables with virtual substitution and to use the CAD algorithm efficiently for lower-dimensional polynomials with higher degrees.

Acknowledgement

I would like to thank Erika Ábrahám and Ulrich Loup for supporting me during the development of this Bachelor thesis.

References

- [BPR06] S. Basu, R. Pollack, and M.F. Roy. *Algorithms in real algebraic geometry*, volume 10. Springer-Verlag New York Inc, 2006.
- [Bro03] Christopher W. Brown. QEPCAD B: A program for computing with semi-algebraic sets using CADs. *SIGSAM BULLETIN*, 37:97–108, 2003.
- [CA11] Florian Corzilius and Erika Ábrahám. Virtual substitution for SMT-solving. In *Proceedings of the 18th international conference on Fundamentals of computation theory, FCT'11*. Springer Verlag, 2011.
- [DS96] Andreas Dolzmann and Thomas Sturm. REDLOG Computer Algebra Meets Computer Logic. *ACM SIGSAM Bulletin*, 31:2–9, 1996.
- [LÁ11] U. Loup and E. Ábrahám. GiNaCRA: A C++ library for real algebraic computations. *NASA Formal Methods*, pages 512–517, 2011.
- [Mis93] B. Mishra. *Algorithmic algebra*. Springer, 1993.
- [Str] Adam Strzebonski. Cylindrical Algebraic Decomposition. <http://mathworld.wolfram.com/CylindricalAlgebraicDecomposition.html>.
- [Tar48] A. Tarski. *A decision method for elementary algebra and geometry*. RAND report. RAND Corp., 1948.

Distant Supervised Relation Extraction

Marcel Ackermann

TU Darmstadt
ackermann@tk.informatik.tu-darmstadt.de

Abstract: In this paper we discuss a new approach to extract relational data from unstructured text without the need of hand labeled data. So-called distant supervision scales large amounts of web data and therefore fulfills the requirement of current information extraction tasks. As opposed to supervised learning, we train generic, relation- and domain-independent extractors on the basis of data base entries. We use Freebase as a source of relational data and a Wikipedia corpus tagged with unsupervised word classes. In contrast to previous work in the field of distant supervision, we do not rely on preprocessing steps that involve supervised learning, having the advantage of staying independent of language or domain while scaling to web size corpora.

1 Introduction

Banko & Etzioni (2008) define Relation Extraction (RE) as “the task of recognizing the assertion of a particular relationship between two or more entities in text”. Extracting relational facts from unstructured text is a highly relevant topic, as it has many applications, such as Information Retrieval, Information Extraction (IE), Text Summarization, Question Answering, Paraphrasing and Word Sense Disambiguation. The common practice is to use supervised methods, learning extractors for entities and their relations from hand-labeled corpora, using lexical, syntactic and semantic features. At this, any kind of resources are employed, such as large corpora, dictionaries or lexical-semantic resources like WordNet. The main issues with these approaches are expensive and time consuming labeling and domain-dependence.

When using unsupervised methods for relation extraction, the resulting relations are hard to map on existing knowledge bases. In addition to this, most Open IE systems use subcomponents, such as a tagger, parser or Named Entity Recognizer (NER) that are trained with supervised machine learning.

Mintz (2009) present an alternative approach – distant supervision – combining advantages from the above methods. They train a logistic regression classifier on a large amount of features, obtained from sentences containing instances from the Freebase database¹. Their approach is based on the distant supervision assumption (Riedel, 2010): “If two entities participate in a relation, all sentences that mention these two entities express that relation.” Refer to Mintz (2009), Hoffmann (2010) and Yao (2010) for more information.

¹ www.freebase.com, an open triple store for real world entities and their relations

We follow the setup of Mintz, while staying completely unsupervised. We use the unsupervised parts-of-speech (POS) tagger of Biemann (2009) and train a distant supervised NER. This results in a domain-independent system that scales to Web size corpora, and the output directly maps to canonical names of existing relations as defined in our database or ontology.

2 Method

We use the same terminology as Mintz (2009). We use the term ‘relation’ in the meaning of an ordered, binary relation between two entities. We call these entities, entity A and entity B, allowing us to refer to the order. We refer to individual ordered pairs in the target relation as instances. We use the relation `person/place_of_birth` as an example to explain the steps of our method:

1. Read Freebase² data, create training and test splits for validation.
`person/place_of_birth: (Juan Ramón Jiménez, Moguer)`
2. Grab all sentences in Wikipedia which contain exactly one part A and one part B, not necessarily from the same instance of 1.
`Juan/6 Ramón/6 Jiménez/10 was/222 born/3 in/3 Moguer/8`
3. Train a named entity recognizer (Stanford NER³), which is able to tag entities with three labels entity A, entity B, other O.
`Juan/A Ramón/A Jiménez/A was/O born/O in/O Moguer/B`
4. Train a classifier that separates sentences containing the target relation from those that do not.
5. Use the classifier of 3 and the NER of 4 to find all sentences in Wikipedia that contain the target relation and extract both relation parts.
6. Compare the found relation instances of 5 with the held-out data of 1.

In order to decide, whether or not a sentence contains the target relation, we tried two different approaches. In the first variant we let the RR decide on the relation before tagging entities, in the second we tagged all sentences with our NER and then classified with the RR.

Classifying before tagging has the advantage that classifying is significantly faster than tagging. Yet, the disadvantage is that we can only use words and POS-tags as features, not the presence/absence of named entities. Our features are n-grams of words, tags and letters. The second classifier also takes the classes assigned by the NER as features. This improves the classification but strongly depends on the quality of the NER. The output of our second classifier is a list of patterns, which indicate a high probability that the sentence contains the target relation in case of a pattern match. As features for patterns, we use word and pos n-grams under the condition that they appear before A, between A and B, after A as well as for the case when B appears before A. Regarding only unigrams we get 12 features: $\{\text{word, pos}\} \times \{\text{AB, BA}\} \times \{\text{before, within, after}\}$

² www.freebase.com, an open triple store for real world entities and their relations

³ <http://www-nlp.stanford.edu/software/CRF-NER.shtml> - used features: all combinations of words, tags, positions, classes and their ngrams

We use the 25% best n-grams per section to build patterns in the form $X(A|B)Y(A|B)Z$, joining word and tag n-grams. We allow only one slot to be empty and if there is a pattern with gaps that is part of a pattern without, we only take the one without gaps. After that, we count the support in the positive sentences.

The final step is to combine word/pos-ngrams respectively the patterns with the named entity recognizer to extract entities from positive classified sentences.

3 Evaluation

At first we evaluate the two subsystems and then provide statistics for overall system performance. Our test data is for the complete evaluation the person/place_of_birth relation. Although we chose one of the larger Freebase relations, it turned out that the representation in the Wikipedia data is very low: Out of 40k test instances in Freebase, only 28 appear in our corpus of 10 million Wikipedia sentences. We add to this the 182 sentences in which A or B appear on their own. The F1 measure – defined by the harmonic mean of Precision ($\frac{\#correct}{\#found}$) and Recall ($\frac{\#correct}{\#contained}$) – of the NER is 0.116, the RR with patterns performs at 0.236 and the RR with n-grams at 0.315.

In the evaluation of the NER, a true positive is the case in which an entity is found completely, the true negatives are the ones in which all words are correctly labeled as other O. Although we biased the NER to use A and B more often, we have very low recall and precision values. For finding single parts A and B this might be suitable, but in our case the low recall is the reason why we do not find any correct Freebase instance at all.

In order to avoid the dependence on the NER, the RR is evaluated using the correct labels. We aimed at and reached a high recall in RR, as the idea was to have two independent systems with high recall – resulting in a combined system with medium recall. The difference in the F1 measure between the two RR variants can be explained by their different feature structure. In the patterns case the entities have to be matched before and then a match in unigrams in the right parts is required. In the other case only unigram, bigram and trigrams are matched. The overall RE system performance cannot be evaluated as the data we used is too sparse and our subsystem performance is too low.

4 Conclusion

In this paper we presented a novel approach to distant supervision. Current state-of-the-art distant supervision systems exploit plenty of features gathered by supervised systems, such as POS tagger, NER and parser. We developed a completely unsupervised system, based on state-of-the-art unsupervised POS tagging (Biemann, 2009). Our contribution is a distant supervised NER and Relation Recognizer. Combining those leads to a Relation Extraction system that generalizes enough to work on out-of-domain corpora.

Reference

- Banko, M. & Etzioni, O., 2008. The tradeoffs between open and traditional relation extraction. *Proceedings of ACL*.
- Biemann, C., 2009. Unsupervised Part-of-Speech Tagging in the Large. *Res. Lang. Comput.* 7.
- Finkel, J.R., Grenager, T. & Manning, C., 2005. Incorporating non-local information into information extraction systems by gibbs sampling. *ACL-05*, pp.363-70.
- Hoffmann, R., Zhang, C. & Weld, D.S., 2010. Learning 5000 Relational Extractors. *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics (ACL '10)*, pp.286-95.
- Lafferty, J., McCallum, A. & Pereira, F., 2001. Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data.
- Mintz, M., Bills, S., Snow, R. & Jurafsky, D., 2009. Distant supervision for relation extraction without labeled data. *Proceedings of ACL-IJCNLP 2009*.
- Morgan, A.A., 2004. Gene name identification and normalization using a model organism database. *J. of Biomedical Informatics*, 37, pp.396-410.
- Riedel, S., Yao, L. & McCallum, A., 2010. Modeling Relations and Their Mentions without Labeled Text. *ECML PKDD 2010, Part III, LNAI 6323*, pp.148-63.
- Rink, B. & Harabagiu, S., 2010. UTD: Classifying Semantic Relations by Combining Lexical and Semantic Resources. *Proceedings of the 5th International Workshop on Semantic Evaluation*, pp.256-59.
- Rozenfeld, B. & Feldman, R., 2008. Self-supervised relation extraction from the web. *Knowledge and Information Systems*, pp.17-33.
- Snow, R., Jurafsky, D. & Ng, A.Y., 2005. Learning syntactic patterns for automatic hypernym discovery. *NIPS 17*, pp.1297-304.
- Yan, Y., 2009. Unsupervised Relation Extraction by Mining Wikipedia Texts Using Information from the Web. *Proceedings of the 47th Annual Meeting of the ACL and the 4th IJCNLP of the AFNLP*, pp.1021-29.
- Yao, L., Riedel, S. & McCallum, A., 2010. Collective Cross-Document Relation Extraction Without Labelled Data. *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pp.1013-23.
- Zhu, J., 2009. StatSnowball: a Statistical Approach to Extracting Entity Relationships. *ACM 978-1-60558-487-4/09/04*.

Development of a Cheap Chess Robot: Planning and Perception

Billy Okal and Oliver Dunkley

Jacobs University, Automation Group
Campus Ring 1, 28759 Bremen
{b.okal, o.dunkley}@jacobs-university.de

Type of Work: Bachelor Thesis

Supervisor: Prof. Dr. Andreas Nüchter

1 Introduction

Interest in chess playing robots started as early as 1770 when Wolfgang von Kempelen unveiled what appeared to be an automatic chess player, that astonished and fascinated crowds, even after its fake nature was discovered in the 1820s [Sch99]. Chess playing robots caught more attention when the Deep Blue computer beat Gary Kasparov in a chess game [NN97], making a major breakthrough in artificial intelligence. Despite the fact that development of robots capable of playing chess has been historically known to be a complex task, major attempts have been made over the years. Classroom efforts such as Groen's lab course in sensor integration [GdBvIS92] are an early example. More recently, the Association for Advancement of Artificial Intelligence (AAAI) organized a small scale manipulation challenge with playing of chess as the competition task¹. Today, many chess playing robots exist in various forms ranging from humanoids to industrial grade manipulators. However, the majority of these rely on either modified chess boards in terms using present colors and sizes to simplify the scene interpretation problem and/or modified chess pieces to aid the robot sensors, or use fixed vision sensors e.g. overhead mounted cameras with a fixed boards all of which serve to simplify the problem at a cost of flexibility, generality and reproducibility of designed methodology. Furthermore, most of the existing chess playing robots do not use established motion planning algorithms mainly because only a few degrees of freedom are involved and hence solving Inverse Kinematics equations to locate end effector poses is not considerably difficult although it could get easily expensive. We develop a system based solely on a single off the shelf webcam as a camera sensor, a cheap 4-DOF robotic arm assembled from standard robotics kits as an actuator and standard chess components. The robot arm was powered by Dynamixel AX-12 servos. We also develop a software solution that is both modular and generic allowing for

¹<http://aaai-robotics.ning.com/forum/topics/icra2010-and-aaai2010>

easy interchange of major components such as the perception system, chess engine and motion planner.

2 Methods

To have a robot play a game of chess there are two main problems that need to be solved namely, perceiving the environment in which the chessboard and chess pieces exist and interpreting the state of the game from such perception sensor data and computing a feasible collision free path for moving a given chess piece on the chessboard. We do not delve into strategies and game logic for deciding which piece to move and use a standard chess engine to generate moves based on the given state of the game. For building a modular system, we use of the Robot Operating System (ROS) [QCG⁺09] for coordination and communication. We developed a system that is depicted in Figure 1. We solve the perception and planning problems using the procedures described in the sections that follow.

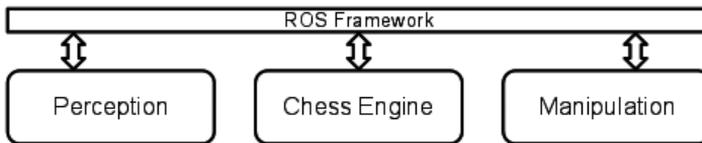


Figure 1: System Overview

2.1 Perception

We use of various first principles like approaches to solve the perception problem. In particular, we search for the chessboard in the scene at every game instance. We take an image of the scene at every game step and compare successive images to determine what has changed i.e. which moves have been played. We use image processing techniques such as line detection, Canny edge detection [SHB07] and image differences to determine the location of the four corners of the chessboard as well as the location of changes on the board and transform the resulting chessboard into a known frame by perspective projection. Once we get the positions of the changes on the chessboard using its size information and scaling, we determine which move to make by querying a standard chess engine. For implementation of the techniques, we use of the OpenCV² library.

²<http://opencv.willowgarage.com>

2.2 Manipulation

With the new move determined, we again use chessboard size information to determine the precise locations to move a given piece, which provides the goal and start pose for the planning process. We use a seasoned motion planning algorithm called BiRRT (Bi-directional Rapidly Exploring Random Trees) [Dia10] a variant of RRT which provides a probabilistically complete algorithm [LaV98]. We compare this method to that of computing the Inverse Kinematics of the end effector (gripper) each time with the belief that the BiRRT motion planning algorithm would take advantage of the topological compactness of the search space as most of the goals are very close in this scenario. Either of the procedures gives us a trajectory to follow. We then execute the computed trajectories on the robot taking into account further joint considerations and reachability of the workspace. We added some constraints into the planning system so that the chess pieces are moved in an 'upright' manner throughout the manipulation process and also to filter the found solutions. For implementation, we make use of OpenRAVE [Dia10] motion planning suite which comes with a set of motion planning algorithms including BiRRT and open interfaces for using the planners.

3 Results

Experiments with the perception procedures showed reliable chess move detection with real-time performance comparable to previous cases with modified chessboards and/or modified chess pieces or fixed cameras. The trajectories generated using seasoned motion planning algorithms were also of better quality compared to those generated by simple Inverse Kinematics. We also noticed that the BiRRT algorithm took considerably shorter time for planning compared to solving Inverse Kinematics for locating the end effector pose. Figure 2 shows the setup of the experiments and a sample trajectory.

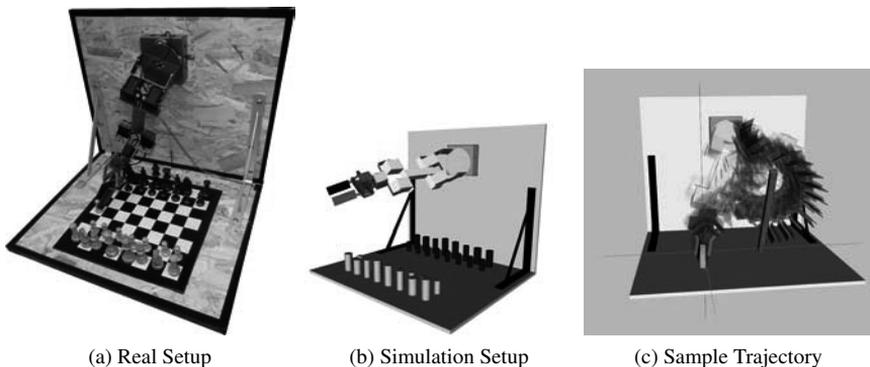


Figure 2: Real and simulated setup and a sample trajectory

Further media including video material depicting the results can be found on our groups

video archive at <http://www.youtube.com/user/AutomationAtJacobs> and further details and discussion of the work is provided in a corresponding thesis document.

4 Conclusion

We demonstrated that the constraint of using a modified chessboard and/or pieces can be relaxed without greatly compromising performance. We have also demonstrated that the constraint of having a fixed camera directly above the chessboard can be done away with while still enjoying reasonable real-time performance. Use of seasoned motion planning algorithms with various optimizations and constraints was also found to produce smoother paths for manipulation chess pieces. We showed all of these using very simple hardware and a limited amount of time. We also noticed that our imprecise hardware and limited DOF posed great difficulties even for seasoned planning algorithms making realization of a full successful game not possible due to overheating of motors. The limitation in DOF can however be easily averted at cheap cost by adding just two more DOF to the robot arm.

References

- [Dia10] R. Diankov. *Automated Construction of Robotic Manipulation Programs*. PhD thesis, Carnegie Mellon University, Robotics Institute, August 2010.
- [GdBvIS92] F. C. A. Groen, G. A. der Boer, A. van Inge, and R. Stam. A chess playing robot: lab course in robot sensor integration. In *Instrumentation and Measurement Technology Conference, 1992. IMTC '92., 9th IEEE*, pages 261–264, May 1992.
- [LaV98] S.M. LaValle. Rapidly-exploring random trees: A new tool for path planning. Technical Report TR-98-11, Iowa State University, Computer Science Department, 1998.
- [NN97] M. Newborn and M. Newborn. *Kasparov Vs. Deep Blue: Computer Chess Comes of Age*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 1997.
- [QCG⁺09] M. Quigley, K. Conley, B. P. Gerkey, J. Faust, T. Foote, J. Leibs, R. Wheeler, and Y. N. Andrew. ROS: an open-source Robot Operating System. In *ICRA Workshop on Open Source Software*, 2009.
- [Sch99] S. Schaffer. Enlightened Automata. In *The Sciences in Enlightened Europe*. (Eds. William Clark, Jan Golinski, and Simon Schaffer), pages 126–165, 1999.
- [SHB07] M. Sonka, V. Hlavac, and R. Boyle. *Image Processing, Analysis, and Machine Vision*. Thomson-Engineering, 2007.

Automatisierte Erkennung von Influencern im Social Web

Dominik Grzelak
E-Mail: dgrzelak@hs-lausitz.de

IBM Deutschland Research & Development GmbH
Schoenaicher Str. 220, 71032 Boeblingen

Hochschule Lausitz
Großenhainer Str. 57, 01968 Senftenberg

Abstract: Dieser Artikel ist ein Auszug der Bachelor-Arbeit, die während des Praktikumsseinsatzes bei IBM (Böblingen), in dem Produktentwicklungsteam von IBM Cognos Consumer Insight, entstanden ist. Diese Arbeit beschäftigt sich mit der Entwicklung eines Verfahrens für die automatisierte Erkennung von Schlüsselnutzern für bestimmte Themen im Social Web – Hauptaugenmerk wurde dabei auf Mikroblogs (in diesem Fall Twitter) und Blogs gelegt. Der erste Schritt bestand darin, diese zwei Technologien in Hinblick auf ihre Eigenschaften zu untersuchen und darauf geeignete Formalismen, Methoden und Scores zu entwerfen, um Schlüsselnutzer zu erkennen. Auf diesen Erkenntnissen aufbauend, wurde eine prototypische Implementierung entwickelt, die unter den technologischen Gegebenheiten von IBM Cognos Consumer Insight arbeitet.

Aufgrund vertraulicher Inhalte, kann in diesem Artikel nur ein grober Überblick, der entwickelten Methoden abgebildet werden.

1 Einführung und Motivation

Social Media wird derzeit von vielen Unternehmen als zusätzlicher Kommunikationskanal für das Reputation Management oder die Kundenakquirierung genutzt sowie für die Steigerung der Markenaffinität und Bekanntheit. Durch den zunehmenden globalen Wettbewerb im Social Web, müssen Unternehmen effektivere als auch effizientere Methoden anwenden, um die Werbebotschaften besser vermitteln zu können.

Die Integration von Werbekampagnen in sozialen Netzwerken ist längst wichtiger Bestandteil vieler Unternehmen. Dabei verschenken viele Unternehmen beim viralen Marketing ihr Effizienzpotenzial. Meinungen und Stimmungen von Konsumenten müssen gezielt aufgenommen und ausgewertet werden, um daraufhin angemessene Schritte einleiten zu können. Häufig besteht das Problem darin, dass die Werbebotschaften nicht zielgerichtet und direkt vermittelt werden. Das virale Marketing kann deutlich verbessert werden, indem Interessenvertreter in sozialen Netzwerken identifiziert werden können – Unternehmensvertreter können sich mit diesen in Verbindung setzen und neue Kontakte knüpfen. Social Media Plattformen, wie Blogs, Twitter, Internetforen oder Facebook, bieten somit ein immer größer werdendes Potenzial, die Stimmungen, Meinungen und Gedanken anderer Nutzer zu beeinflussen.

2 Social Media Analyse

Wie der Name schon ausdrückt, befasst sich die Social Media Analyse mit der Auswertung von Social Media, indem die zwei Hauptelemente *Inhalt* (User Generated Content) und das

soziale Netz¹ untersucht werden (s. Abb. 1). Durch die Social Media Analyse lassen sich

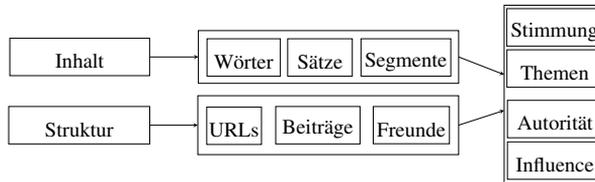


Abbildung 1: Vorgehensweise einer Social Media Analyse nach [Hur08]

dann u. a. folgende Sachverhalte bestimmen:

- Trends und Aktivitäten aus Benutzerunterhaltungen erkennen
- Demographie der Benutzer analysieren und Hauptstandorte von Aktivitäten entdecken
- Positive/Negative Meinungen zu einem Thema finden
- **Schlüsselnutzer oder Meinungsführer für ein Thema entdecken**

Typen von Influencern

Ein Influencer als einzelne Person wird vom Social Media Advertising Consortium (SMAC) wie folgt definiert: „An advocate or detractor who can impact sentiment and holds a level of credibility with a group or connection“ [Con11] Ein Influencer tritt nicht nur als Befürworter auf, sondern fungiert auch als Kritiker. Löst man sich von der Vorstellung einer einzelnen Person, kann unter diesem Begriff auch eine Gruppe aufgefasst werden. Zwei oder mehrere Autoren eines Blogs oder einer Nachrichten-Website können als Influencer agieren. Wird der Begriff Influencer aus einem anderen Blickwinkel betrachtet, kann die Gesamtheit einer Social Media Site als Influencer angesehen werden. Dies kann bspw. ein Blog oder Forum sein, in denen Personen oft und gerne für Ratschläge oder Diskussionen unterwegs sind.

Weitere Typisierung von Influencern sind unter [Ste10] und [Bar11] zu finden.

3 IBM Cognos Consumer Insight

IBM Cognos Consumer Insight ist eine Social Media Analyse Software, die „Marketingexperten zur Förderung der Kundenbeziehungen durch Analysieren der Meinungen, Affinität und künftig relevanter Themen auf Websites sozialer Netzwerke“ [CCI11] dient.

¹Das soziale Netz entsteht durch die Kommunikation mit anderen Nutzern auf Social Media Plattformen (vgl. hierzu [Bak10]).

Für eine Analyse werden Einträge und Kommentare aus Blogs, ganze Unterhaltungen aus Foren oder Nachrichten aus öffentlichen Gruppen sozialer Netzwerke automatisch gesammelt. Für die anschließende Datenanalyse wurde IBM InfoSphere BigInsights eingesetzt.

IBM BigInsights ist eine hadoop-basierende Infrastruktur, die es ermöglicht, die immensen Datenmengen aus dem Internet sehr schnell parallel verarbeiten zu können. Dabei werden die Open Source Komponenten von Apache Hadoop, zusätzlich mit von IBM entwickelten Features erweitert. Die Textanalyse erlaubt es, neue Erkenntnisse aus den Datenmengen zu extrahieren. Dafür wird die Abfragesprache Jaql verwendet.

Jaql ist eine von IBM entwickelte Open Source Abfragesprache. Dabei weist Jaql Ähnlichkeiten zu anderen Abfragesprachen (SQL und XQuery) auf. Die primären Vorteile liegen in der schnellen Auswertung von großen und unstrukturierten Datenmengen. Hierbei werden die Vorzüge der Parallelisierung durch MapReduce Jobs genutzt. Über User-Defined-Functions hat man die Möglichkeit, Jaql-Funktionalitäten bspw. mit Java, Python oder Perl zu erweitern.

4 Schwierigkeiten

Das Vorkommen von Bots² stellt, wie auch in vielen technischen Bereichen, ein Problem dar. Gleichzeitig schloss sich deshalb ein Entwurf und Implementierung einer automatisierten Erkennung von Twitter-Bots an, welches als selbstständiges Modul vor die Influencerberechnung geschaltet wurde. Für Blogs wurde dieser Aspekt nicht betrachtet, da es zum einen nicht in dem zeitlichen Rahmen hätte geschehen können und Spam-Blogs (Splogs), durch die verwendete Methodik in der Arbeit, keinen negativen Einfluss auf die Ergebnisse hatten. Der interessierte Leser sei an dieser Stelle auf [KJF⁺06] verwiesen, welches ein komplexes System zur Erkennung von Splogs präsentiert.

Weitere Schwierigkeiten bestanden darin, die ermittelten Ergebnisse angemessen auf ihre Güte mit anderen Tools zu vergleichen. Nicht alle Online-Dienste hatten Influencer zu den gleichen Themen berechnet, wie es bspw. in der Arbeit der Fall war.

5 Ergebnisse

Für einen Zeitraum von drei Monaten wurden Daten über drei Themenkategorien (aus den Bereichen Mobiltelefonie, Sport und Kosmetik) für Twitter und Blogs gesammelt. Für die Influencerberechnung auf Twitter – wurden insgesamt vier Scores entwickelt, welche die unterschiedlichen Charakteristiken von Influencern auf Twitter quantifizieren. Auf Grundlage der Datenmenge wurde jeder Score zusätzlich kalibriert; außerdem wurde ein Verfahren zur Detektion von Bots auf Twitter entwickelt. Die Methode zeigte eine hohe Klas-

²Bots werden eingesetzt, um wiederkehrende Aufgaben selbstständig durchführen zu lassen. Die primäre Aufgabe von Twitter-Bots besteht also darin, Tweets automatisch zu veröffentlichen.

sifikationsgüte: Der Anteil der korrekt klassifizierten Bots betrug 93 %. Insgesamt wurde eine Übereinstimmung der gefundenen Influencer mit den Ergebnissen anderer Tools³ (für ähnliche oder gleiche Themenbereiche) von bis zu 90 % festgestellt. Die zwei zuletzt genannten Maßnahmen trugen zur Verbesserung der Ergebnisqualität bei.

Bei der Influencerberechnung in der Blogosphäre, wurden sieben Methoden mit unterschiedlicher Gewichtung verwendet, die ein Ranking von Blogs für ein bestimmtes Thema erlaubten. Die Klassifikation durch den Algorithmus wurde mit den Ergebnissen der zwei etablierten Dienste – Konector und Technorati – verglichen. Es konnte für alle Themen eine Übereinstimmung von bis zu 70 % festgestellt werden.

6 Zusammenfassung und Ausblick

Schließlich kann gesagt werden, dass sich der Ansatz einer Social Media übergreifenden Influencer-Analyse lohnt. Wird bspw. auf Twitter und einem Blog der gleiche Nutzer festgestellt, so besitzt dieser eine größere Influence – als ein Nutzer, bei dem dies nicht der Fall ist. Hierfür muss für einen Influencer im Social Web eine geeignete Datenstruktur geschaffen werden, die alle verwendeten Social Media Plattformen auflistet.

Der Ansatz für den Mikroblog Twitter, kann zudem problemlos auf andere Mikroblogging-Dienste⁴ angewendet werden.

Literatur

- [Bak10] Thomas Sammer Bakk. Social Media Analysis: Einsatzmöglichkeiten im Unternehmen anhand eines Fallbeispiels. Diplomarbeit, Jan 2010. <http://en.calameo.com/read/0003261003d3e7201521c>.
- [Bar11] Lisa Barone. The 5 Types of Influencers On The Web. Online-Quelle, Mar 2011. <http://smallbiztrends.com/2010/07/the-5-types-of-influencers-on-the-web.html>.
- [CCI11] CCI. IBM Cognos Consumer Insight. Online-Quelle, 2011. <http://www-142.ibm.com/software/products/de/de/cognos-consumer-insight/>.
- [Con11] Social Media Advertising Consortium, May 2011. <http://wiki.smac.org/social-consumers/influencer>.
- [Hur08] Matthew Hurst. Social Media Analysis: Present and Future. Online-Quelle, 2008. <http://www.slideshare.net/matthewhurst/social-media-analysis-present-and-future>.
- [KJF⁺06] Pranam Kolari, Akshay Java, Tim Finin, Tim Oates und Anupam Joshi. Detecting Spam Blogs: A Machine Learning Approach. In *Proceedings of the 21st National Conference on Artificial Intelligence (AAAI 2006)*. Computer Science and Electrical Engineering, University of Maryland, Baltimore County, July 2006.
- [Ste10] Jim Sterne. *Social Media Metrics: How to Measure and Optimize Your Marketing Investment (New Rules Social Media Series)*. Wiley, 2010.

³Twitalyzer (<http://twitalyzer.com/>), Klout (<http://klout.com/>)

⁴Beispielsweise für *Identi.ca* – dieser besitzt die gleichen Features, wie auch andere Mikroblogging-Dienste.

Implementing an AI for Mayagarden, a strategic 2 player game

Felix Kerger

contact@felix-kerger.de

1 Introduction

Artificial Intelligence (AI) is an important part of many computer games and can greatly enhance the fun and immersion of a game if done well. Depending on the complexity of the game and the proficiency of the game designer in computer programming, one of several paradigms for the implementation of the AI strategy has to be chosen. Possible paradigms are e.g. logic-based approaches, behaviour trees[1], neural networks[2] or other machine learning algorithms[3]. Using the example of the Mayagarden game, which was developed as a student project for a game design class, we discuss several paradigms and lay out our decision process, which led to a realization using a finite state machine (FSM).

2 The game

In the game two Aztec gods are fighting for dominance over a planet through planting seeds in flowerbeds. Each god can plant a flower using seeds he carries, moreover he can harvest a flowerbed to retrieve new seeds. But he can not carry more than five seeds at the same time. After planting, the flowerbed grows until it reaches its highest level of 5 flowers, if harvested it will generate up to five seeds, one for each flower. If a god is over an enemy field he can sacrifice a seed to destroy one flower. This implicates that if a field is fully grown it can not be destroyed and planted at the same time, because the five seeds a god carries can only destroy the five flowers in the flowerbed. So the flowerbed will be empty and the god has to retrieve new seeds to conquer the flowerbed, leaving the other god time to reclaim the field for itself. With this description it is clear that there are several actions a god can do, turn left and right, seed a plant, harvest a flowerbed or sacrifice a seed to destroy an enemy plant. Because planting and sacrificing are context dependant and mutual exclusive, they are mapped to the same action/key.

3 Possible approaches

There are several approaches which could be used to implement the AI for this game. [1] lists several of these and was used as the main source of information during the implementation of the AI, as source for rule based AI another book was used[4].

3.1 Rule based decisions

A rule based way to implement an AI for Mayagarden would be to encode the constraints of the game and the winning strategy into Propositional logic, see [4]. The game would provide the values for symbols and a logic solver could be used to determine which actions to take, based on the truth value of rules. To decide if the AI needs more seeds a simple rule can be used. 'Sx' is the Symbol that is true if the player has x seeds. When S0 is true the player does not have any seeds, if S1 is true the player has one seed and so on. The AI needs to pick up new seeds when it does not has any seeds left, this could be expressed with the following rule.

$$S0 \Rightarrow NeedSeeds$$

Defining a rule which selects a field to use for harvesting is a lot more complex and needs a lot of symbols. Each field has two states, one defines to which player the field belongs and the other how many flowers are on the field. We define the symbol Hp,x for each field, where $x \in \{1, 2\}$, and $p \in \{1, \dots, 12\}$. H1,1 means that player 1 can harvest field 1. Pp,x with $x \in \{1, 2\}$, and $p \in \{1, \dots, 12\}$ means that field p belongs to player x. Lp,l with $p \in \{1, \dots, 12\}$ and $l \in \{1, \dots, 5\}$ is true when the field at position p has l flower(s).

$$Hp, x \Leftrightarrow Pp, x \wedge (Lp3 \vee Lp, 4 \vee Lp, 5)$$

This rule is a lot more complex and selecting a field to attack is not simpler. Describing the complete game with rules and using a logic solver would work, but would result in a rather complex system of rules. Such a rule base is hard to maintain and hard to extend with e.g. settings of game difficulty. Therefore, we decided to take a different route.

3.2 Behaviour trees

Behaviour trees are, like the name suggest, a kind of tree. There are several ways to implement behaviour trees, but the core principle is always the same. Leaf nodes are actions which should be executed and inner nodes can either be a selector or a sequence. A selector selects one of its possible children based on a condition and a sequence executes all its children in a defined order. With this simple setup, which might be extended when needed, all AI behaviours can be implemented. With a GUI behaviour trees are an excellent choice to enable people without programming knowledge, like game designer, to create AIs for games. But this involves a lot of ground work, which needs to be done first, including creating the base classes for the tree and all the interface classes needed to access the data

stored in game objects. With a simple AI as needed for Mayagarden, this effort did not seem worthwhile, especially when no Game Designer was involved in creating the AI and changing the AI was not very likely once it was implemented.

3.3 Finite State Machine

Another way to implement an AI is to use Finite State Machines(FSM). A FSM is described by its transition function, which produces an output and changes its current state, based on the previous state and an input. Creating FSMs can be easier than other approaches if it is clear which states are needed and which transitions exist. This was the case with Mayagarden due to its simple nature in game design. The downside is that most FSMs must be implemented directly in source code or with a complex GUI to manage states, transactions and output actions. This makes the FSMs approach harder to use for game designers and games where the AI has to change depending on the level or difficulty setting.

4 Evaluation

An AI implemented with a FSM proves to be effective and almost impossible to beat for this game. Without a pause between two plant or harvest actions, it is not possible to win against the AI. If the player is good enough, the game will go on forever. When a small time delay between each harvest or plant action is added, the AI is still a strong opponent but can be beaten by a good player if he plays in such a way that he always has more fields than the AI when a preset time limit is reached, in this test 5 minutes. Another interesting test for the AI is to play against itself. If both AI player do not pause between two harvest or plant actions, the game ends in a tie after the time limit has been reached. Now if one AI player has a delay between two actions the result changes depending on the length of the delay. If the delay is 0.8 seconds or greater, the player without a delay wins after the time limit is reached. If the delay is 12.8 seconds or greater, the player without a delay wins in under 5 minutes.

5 Conclusion

This result shows that with a rather simple FSM it was possible to implement an AI for this small game. All in all this approach proves that using a simple construct can result in an artificial intelligences which can be used instead of human players, without the implication that they are weaker, depending on the complexity of the game. Of course the fun of playing the game side by side with a friend can not be recreated with a software solution.

References

- [1] Ian Millington, *Artificial Intelligence for Games*. Morgan Kaufmann, 2005, ISBN-10: 0124977820
- [2] David J. C. MacKay, *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press; 10th edition (2011), ISBN-13: 9780521642989
- [3] Christopher M. Bishop, *Pattern Recognition and Machine Learning*. Springer (2006), ISBN-13: 9780387310732
- [4] Stuart Russell, Peter Norvig, *Artificial Intelligence: A Modern Approach*. Prentice Hall; 3 edition (December 11, 2009), ISBN-10: 0136042597

3D Objekt Kategorisierung in häuslichen Umgebungen

Christian A. Mueller

christian.mueller@smail.inf.h-brs.de

Aufgaben in der Service Robotik, die einem Roboter gestellt werden, involvieren häufig Objekte – beispielsweise müssen Objekte gesucht, erkannt oder auch manipuliert werden. Daher ist es notwendig, dass ein Service Roboter mit einer Objektwahrnehmungskomponente ausgestattet ist, die objektrelevante Informationen – auch unter schwierigen realen Bedingungen wie in häuslichen Umgebungen – effizient extrahiert. Neben Objektwahrnehmungen wie die Objekterkennung, welche die Objektidentität analysiert, ist die Objektkategorisierung eine Art, die für bekannte und unbekannte Objekte die zugehörige Kategorie feststellt. Dabei werden charakterisierende Eigenschaften basierend auf der Objekterscheinung oder -form von Kategorien extrahiert um diese später in dem zu kategorisierenden Objekt wieder zu erkennen. In dieser Arbeit sind Kategorien von Interesse, die in häuslichen Umgebungen auftreten wie z.B. Tasse, Dose, Box, Flasche, Teller oder Schüssel. Durch das Assoziieren einer Kategorie zu einem Objekt lassen sich weitere Informationen rückschließen, z.B. im Fall bei einer Tasse: ist beweglich, kann Flüssigkeiten behalten, das ungefähre Gewicht oder auch die optimalen Greifpositionen. Solche Annahmen bzw. Informationen über ein detektiertes Objekt in der Umgebung können genutzt werden um angemessen mit dem Objekt zu interagieren oder auch um höhere Aufgabenstellung, die über die einfache Objektwahrnehmung hinausgehen, zu bewältigen.

In der vorgestellten Arbeit wird eine Pipeline präsentiert von der Detektierung von Objekten in häuslichen Szenen über die Objektdeskription bis hin zum Lernen von Vorhersagemodellen, die für die Kategorisierung der detektierten Objekten angewandt wird. Die Szene ist durch eine 3D Punktwolke repräsentiert, die durch eine 3D Kamera von einer realen Szene erfasst worden ist. Folgende Eigenschaften des Systems fallen besonders in den Fokus: *Robustheit*, *Skalierbarkeit* und *Anwendbarkeit*. D.h., die Robustheit der Kategorisierung bezüglich des Rauschens in den Kamerainformationen; Skalierbarkeit bezüglich der effizienten Erweiterung von neuen Kategorien; Anwendbarkeit des Systems auf einem Service Roboter, welches eine kurzen Antwortzeit der Kategorisierung erfordert.

Um in komplexen häuslichen Szenen objektrelevante Informationen zu extrahieren wird ein Objekt-Detektor angewandt, der auf der Extraktion von *Flächennormalen* [HHRB11] und *RANSAC* (Random Sample Consensus) [FB81] basiert. Der Detektor ist in der Lage Objektkandidaten auf unterschiedlichen Flächen in der Szene zu erkennen wie beispielsweise auf Regalen – jeder Kandidat wird als eine 3D Punktwolke repräsentiert.

Im nächsten Schritt wird für jeden Objektkandidat die Objektfläche aus der jeweiligen Punktwolke rekonstruiert. Dieser Schritt spielt eine besonders wichtige Rolle, da die rekonstruierte Oberfläche – in Form eines Netzes – die geometrischen Eigenschaften des

Objekts widerspiegelt, die in den folgenden Schritten extrahiert werden, um die Kategorie für das Objekt festzustellen. Standard Triangulierungstechniken verbinden die Punkte in der Wolke auf eine Weise, dass ein delaunay-artiges Netz entsteht. Dabei ist zu beachten, dass das Rauschen in der Punktwolke ebenfalls in einer solchen Rekonstruktion der Oberfläche involviert wird. Im Gegensatz dazu wird in dem vorgestellten Ansatz zur Oberflächenrekonstruktion die Punktwolke als eine Verteilung von Punkten im drei Dimensionalen Raum betrachtet. Dabei wird der Fokus auf das Lernen der Verteilung diese Punkte im Raum gelegt um markante Eigenschaften der Objektform – speziell die *Topologie* – des Objekts, welches durch die Punktwolke repräsentiert, abzubilden. Eine unüberwachte Lernmethode (Unsupervised Learning), nämlich ein modifiziertes *Growing Neural Gas* (GNG) [Fri94, MHP11] wird eingesetzt, um die Verteilung der Punkte in einem *Hebbian Learning* ähnlichen Lernverfahren zu lernen, welches zu einer netzartigen Repräsentation der Punktwolke führt: Prinzipiell werden iterativ zufällige Punkte aus der Wolke selektiert und in das GNG generierte Netz “gefeuert”. Dabei passt sich das Netz sukzessiv an die Verteilung bzw. Topologie der Punktwolke an. Letztendlich repräsentiert ein Knoten im Netz nicht ein Punkt in der Punktwolke sondern eine markante Position in der Verteilung der Punkte in der Wolke. Folgende Eigenschaften dieses Verfahrens können beobachtet werden, dass z.B. das Rauschen in der Punktwolke erheblich gemindert in dem generierten Netz bzw. der rekonstruierten Oberfläche abgebildet wird oder einen reorganisierenden Effekt auf die Verteilung der Punkte. Dieser Effekt führt dazu, dass z.B. strukturell ähnliche Objekte, die jedoch z.B. durch einen höheren oder niedrigeren Rauschanteil in der Punktwolke unterschiedlich sind, dennoch zu einer ähnlichen Oberfläche führen. Weiterhin ist es nicht erforderlich Annahmen über die Objektform zu definieren. Oberflächen von Objekten, die aus unterschiedlichen Formen bestehen (z.B. flachen, konvexen, konkaven oder sphärischen) können durch das modifizierte GNG verfahren rekonstruiert werden.

Für jedes detektierte Objekt wird die Oberfläche durch den GNG-Ansatz rekonstruiert. Im folgenden Schritt wird ein Deskriptor auf das Netz der Oberfläche angewendet, der spezielle Verteilungen [HK06] aus dem Netz extrahiert. Diese Verteilungen bilden einen einzelnen Vektor (Objekt-Deskription Vektor), wodurch das Objekt schließlich kodiert und letztlich repräsentiert wird. Zwei Verteilungen bezüglich *Topologie* und *Krümmung* werden aus dem Netz extrahiert. Die Verteilung bezüglich der Topologie basiert auf der Extraktion der sogenannten *All-Pair-Shortest-Path* Distanzen im Netz. Dabei werden von jedem Knoten die kürzesten Pfade zu allen anderen Knoten im Netz ermittelt. Für die ermittelten Distanzen wird die Wahrscheinlichkeitsverteilung geschätzt. Die Verteilung der Krümmung basiert auf der Extraktion der Flächennormalen für jeden Knoten im Netz. Auch hier wird die Wahrscheinlichkeitsverteilung geschätzt. Schließlich werden beide geschätzte Verteilungen zu dem Objekt-Deskription Vektor konkateniert. Experimente zeigen, dass zwischen verschiedenen Kategorien eine Diskrimination besteht, d.h. Deskriptionen von Objekten, die von ähnlichen Kategorien stammen, sind ähnlich und vice versa.

Der letzte Schritt fokussiert sich auf eine *robuste* Klassifizierung der Objektdeskriptionen zu einer zugehörigen Kategorie, wobei die vorherig genannte *Skalierbarkeit* im Vordergrund steht. Um eine Skalierbarkeit zu begünstigen wird das Konzept des Wörterbuchs (*Dictionary* oder auch *Codebook* genannt) [JT05] aufgegriffen um einen gemeinsamen Merkmalsraum (*Feature Space*) zu erzeugen. Das vorgeschlagene Verfahren zur Generie-

rung des Wörterbuchs basiert nicht auf Standard Algorithmen wie z.B. *K-Means* [Mac67] oder *Mean-Shift* [Che95] Clustering im sogenannten *Bag-of-Words* [NHK09, NJT06] Verfahren. In dem vorgeschlagenen Verfahren wird zuerst für jede Kategorie ein *probabilistisches Neurales Netzwerk* (PNN), das aus mehreren Objektdeskriptionen der jeweiligen Kategorie (Prototypen) besteht, erzeugt. Wird eine Objektdeskription, bei der die Kategorie unbekannt ist (Query), an ein solches PNN geleitet, so wird die Query mit jedem Prototypen des PNNs durch einen Gaussischen Kernel verglichen und die Resultate in einem einzigen Stimulus akkumuliert. D.h. wenn die Query ähnlich zur Kategorie des zu vergleichenden PNNs ist, wird die Magnitude des Stimulus hoch sein, und vice versa. Diese PNNs werden als Feature für das neu vorgeschlagene Verfahren zur Generierung des Wörterbuchs eingesetzt. Das eigentliche Wörterbuch setzt auf das *Random Forest* Konzept [Bre01, MNJ08, SCZ08] auf. Der Wald besteht aus einer Menge von Entscheidungsbäumen, wobei die Knoten in den Bäumen die PNNs repräsentieren. Jeder Baum kann als Wort im Wörterbuch betrachtet werden. Das Wörterbuch wird auf jede Objektdeskription angewandt. Dabei wird die Objektdeskription als Input für jeden Entscheidungsbaum gegeben. Schließlich reagiert jeder Baum mit dem *Stimulus* des am stärksten stimulierten PNN und dem zugehörigen *Label* der Kategorie des PNNs. Diese Reaktionen werden in einem Vektor (Wörterbuch-Deskription Vektor) zusammengeführt. Das Wörterbuch wird zu Beginn mit Objekten von einer Reihe von Kategorien (Baskategorien) trainiert und bleibt danach unverändert, auch wenn neue Kategorien hinzugefügt werden. Dieses Wörterbuchverfahren hat mehrere Vorteile – wie Experimente zeigen. Es begünstigt das Lernen von neuen Kategorien, denn jedes Objekt – stamme es von den Basiskategorien oder von neuen unbekanntem Kategorien – wird auf die selbe Weise durch das Wörterbuch beschrieben, in Form des Wörterbuch-Deskription Vektor, d.h. dieser Vektor liegt immer im selben Merkmalsraum. Weiterhin beinhaltet dieser Vektor unterschiedliche Reaktionen zu Kategorien, bezüglich einer Query. Diese *Mixtur* von unterschiedlichen Kategoriereaktionen können durch ein maschinell überwachtetes Lernen (Supervised Learning) effizient gelernt werden. Eine online *Support Vector Machine* (SVM) [LBRF11] wird als Lerner eingesetzt, der für jede Kategorie ein Vorhersagemodell lernt. Diese Modelle werden später verwendet um Objekte, die durch einen Wörterbuch-Deskription Vektor beschrieben sind, zu klassifizieren. Ein Kreuzvalidierungsfehler von 5,32%, unter Berücksichtigung von 7 Kategorien (Tasse, Dose, Box, Flasche, Schüssel, Teller, Ball), ist erzielt worden. Die durchschnittliche Kategorisierungszeit beträgt ca. 1,8 Sekunden, welche für den Einsatz auf einem Service Roboter ausreichend ist.

Mit diesem vorgestellten Verfahren zeigt sich in weiteren Experimenten, dass weitere Kategorien (z.B. Hefter, Spray, Hammer oder Schuh) effizient hinzugefügt werden können, ohne zeitintensive Vorhersagemodellaktualisierungen zu benötigen und dennoch ein diskriminierendes Klassifizierungsverhalten zu erzielen. Die Modellaktualisierung durch die SVM beträgt ca. 2 Minuten für das Hinzufügen einer neuen Kategorie. Dabei fluktuiert der Kreuzvalidierungsfehler nur gering und erreicht, bei insgesamt 11 Kategorien, 7,3 %.

Das vorgestellte Verfahren ist ein weiterer Schritt hin zu dem ultimativem Ziel eines autonomen Service Roboter, der unter realen Bedingungen Objekte wahrnimmt, neue Objekte lernt mit Hinblick auf vorherig gelerntes Objektwissen und Schlussfolgerungen aus wahrgenommenen Objekten ziehen kann, um komplexe Aufgabenstellungen zu lösen.

Literatur

- [Bre01] Leo Breiman. Random Forests. *Machine Learning*, 45(1):5–32, 2001.
- [Che95] Yizong Cheng. Mean Shift, Mode Seeking, and Clustering. *IEEE Trans. Pattern Anal. Mach. Intell.*, 17(8):790–799, 1995.
- [FB81] Martin A. Fischler und Robert C. Bolles. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. 24(6):381–395, Juni 1981.
- [Fri94] Bernd Fritzsche. A Growing Neural Gas Network Learns Topologies. In *NIPS*, Seiten 625–632, 1994.
- [HHRB11] Dirk Holz, Stefan Holzer, Radu Bogdan Rusu und Sven Behnke. Real-Time Plane Segmentation using RGB-D Cameras. In *Proceedings of the 15th RoboCup International Symposium*, Istanbul, Turkey, July 2011.
- [HK06] A. Ben Hamza und Hamid Krim. Geodesic Matching of Triangulated Surfaces. *IEEE Transactions on Image Processing*, 15(8):2249–58, August 2006.
- [JT05] Frederic Jurie und Bill Triggs. Creating Efficient Codebooks for Visual Recognition. In *IEEE International Conference on Computer Vision*, Seiten 604–610, Washington, DC, USA, 2005. IEEE Computer Society.
- [LBRF11] K. Lai, L. Bo, X. Ren und D. Fox. A Scalable Tree-based Approach for Joint Object and Pose Recognition. In *Conference on Artificial Intelligence*, August 2011.
- [Mac67] J. B. MacQueen. Some Methods for Classification and Analysis of MultiVariate Observations. In *Proc. of the fifth Berkeley Symposium on Mathematical Statistics and Probability*, Jgg. 1, Seiten 281–297, 1967.
- [MHP11] Christian A. Mueller, Nico Hochgeschwender und Paul G. Ploeger. Surface Reconstruction with Growing Neural Gas. *Active Semantic Perception and Object Search in the Real World Workshop on Intelligent Robots and Systems(IROS)*, 2011.
- [MNJ08] F. Moosmann, E. Nowak und F. Jurie. Randomized Clustering Forests for Image Classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(9):1632–1646, 2008.
- [NHK09] Hideki Nakayama, Tatsuya Harada und Yasuo Kuniyoshi. Dense Sampling Low-Level Statistics of Local Features. Seiten 1–8, 2009.
- [NJT06] Eric Nowak, Frédéric Jurie und Bill Triggs. Sampling Strategies for Bag-of-Features Image Classification. *European Conference on Computer Vision*, Seiten 490–503, 2006.
- [SCZ08] F. Schroff, A. Criminisi und A. Zisserman. Object Class Segmentation using Random Forests. In *Proceedings of the British Machine Vision Conference*, 2008.

Soft String-to-Dependency Hierarchical Machine Translation

Jan-Thorsten Peter, Matthias Huck, and Hermann Ney
Human Language Technology and Pattern Recognition Group
RWTH Aachen University, Aachen, Germany
`{peter,huck,ney}@i6.informatik.rwth-aachen.de`

Daniel Stein
Fraunhofer IAIS
St. Augustin, Germany
`daniel.stein@iais.fraunhofer.de`

Abstract: In this work, we dissect the influence of several target-side dependency-based extensions to hierarchical statistical machine translation, including a dependency language model (LM). We pursue a non-restrictive approach that does not prohibit the production of hypotheses with malformed dependency structures. Since many questions remained open from previous and related work, we offer in-depth analysis of the influence of the language model order, the impact of dependency-based restrictions on the search space, and the information to be gained from dependency tree building during decoding. The application of a non-restrictive approach together with an integrated dependency LM scoring is a novel contribution which yields significant improvements for two large-scale translation tasks for the language pairs Chinese–English and German–French.

1 Introduction

String-to-dependency hierarchical machine translation employs target-side dependency features to capture syntactically motivated relations between words even across longer distances. It is based on the hierarchical phrase-based-translation paradigm [Chi07] and implements enhancements that allow for an integration of knowledge obtained from dependency parses of the training material. Dependency trees over translation hypotheses are built on-the-fly during the decoding process from information gathered in the training phase and stored in the phrase table. A dependency language model can be applied to rate the quality of the constructed tree structures.

In initial publications on the topic [SXW08, SXW10], a restriction of the phrase inventory to a subset of phrases which meet certain validity conditions concerning the dependency relations is proposed. Phrases with dependency structures that are not suitable for the construction of a well-formed dependency tree are excluded beforehand. Additional merging constraints apply during decoding. In later works [SPVN10, H⁺11], heuristics are proposed that enable assembling of malformed dependency structures as well, thus permitting

the utilization of the full phrase inventory of the standard hierarchical approach. Validity and tree well-formedness conditions are modeled in a soft way as features in the log-linear model. Here, the dependency language model is however included in an n -best reranking framework only.

2 Scope

This work aims at filling the gap by investigating string-to-dependency hierarchical translation with and without restrictions, and by comparing dependency LM reranking methods with dependency LM scoring integrated into the decoder. In particular, we explore the following aspects:

- In an n -best reranking framework, only a limited amount of fully generated sentences is presented to the reranking models. We evaluate whether the dependency LM works better in decoding or in reranking.
- The constructed dependency tree is probably erroneous, but so is a parse obtained directly with a dependency parser on a grammatically malformed hypothesis. We analyze whether in a dependency LM reranking framework a direct parsing of the n -best hypotheses performs better than tree building during decoding.
- Restrictions on the phrase table entries as well as on the allowed combination of phrases during decoding might prevent possibly beneficial hypotheses. We investigate whether a soft, i.e. feature-based, approach yields improvements over a restrictive method that guarantees tree well-formedness. We analyze which limitations, if any, are more useful when compared to a non-restrictive approach.
- In the soft approach, the feature set of the log-linear model of the baseline hierarchical system is augmented with additional dependency-based features that can be categorized in two groups: those associated with the tree building process and those related to the dependency LM. We study how dependency tree building features and dependency LM each perform in isolation.
- Usually trigrams are used for the dependency language model. We analyze the typical dependency tree structures found in our data and, based on the findings, explore which dependency language model order is appropriate.

An empirical evaluation is conducted based on the two metrics BLEU [PRWZ02] and TER [SDS⁺06] on a Chinese–English and on a German–French translation task.

3 Methods

String-to-dependency machine translation demands the creation of dependency structures over hypotheses produced by the decoder. Target-side dependency trees are also necessary

Chinese	经长速能缓.
Reference	japanese economic growth may slow down next year .
Baseline	japan 's economy may be slowing in the pace of growth next year .
Dependency	japan 's economic growth may slow down next year .

Table 1: Translation examples, *Reference* denotes a human generated translation

for dependency LM scoring.

To obtain these trees we parsed the training material. The dependency information is carried over to the hypotheses by augmenting the entries in the phrase table with dependency structures. However, the dependency structures seen on phrase level during phrase extraction are not guaranteed to be applicable for the assembling of a dependency tree during decoding. Many of the extracted phrases may be covered by structures where some of the dependencies contradict each other. A standard solution is to restrict the phrase table to only those entries that possess *valid* dependency structures, i.e. structures that comply with certain well-formedness requirements [SXW08]. In an approach without hard restrictions, all kinds of structures are allowed, but invalid ones are penalized [SPVN10]. Merging heuristics allow for building of trees from malformed dependency structures as well.

Given a dependency tree of the target language, we are able to introduce language models that span over longer distances than shallow n -gram language models. The here used dependency language model works on dependency tree levels instead of the plain sentence.

4 Results

We showed that information derived from dependencies can significantly improve the translation performance on both a Chinese–English and a German–French task. Table 1 presents a translation example. By focussing on the individual aspects of the dependency features, we were able to dissect the influences that contribute to this improvement. It seems that parsing the output as compared to constructing dependency trees from structures extracted from parsed training material yields no positive impact on the result. Utilizing a dependency language model during decoding produces better results than when employed in a reranking step. Additionally, in a non-restrictive dependency tree construction process, tree building features give valuable information that can guide the translation to a grammatically more sound direction. The n -gram order of the dependency language model seems to have only a marginal influence. More interestingly, the translation performance is better whenever the decoder search space is not restricted. Even if merging errors are to be expected, it seems to be more important to offer these informations as soft features rather than to exclude certain phrases. A more detailed description of this work and its results can be found in [Pet11] and [PHNS11].

References

- [Chi07] David Chiang. Hierarchical Phrase-Based Translation. *Computational Linguistics*, 33(2):201–228, June 2007.
- [H⁺11] Matthias Huck et al. The RWTH Aachen Machine Translation System for WMT 2011. In *Proc. of the EMNLP/WMT*, pages 405–412, Edinburgh, UK, July 2011.
- [Pet11] Jan-Thorsten Peter. Soft String-to-Dependency Hierarchical Machine Translation. Diploma Thesis, RWTH Aachen University, Aachen, Germany, November 2011.
- [PHNS11] Jan-Thorsten Peter, Matthias Huck, Hermann Ney, and Daniel Stein. Soft String-to-Dependency Hierarchical Machine Translation. In *International Workshop on Spoken Language Translation*, pages 246–253, San Francisco, California, USA, December 2011.
- [PRWZ02] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. Bleu: a Method for Automatic Evaluation of Machine Translation. In *Proc. of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 311–318, Philadelphia, PA, July 2002.
- [SDS⁺06] Matthew Snover, Bonnie Dorr, Richard Schwartz, Linnea Micciulla, and John Makhoul. A Study of Translation Edit Rate with Targeted Human Annotation. In *Proc. of the AMTA*, pages 223–231, Cambridge, MA, August 2006.
- [SPVN10] Daniel Stein, Stephan Peitz, David Vilar, and Hermann Ney. A Cocktail of Deep Syntactic Features for Hierarchical Machine Translation. In *Proc. of the AMTA*, Denver, CO, October/November 2010.
- [SXW08] Libin Shen, Jinxi Xu, and Ralph Weischedel. A New String-to-Dependency Machine Translation Algorithm with a Target Dependency Language Model. In *Proc. of the ACL/HLT*, pages 577–585, Columbus, Ohio, June 2008.
- [SXW10] Libin Shen, Jinxi Xu, and Ralph Weischedel. String-to-Dependency Statistical Machine Translation. *Comput. Linguist.*, 36:649–671, 2010.

Bestimmung von semantisch ähnlichen Worten mit Hilfe von Kookkurrenzen und Wortstämmen

Johannes Schwandke
Technische Universität Darmstadt
jojo-schwandke@web.de

Abstract: Dieses Paper beschreibt eine Methode aus dem Bereich des unsupervised natural language processing (NLP), die zueinander semantisch ähnliche Wörter bestimmt. Das vorgestellte Verfahren arbeitet mit zu Wörtern generierten Wortstämmen und wertet Kookkurrenzdaten aus. Außerdem werden mehrere Ansätze aufgezeigt, welche die Signifikanz der gefundenen Analogien bewerten.

1 Einleitung

Die Frage nach semantischer Ähnlichkeit von Wörtern, unabhängig von der Sprache ist von einer großen Bedeutung für verschiedene Anwendungsbereiche. Bei der Interaktion zwischen Mensch und Computer hilft das Erkennen semantischer Ähnlichkeiten, die Vision vom Sprache-verstehenden Computer realer werden zu lassen. Unüberwachtes Lernen hat den wesentlichen Vorteil, dass keine massenweise “per Hand“ erstellte Daten benötigt werden. In diesem Paper ist ein Ansatz beschrieben mit der Idee, dass Wörter eine analoge Bedeutung haben, wenn sie einen gemeinsamen Wortstamm haben und zusätzlich in einem vergleichbaren Kontext auftreten. Es wird eine Methode vorgestellt, relevante Kandidaten zu ermitteln. Im Weiteren wird die Auswertung der Kandidaten durch verschiedene Kriterien diskutiert.

2 Vorgehensweise

Zur Wortstammgewinnung dient das morfessor-Tool.[CRL2] Der Morfessor Baseline Algorithmus ist ein rekursives Minimum Description Length (MDL) Verfahren. Erweiterungen [CR], [CRL1] dieses Verfahrens benutzen ein Maximum Likelihood Modell und ein hierarchisches Lexikon.

Für diese Arbeit wurde ein Textcorpus [LCC] mit einer Million Sätzen ausgesucht. Zu dem gewählten Corpus werden zunächst für alle vorkommenden Worte die Wortstämme berechnet. Anschließend werden Kookkurrenzdaten verwendet, um Kandidaten für semantische Ähnlichkeit zu gewinnen. Diese Kandidaten werden dann in einem letzten Schritt mit verschiedenen Methoden ausgewertet.

Ein Kandidat beschreibt eine potentielle Ähnlichkeit zweier Wörter. Es werden nur Worte

in Betracht gezogen, die mindestens einen gemeinsamen Wortstamm und eine gemeinsame Kookkurrenz haben. Die Qualität eines Kandidaten, also die Relevanz für semantische Äquivalenz des Wortpaares, hängt von mehreren Faktoren ab.

Es gibt vier Faktoren, die den Kandidaten direkt betreffen: jeweils die Frequenz beider Kandidatenworte im Textcorpus, die Anzahl der gemeinsamen Kookkurrenzen¹ und die Anzahl der gemeinsam genutzten Stämme. Weiterhin wird zu jedem Wortstamm abgelegt, in wie vielen verschiedenen Worten er vorkommt. Zu jeder gemeinsamen Kookkurrenz werden jeweils drei Werte gespeichert, die Frequenz im Textcorpus und jeweils ein Signifikanzwert zu beiden Worten des Kandidaten. (log-likelihood Signifikanzmaß.)

Die Berechnung der Kandidaten geschieht wie folgt: Zuerst wird der Suchraum für die Kookkurrenzen festgelegt, also ob man unter benachbarten Wörtern sucht (con), oder im ganzen Satz (cos). Dann wird zu jedem Wortstamm ermittelt, in welchen Worten dieser vorkommt. Der Algorithmus nutzt die von morfessor ermittelten Wortstämme, um Kandidaten zu finden. Zu jedem Stamm gibt es eine Wortmenge W aller Wörter, die diesen enthalten. Da die Wortpaare aller Kandidaten jeweils mindestens einen gemeinsamen Stamm haben, reicht es für jeden Stamm aus, innerhalb dieser Menge W alle möglichen Kombinationen abzarbeiten. Für die gebildeten Kombinationen muss anschließend noch überprüft werden, ob es gemeinsame Kookkurrenzen gibt. (Schnittmengenbildung)

2.1 Baseline

Als Baseline-Methode wurde berechnet, welcher Buchstabenanteil der beiden Worte eines Kandidaten durch dessen Stämme abgedeckt wird. Zwei Wörter gelten somit als ähnlich, wenn sie in ihrer Länge vom gemeinsamen Stamm nur wenig abweichen. Ein Blick auf zwei der Testworte macht deutlich, dass diese Methode stark von der Qualität der Stämme abhängig ist. Das Wort "verzicht" z.B. hat die ersten drei Ähnlichkeiten "verzichte", "verzichtet" und "verzichten", wozu hingegen das Wort "darstellungen" mit "stellung", "umstellung" und "zustellung" in Verbindung gebracht wird. Generell ist es beim Baselineverfahren sehr auffällig, dass zu einem Testwort meist entweder alle als ähnlichen gefundenen Worte passen oder gar keins. Das ist darauf zurückführbar, dass der gemeinsame Kontext der Worte gänzlich außer Acht gelassen wird und nur die Schreibweise für die Bestimmung von Ähnlichkeit genutzt wird.

2.2 Weitere Methoden

Die "Inverse-Frequenz-Methode" (if) iteriert über alle gemeinsamen Kookkurrenzen eines Kandidaten. Versucht man nun die Relevanz zu ermitteln, so ist die Bindung zu diesem Wort umso stärker, je seltener es auftritt. Seltenerere Worte werden stärker gewichtet als häufigere. Die Summe über die einzelnen Kookkurrenzen bildet das Endergebnis für den Kandidaten. Die Wortstämme gehen hier nicht in die Berechnung ein. Es wird angenom-

¹eine gemeinsame Kookkurrenz ist ebenfalls ein Wort aus dem Corpus

men, dass die gemeinsame Verwendung der Stämme den Aspekt ausreichend berücksichtigt. Eine weitere Auswertungsvariante nutzt die Signifikanzwerte der gemeinsamen Kookkurrenzen. Die Kookkurrenz zweier Worte kann durch diesen Wert klassifiziert werden. Da die Kookkurrenzen eines Kandidaten zweistufig sind, werden zu jedem Wort der Kookkurrenzliste eines solchen auch zwei Signifikanzwerte abgelegt. Dabei ist zu beachten, dass eine gemeinsame Kookkurrenz nur dann aussagekräftig ist, wenn sie zu beiden Worten des Kandidaten eine starke Bindung hat. Die verwendete “*Signifikanz-gewichtete Methode*“ (Msw) berechnet zu einem Kandidaten dessen besten Wert für eine gemeinsame Kookkurrenz.

2.3 Auswertung der Methoden

Die vorgestellten Methoden wurden jeweils auf 100 Testwörter angewendet. Dabei wurden die Testwörter für die Suchräume *cos* und *con* unterschiedlich gewählt. Für jedes der Verfahren wurde eine Liste von Testwörtern produziert, sowie die zu jedem Testwort gehörigen n besten Ähnlichkeiten in absteigender Ordnung. Dann wurde der Anteil korrekt erkannter Analogien ermittelt. Folgende Worte wurden als ähnlich anerkannt: Synonyme, Worte, die sich nur aufgrund ihres Kasus, Numerus oder Genus unterscheiden und korrespondierende Nomen und Verben (z.B. “*fliegen*” und “*Flug*”). Die nachfolgenden Tabellen zeigen die Ergebnisse für n = 5 und n = 2. Es wurden zusätzlich Kandidaten herausgefiltert, deren Wert unter dem arithmetischen Mittel lag.

n = 5 (mit Filter)				n = 2 (mit Filter)			
Suchraum	Methode			Suchraum	Methode		
	Baseline	if	msw		Baseline	if	msw
cos	~70%	~65%	~87%	cos	~70%	~69%	~88%
con	~72%	~60%	~75%	con	~78%	~65%	~84%

3 Zusammenfassung

Die Auswertung der Testwörter hat ergeben, dass die vorgestellten Methoden einige sehr gute semantische Analogien finden. Die Ergebniswerte sind eine gute Basis für weitergehende Ansätze. Denkbare Ansätze zur Verbesserung liegen in der Kombination der Auswertungsmethoden oder der Erfindung komplexerer Varianten zur Bestimmung des Ähnlichkeitsniveaus eines Kandidaten. Außerdem zeigt die Arbeit, dass die Idee, über Wortstämme und Kookkurrenzdaten semantische Ähnlichkeiten zu berechnen, ein guter Ansatz ist. Die Möglichkeiten verschiedener Bewertungsmethoden für Kandidaten ist, wie erläutert, nicht vollständig erschöpft, daher sind bessere Ergebnisse durch zusätzliche oder kombinierte Verfahren zu erwarten. Durch aufwändigere Auswertungsstrategien und/ oder einen umfangreicheren Corpus kann man vermutlich das Niveau der Ergebnisse weiter verbessern.

Literatur

- [CR] M. Creutz. 2003. *Unsupervised segmentation of words using prior distributions of morph length and frequency*. In Proc. ACL'03, Sapporo, Japan.
- [CRL1] M. Creutz, K. Lagus. 2002. *Unsupervised discovery of morphemes*. In Proc. Workshop on Morphological and Phonological Learning of ACL'02 Philadelphia, Pennsylvania, USA.
- [CRL2] M. Creutz, K. Lagus. 2005. *Inducing the morphological lexicon of a natural language from unannotated text*.
- [LCC] C. Biemann, G. Heyer, U. Quasthoff, M. Richter. 2007. *The Leipzig Corpora Collection: Monolingual Corpora of Standard Size*.

SOLVE – Geschlechterspezifisches Problemlösungsverhalten

B.Eng. Anke Tallig

Hochschule Merseburg
FB Informatik und Kommunikationssysteme (IKS)
anke.tallig@stud.hs-merseburg.de

Art der Arbeit: Projekt

Teilnehmer: B.Eng. Nicki Granz, B.Eng. Sebastian Renner, B.Eng. Jens Schreck

Betreuer: Prof. Dr.-Ing. Karsten Hartmann (karsten.hartmann@hs-merseburg.de)

Das Projekt SOLVE beschäftigt sich mit geschlechterspezifischem Problemlösen, um die Problemlösungsstrategien von Frauen und Männern zu erforschen. Die Ergebnisse aus den Usability-Tests sollen durch weitere Tests, unter verschiedenen Aspekten, erweitert und bestätigt werden. Ziel der Arbeit ist es, mit den Erkenntnissen der Versuche intelligente Assistenzsysteme so zu erweitern, dass sie die Nutzergemeinde bei der Lösung von Aufgaben, die mit Schwierigkeiten verbunden sind, unterstützen.

Das Problemlösungsverhalten von Menschen ist vielfältig. Genauso vielfältig sind aber auch die Bereiche für Soft- und Hardware, in denen aufgearbeitete Ergebnisse des Problemlösungsverhaltens von Personen eingearbeitet und eingesetzt werden können.

Problemlösungsverhalten? DAS Problemlösungsverhalten gibt es nicht!

Das Problemlösungsverhalten von Menschen lässt viele Fragen aufkommen, wie zum Beispiel: Lösen Menschen Aufgaben, die mit Schwierigkeiten verbunden sind, gleich? Wenn nicht, wie unterscheidet sich das Lösen? Wird das Lösungsverhalten von der Vorbildung, vom Vorwissen beeinflusst? Vielleicht vom Alter? Ist das Geschlecht der einzelnen Person entscheidend oder einfach nur der Blickwinkel aus dem sie das Problem betrachten? Hat der Problemlösungsvorgang weitere beeinflussende Faktoren? Welche sind das? Wie beeinflussen sie die problemlösenden Personen? Wie beeinflussen sich Personen, die in einer Gruppe Probleme lösen sollen, gegenseitig, durch ihr unterschiedliches Problemlösungsvorgehen?

Kann uns das Wissen um die Problemlösungsvorgänge beeinflussenden Faktoren helfen, intelligente Assistenzsysteme zu modifizieren und zu erweitern, damit die Schwierigkeiten der User bei der Lösung ihrer Probleme auf ein Minimum reduziert werden, im Idealfall überhaupt nicht mehr auftreten?

Ja, es kann helfen!

Doch bevor den Nutzern von technischen Systemen geholfen werden kann, muss sich im Rahmen der Künstlichen Intelligenz mit dem Problemlösungsverhalten von Menschen beschäftigt werden.

Hierbei wird die Interaktion Mensch-Problem bzw. Mensch-Mensch-Problem bei der Problemlösungssuche in Gruppen, nach bestimmten Aspekten beobachtet und ausgewertet. Die Ergebnisse dieser Untersuchungen werden, nach der empirischen Aufarbeitung, für die Modifizierung von intelligenten Assistenz- bzw. Hilfesystemen, in puncto Hilfe bei Aufgaben, deren Lösung mit Schwierigkeiten verbunden sind, verwendet. Ein weiterer Schritt könnte sein, nicht nur die Interaktion Mensch-Maschine zu verbessern, sondern auch die Maschine-Maschine-Interaktion für diesen Themenbereich aufzubauen und zu erweitern. Damit könnte die Weitergabe von aufgenommenen und analysierten Problemlösungsverhalten zwischen verschiedenen Systemen realisiert werden. Technische Systeme könnten so verschiedene Problemlösungswege erlernen, diese aufbereiten und an Systeme, die sie für die Mensch-Maschine-Interaktion benötigen, weitergeben. Dies sind jedoch erst Ausblicke in die Zukunft. Das Fundament für diese Ideen wird durch Beobachtungen von Problemlösungsvorgängen, Usability-Tests, in der Gegenwart gelegt.

Was ist der Untersuchungsgegenstand vom Projekt SOLVE?

Innerhalb des Projekts wurde das typische Problemlösungsverhalten von Einzelpersonen und Gruppen getestet. Dabei wurde untersucht, ob sich die beiden Geschlechter bei Vorlage einer gleichen Problemstellung unterschiedlich verhalten und wie sich die Zusammensetzung der Gruppe auf den Problemlösungsablauf auswirkt. Darüber hinaus wurden Untersuchungen zur Gruppendynamik vorgenommen.

Wie wurden die Versuche durchgeführt?

Für die Durchführung der Feldversuche, Introspektionen und Interviews wurden die Probanden_innen in mehrere Gruppen eingeteilt. Zum einen wurden jeweils eine rein weibliche Gruppe (4 Personen) und eine rein männliche Gruppe (4 Personen) gebildet. Zum anderen wurden die Geschlechter in einer fairen (2 Frauen, 2 Männer) und einer unfairen (1 Frau, 3 Männer) Gruppe gemischt. Drei Personen (1 weibliche, 2 männliche) wurden in Einzelsitzungen mit Hilfe eines „Lautes-Denken-Interviews“ (aus der Introspektion abgeleitet) anhand der zweiten Aufgabe getestet.

Den zusammengestellten Gruppen wurden drei verschiedene Aufgaben gestellt:

- I. Gründung einer Firma
 1. Gründung einer Kfz-Werkstatt für die rein weibliche Gruppe
 2. Gründung eines Friseursalons für die rein männliche Gruppe
 3. Gründung einer Kneipe, Aufgabe der fair gemischten Gruppe
 4. Gründung einer Kfz-Werkstatt (< Frauen) oder eines Friseursalons (< Männer)

- II. Terra incognita

Ein Spiel, in dem der Weg von einem Ausgangspunkt zu einer entfernten Stadt gefunden werden musste.

III. Die Umsetzung des Gender-Mainstreams in der oben gegründeten Firma (siehe Aufgabe I).



Abbildung 1: SOLVE-Team beim Test „Terra Incognita“
(Gestellt, zum Schutz der Versuchspersonen)

Der gesamte Vorgang der Problemlösung der Gruppen wurde auf Video aufgezeichnet. Während des Versuches wurden die Personen über einen einseitigen Spiegel beobachtet und ihr Verhalten protokolliert. Der betreuende Professor oder einer der am Projekt teilnehmenden Studierenden befand sich im gleichen Raum wie die Versuchspersonen, um bei Bedarf Hilfestellungen zu geben. Den Probanden_innen waren zur Anonymisierung Nummern zugeordnet.

Welche Methoden wurden für die Auswertung eingesetzt?

Die Protokolle und Videoaufnahmen wurden von den beteiligten Studierenden zur Erstellung des Problemlösungsweges und des Aufbaus der Gruppenstruktur ausgewertet. Dazu wurden im Voraus Hypothesen aufgestellt, die zum einen durch die Untersuchungen bestätigt werden sollten und zum anderen als Prämissen für empirische Aussagen dienen. Für die Analyse wurden die Methoden Introspektion, Interview und Feldversuch eingesetzt. Diese Methodenvielfalt sollte für unterschiedliche Blickwinkel in der Auswertung sorgen. Die Analyse erfolgte unter zwei Gesichtspunkten. Erstens wurde die Situation analysiert. Dabei erfolgte die Zerlegung der Situation in ihre Bestandteile und eine detaillierte Darstellung der Subjekt-Objekt-Beziehung. Zweitens wurde das Ergebnis analysiert. Nachdem ein Vernetzungsplan aller Einzelelemente aufgestellt wurde, konnten die Einzelergebnisse in das Gesamtbild eingegliedert werden. Dabei wurde auch die Häufigkeit, wie oft eine bestimmte Probandin bzw. ein bestimmter Proband zu Wort kam, aufgenommen. Ob sie bzw. er die Problemlösung maßgeblich beeinflusst hat und auf welche Art und Weise, wurde genauso statistisch ausgewertet, wie Redeversuche, unterbrochene verbale Einwürfe, welche Einblicke auf die Gruppenkonstellation und die Person an sich zulassen.

Im letzten Schritt für eine nutzbringende Anwendung wurden die erkannten Regelmäßigkeiten in den Problemlösungsstrategien und dem Verhalten zusammengefasst und Kategorien gebildet (induktive Verallgemeinerung).

Zu welchen Ergebnissen führten die Untersuchungen?

An dieser Stelle sollen drei markante Ergebnisse, zwei Ergebnisse für das Problemlösungsverhalten und ein Ergebnis für das Gruppenverhalten, aufgeführt werden. Das Problemlösungsverhalten von Frauen und Männern unterscheidet sich wesentlich im Vorgang des Informationensammelns. Dabei arbeiten Frauen strukturierter als Männer. Sie teilen die gewonnenen Informationen direkt in Themengebiete ein (ähnlich Mind Mapping), während Männer im Gegensatz dazu die Daten erst einmal, wie beim Brainstorming, sammeln und in einem zweiten Schritt sortieren. Des Weiteren wurde festgestellt, dass Frauen wie Männer in einem Spielszenario den rechten Weg bevorzugen. Die Beobachtung deckt sich mit der Wegeanordnung in Supermärkten, die auch in den meisten Fällen rechts beginnen und die Einkäufer zu Linkskurven veranlassen. Genauso zeigt sich das Phänomen bei Stadionrunden im Sport. Die Analyse des Verhaltens in der Gruppe zeigte Unterschiede in der „Vergabe“ der Führungsrolle. Bei Männer-Gruppen hat von Anfang an eine Person die Führungsrolle inne und behauptet diese unabhängig von der Kompetenz. Die Führungsrolle bei Frauen-Gruppen hingegen wird von der Kompetenz der Person für ein Thema bestimmt. Zeigt eine Frau während des Lösungsprozesses mehr Kompetenz als eine andere, kann die Führungsrolle wechseln, da so die Erfolgchancen höher sind.

Alle durch die Untersuchungen erhaltenen Ergebnisse werden in der Fortführung des Projektes als Grundlage genommen. Es sollen Szenarien programmiert werden, die die Beobachtungen bedienen und somit weitere statistische Auswertungen ermöglichen. Im Weiteren werden aber auch neue Tests entworfen, diese sollen die nicht eindeutigen Beobachtungen in den Fokus rücken und zu einem verwertbaren Ergebnis führen.

Ziel ist es, wie oben schon erklärt, Daten zu sammeln, die zur Erweiterung von intelligenten Assistenzsystemen, in Bezug auf die Hilfe bei Problemlösungsvorgängen, dienen.

Literaturverzeichnis

- [Hussy93] Hussy, Walter: Denken und Problemlösen. Kohlhammer, Stuttgart, 1993.
- [Vogt98] Vogt, Hubert M.: Persönlichkeitsmerkmale und komplexes Problemlösen. Rainer Hampp Verlag, München, 1998.
- [Wessells94] Wessells, Michael G.: Kognitive Psychologie. Reinhardt, München, 1994.
- [Winkelholz06] Winkelholz, C.: Analyse, Modellierung und Simulation der visuell-räumlichen Kognition bei der Mensch-Maschine-Interaktion. Shaker, Aachen, 2006.

Simbac - Ein Analysetool zur Simulation von Bachelorabschlussquoten

Alexander Brömmer, Marius Fink und Till Schomborg
Studienprojekt des Instituts für Logistik und Transport (ILT), Universität Hamburg

a.broemmer@googlemail.com
simbac@mariusfink.org
8schombo@informatik.uni-hamburg.de
Projekthomepage: <http://simbac.mariusfink.org>

Abstract: Die Abschlussquoten der neuen Bachelorstudiengänge sind schwer abzuschätzen. Dennoch werden aussagekräftige Daten benötigt, um Studiengänge studierbarer zu machen. Diese Arbeit umfasst einen Prototypen zur generischen, agentenbasierten Simulation von Studierenden. Dieser ist durch ein umfassendes Konzept geprägt, das mathematische Modelle und ereignisgesteuerte Simulationen beinhaltet. Das Ergebnis ist ein Werkzeug zur Überprüfung der Validität von tatsächlichen Abschlussquoten im Verhältnis zu den Eingabeparametern.

1 Motivation

Im Rahmen des Bologna-Prozesses kam es in den meisten Ländern zur Einführung von Bachelor- und Masterstudiengängen mit dem Ziel eines einheitlichen europäischen Hochschulraums mit vergleichbaren Abschlüssen (vgl. [Wal06]). Die Einführung der Bachelor- und Masterregelung hat neue Studiengänge hervorgebracht und klassische Studiengänge wurden an diese neuen Vorgaben angepasst. Folglich bekamen die Studiengänge völlig neue Strukturen, wie z.B. die Festlegung des Studienverlaufs, die Einführung von Pflicht- und Wahlpflichtmodulen, die Einführung des European Credit Transfer Systems (ECTS) oder einer Regelstudienzeit. Die Erfolgsquoten der einzelnen Studiengänge können bislang nur schlecht vorausgesagt werden. Das Ziel dieses Projektes ist die Erarbeitung eines Konzepts zur Abschätzung der Erfolgsquoten von Bachelorstudiengängen. Auf Basis des erarbeiteten Konzepts wurde ein Prototyp eines Simulationswerkzeugs entwickelt. Die Simulation soll in Abhängigkeit von den Einflussgrößen Erfolgsquoten für einen abgebildeten Studiengang ermitteln. In der Simulation wird der Studierende¹ als Softwareagent modelliert. Dieses Modell erfüllt nahezu alle Bedingungen der *weak notion of agency* (vgl. [WJ95]) indem es über die Eigenschaften *Autonomie*, *Proaktivität* und *Reaktivität* verfügt.

¹Wir sehen keine relevanten Gründe, Geschlechter abzubilden. Wann immer von „der Studierende“ gesprochen wird, ist die geschlechtsunabhängige *Entität* „Studierender“ gemeint.

2 Konzept der Simulation

Eingabe Die Eingabe umfasst eine Menge von Parametern und Daten, die einerseits den nötigen Input für die Simulation darstellen und andererseits auch die Einflussstärke der Parameter in der Verarbeitung bestimmen. Folgende Eingabeparameter sind möglich: *Motivationsfaktor* (als Verteilung, beeinflusst die Anzahl der Module pro Semester, Prüfungen pro Termin sowie die Prüfungsnoten), *Vorbildungsanteile* (Abitur, Fachabitur, Realschulabschluss), *Einflussfaktoren* (z.B. Jobs, werden mit dem Motivationsfaktor verrechnet), *Anzahl der zu simulierenden Studierenden*, die *Regel- und maximale Studienzeit* sowie folgende *Korrelationskoeffizienten*: Motivationseinfluss auf Prüfungsnoten, Prüfungseinfluss durch korrelierende oder vorausgesetzte Prüfungen, Bildungseinfluss auf Motivation sowie der Anstieg der Motivation bei nicht bestandener Prüfung. Außerdem kann die Prüfungsordnung abgebildet werden. Dazu gibt es Prüfungen in Modulen, die wiederum in Bereiche gegliedert sind. Zu jeder Prüfung können die *Notenverteilung* (und damit die Durchfallquote), die *Prüfungstermine* (Winter- bzw. Sommersemester), *maximale Versuche* und *Korrelationen mit anderen Prüfungen* definiert werden. Neben den Prüfungen enthält ein Modul eine Anzahl Credits, ein Referenzsemester, zwingende oder empfohlene Voraussetzungen sowie „Credits in Bereich“-Voraussetzungen. Der Bereich enthält die Module und die Anzahl der in diesem Bereich zu absolvierenden Credits. Diese Eingabeparameter orientieren sich an der Prüfungsordnung für den Bachelorstudiengang *Wirtschaftsinformatik* der Universität Hamburg.

Verarbeitung Am Anfang der Verarbeitung steht die Generierung eines Studierenden, bei der er zufällige Zustandsvariablen erhält. Dazu gehört die Vorbildung sowie eventuelle weitere Einflüsse und der Motivationsfaktor. Danach muss bestimmt werden, welche Module der Studierende im laufenden Semester belegt. Wir gehen davon aus, dass der Studierende sich nach rationalen Kriterien, wie dem Referenzsemester oder den Modulvoraussetzungen entscheidet (siehe Skala). Wie viele Module er belegt, ist dabei von einer

hohe Priorität ↑ ↓ niedrige Priorität	1) Das Referenzsemester liegt im letzten oder vorletzten Fachsemester.
	2) Das Modul ist eine zwingende Voraussetzung für Module mit Referenzsemestern in der Vergangenheit oder im aktuellen Semester.
	3) Das Referenzsemester des Moduls liegt in diesem Fachsemester.
	4) Das Modul ist eine zwingende Voraussetzung für Module mit Referenzsemester in den nächsten zwei Semestern.
	5) Es ist eine zwingende Voraussetzung für Module, die nicht von 2 oder 4 betroffen sind.
	6) Das Referenzsemester liegt im nächsten oder in einem späteren Fachsemester.
	7) Das Modul ist eine empfohlene Voraussetzung für ein noch nicht belegtes Modul.
	8) Der Student hatte gute Noten in den empfohlenen Voraussetzungen des Moduls.
	9) Der Student hatte gute Noten in korrelierenden Prüfungen.
	10) Das Modul hat nur einen Prüfungstermin im aktuellen Semester.
	11) Das Modul hat nur zwei Prüfungstermine im aktuellen Semester.
	12) Das Modul hat mehr als eine Prüfung.

Tabelle 1: Entscheidungsskala eines Studierenden zur Priorisierung der Module

Kombination aus Motivation und dem Zufall abhängig. Hier soll ein möglichst intelligentes Verhalten des Studierenden simuliert werden, bei dem jeder Studierende individuell die für ihn besten Module wählt. Falls für ein Modul mehrere Prüfungstermine möglich sind, wird zufällig ein Termin gewählt, sofern dadurch nicht zu viele Prüfungen in einer Prüfungsphase belegt werden. Der nächste Schritt ist die Bestimmung der Prüfungsnoten. Dabei wird zunächst ein für die Prüfung angepasster Motivationsfaktor ermittelt. Dieser kann vom Motivationsfaktor des Studierenden abweichen, wenn der Studierende die Prüfung bereits einmal nicht bestanden hat. Anschließend wird mit Hilfe der Korrelation zwischen Motivation und Prüfungen eine motivationsbeeinflusste Zufallszahl generiert. Dann werden weitere Einflussfaktoren auf diese Prüfung ermittelt. Dazu zählen neben den Noten der korrelierenden Prüfungen auch alle Prüfungsnoten in vorausgesetzten Modulen. Aus der motivationsbeeinflussten Zufallszahl und den weiteren Einflussfaktoren wird dann ein Faktor ermittelt, aus dem, mittels inverser Transformation auf die Notenverteilung der Prüfung, die endgültige Prüfungsnote ermittelt wird. Eine Prüfung gilt als bestanden, wenn sie mit einer anderen Note als 5,0 abgelegt wurde. Eine nicht bestandene Prüfung kann wiederholt werden, sofern dies die Prüfungsordnung gestattet. Nachdem alle Prüfungsnoten eines Semesters ermittelt wurden, wird überprüft, ob der Studierende sein Studium fortsetzen wird. Dazu dürfen keine Modulfristen oder maximale Versuche überschritten worden sein. Außerdem kann der Studierende sich freiwillig für einen Abbruch des Studiums entscheiden. An dieser Stelle wird auch geprüft, ob der Studierende sein Studium bereits erfolgreich abgeschlossen hat.

Ausgabe Die Ausgabe erfolgt in einem eigenen Programm, dem sog. *Viewer*. Er kann neben dem Export und der Ansicht von Eingabedaten der Simulation alle Aktionen der Studierenden in einem Baum anzeigen. Dieses Feature ist aus dem Wunsch nach Nachvollziehbarkeit und Debugging entstanden. Des Weiteren verfügt der Viewer über folgende (grafische) Auswertungen: Einen *Notenspiegel* (zeigt die Abschlussnoten der simulierten

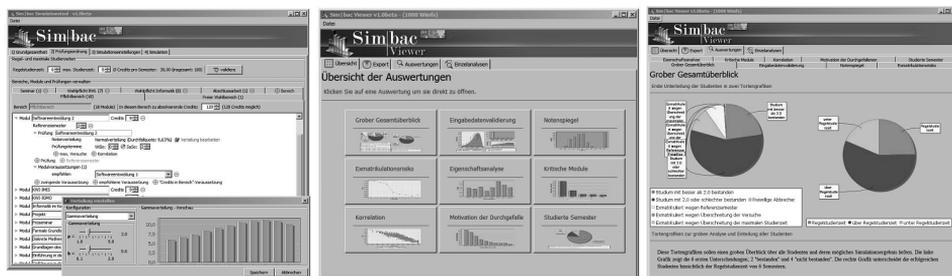


Abbildung 1: Bildschirmfotos der Eingabe und Auswertungen

Studierenden an, wobei die genauen Durchschnittsnoten auf das Notenintervall gerundet werden), eine *Eigenschaftsanalyse der Durchgefallenen* (betrachtet, wie hoch der Anteil der durchgefallenen Studenten mit gewissen Eigenschaften ist), eine *Eingabedatenvalidierung* (durch Soll-Ist-Vergleiche, Standardabweichungsberechnungen der Grundmotivation, der tatsächlichen Bildungsverteilung sowie der Prüfungsnotenverteilung jeder Prüfung

mit den tatsächlichen Noten), eine „*Kritische Module*“-Analyse (also die Identifizierung solcher Module, die Auslöser für die meisten Exmatrikulationen sind. Sie gibt der/dem Benutzer/in Aufschluss darüber, welche Module für eine Lockerung der Fristen oder Versuche in Frage kommen, um eine höhere Abschlussquote zu erzielen), eine *Exmatrikulationsrisiko-Analyse* (stellt die Risiken, ein Studium nicht abzuschließen in Abhängigkeit der studierten Semester dar), eine *Korrelationsanalyse zwischen Motivation und Note* sowie eine Auswertung „*Studierte Semester*“ (zeigt eine Übersicht über die Menge aller insgesamt absolvierten Studiensemester und eine Kostenanalyse durch das Überschreiten der Regelstudienzeit).

3 Kritik und Ausblick

Sowohl der Umfang der in einem Semester belegten Module als auch die Prüfungsnoten hängen stark von der Motivation des Studierenden ab. Wir gehen zur Zeit noch von der unrealistischen Annahme aus, dass ein guter Studierender immer sowohl viele Module belegt, als auch gute Noten erreicht. Es ist daher beispielsweise nicht möglich, einen Studenten abzubilden, der zwar nur wenige Module absolviert, diese dafür aber mit sehr guten Noten besteht. Grundsätzlich fehlen im aktuellen Stand noch valide Eingabedaten. Motivation ist als einzelner Faktor ein Anfang, aber noch weit davon entfernt, die Entscheidungen von Studierenden real abzubilden. Dabei gibt es Daten genug: Das Projekt *ZEITLast* (vgl. [SM11]) des Zentrum für Hochschul- und Weiterbildung (ZHW) befasst sich mit der Lern- und Studienzeit von Studenten, die sie ggf. mit den Prüfungsnoten vergleichen können. Eine erste Analyse hat jedoch keinen signifikanten Zusammenhang aufzeigen können. Eine andere Datenquelle sind Portale zur Bewertung von Studiengängen, Kursen und DozentInnen. Auf meinProf.de werden z.B. Faktoren wie „Fairness in Prüfungen“ oder „Zeitaufwand“ für Kurse erhoben. Ein Angleichen des Konzeptes an diese und ähnliche Parameter wäre ein möglicher Meilenstein des Simbac-Projektes. Eine andere Möglichkeit ist der Einsatz von zwei unabhängigen Faktoren: *Zeitinvestitionswille* (Quantität) und *Leistungsmotivation* (Qualität). Dabei könnte der *Zeitinvestitionswille* über die Anzahl der Kurse und Prüfungen pro Semester und die *Leistungsmotivation* über Prüfungsnoten entscheiden.

Literatur

- [SM11] Rolf Schulmeister und Christiane Metzger. *Die Workload im Bachelor: Zeitbudget und Studierverhalten. Eine empirische Studie*. Waxmann, Münster, Westfalen, 2011.
- [Wal06] Thomas Walter. *Der Bologna-Prozess: ein Wendepunkt europäischer Hochschulpolitik*. VS Verlag für Sozialwissenschaften, Wiesbaden, 2006.
- [WJ95] Michael Wooldridge und Nicholas Jennings. Agent theories, architectures, and languages: A survey. In Michael Wooldridge und Nicholas Jennings, Hrsg., *Intelligent Agents*, Jgg. 890 of *Lecture Notes in Computer Science*. Springer, Berlin/Heidelberg, 1995.

Entwicklung eines Konfigurationsvalidierers

Christian Cikryt

Freie Universität Berlin
Arbeitsgruppe Softwaretechnik
in Zusammenarbeit mit tarent solutions GmbH
c.cikryt@fu-berlin.de

Art der Arbeit: Bachelorarbeit in Informatik

Betreuer der Arbeit: Prof. Dr. Lutz Prechelt, Prof. Dr. Elfriede Fehr, Florian Wallner

Abstract: Das Ziel dieser Arbeit ist es, ungültige Konfigurationen von Softwareartefakten zurückzuweisen, um eine aufwendige Fehlersuche während des Betriebs zu vermeiden. Dazu wird – nach Evaluation existierender Technologien – eine Beschreibungssprache (DSL) zum Erfassen von Konfigurationsanforderungen sowie ein Validierungsprogramm entwickelt. Die erstellten Beschreibungen werden mittels eines Parsergenerators eingelesen und auf Erfüllbarkeit und Widersprüche geprüft. Anschließend können konkrete Konfigurationen anhand der formulierten Kriterien validiert werden.

1 Motivation

Bei der Konfiguration von komplexen, aus vielen Komponenten bestehenden Systemen stellen fehlerhafte Einstellungen keine Seltenheit dar, ziehen aber mitunter eine aufwendige Fehlersuche nach sich, weil sie in der Regel erst zur Laufzeit bemerkt werden. Es reicht ein Tippfehler in einer Server-URL, ein „hpt“ statt „http“, um eine Anfrage mit einer Fehlermeldung zu beantworten. Kostspielig wird dies, weil für die Qualitätssicherung nicht erkennbar ist, ob es sich um einen Bug oder eine Fehlkonfiguration handelt. Anders als bei Mängeln im Code existieren noch keine automatischen Tests für Konfigurationen, die unmittelbar Rückmeldung liefern. Die in dieser Arbeit vorgestellte Lösung setzt genau hier an: Konfigurationen sollen noch vor dem Einspielen automatisch überprüft und die dafür benötigten Informationen mit der Software ausgeliefert werden. Das bedeutet, dass für die einzelnen Komponenten die Anforderungen in einer zu entwickelnden Beschreibungssprache niedergeschrieben werden müssen. Im Folgenden wird der Entwurfsprozess dieser Sprache beleuchtet und dabei Fokus auf die Technologieauswahl, speziell auf die Abwägung zwischen externer und interner DSL gelegt.

2 Entwicklung einer Beschreibungssprache

Bei dem Entwurf der Beschreibungssprache dienten die häufigsten Konfigurationsfehler als Hauptgrundlage, die durch eine Befragung der beteiligten Administratoren und Entwickler¹ gesammelt worden sind. Damit konnten die Anforderungen an die Ausdrucksstärke der Sprache ermittelt und in einem Domänenmodell (siehe [Eva03]) festgehalten werden, so dass auf dieser Grundlage die Sprachsyntax skizziert und damit Technologien auf ihre Eignung untersucht werden konnten. Das entstandene Domänenmodell stellt eine solide Schnittstelle zwischen Sprache und Validierungslogik dar und bleibt von Änderungen an der Beschreibungssprache unbeeinträchtigt. Gleichzeitig bildet es das Fundament für die Evaluierung, weil sich eine konkrete Sprache daran testen lässt, ob sie alle Objekte des Modells mit Daten befüllen kann.

2.1 Evaluation der Sprachtechnologie

Der Evaluation einer Technologie für die Beschreibungssprache wurden die folgenden Maßstäbe zugrunde gelegt: Weil die Sprache einfach erlernbar sein und eine präzise und kompakte Beschreibung erlauben soll, sind eine verständliche und konsequente Syntax sowie eine intuitive Semantik Pflicht. Auch soll sie für zukünftige Anforderungen erweiterbar sein. Eine gute IDE-Unterstützung mit Syntaxvervollständigung und -hervorhebung ist wünschenswert, aber bei der angestrebten Einfachheit der Sprache nicht zwingend nötig.

Man bezeichnet eine derartige Sprache aufgrund ihrer Beschränktheit auf eine Anwendungsdomäne auch als Domänenspezifische Sprache (Domain-specific Language / DSL). Im Gegensatz zu generischen Programmiersprachen beschränkt sich eine DSL auf die für den Anwendungsfall minimal benötigten Funktionen, z. B. verzichten DSLs für gewöhnlich auf Kontrollstrukturen und Variablen. Für einen Überblick über die Publikationen in diesem Bereich sei auf [vDKV00] und [FP11] verwiesen.

Es wird zwischen internen und externen DSLs unterschieden. Interne DSLs sind in der Hostsprache geschrieben, während externe DSLs in einer separaten Sprache verfasst sind und geparkt werden müssen. Interne DSLs haben den Vorteil, dass keine zusätzliche Einarbeitung in externe Technologien erforderlich ist und man die gewohnte IDE-Unterstützung in Form von Syntaxvervollständigung und -hervorhebung während des Verfassens erhält. Dagegen ist man in der Syntaxwahl je nach verwendeter Sprache mitunter sehr eingeschränkt. Externe DSLs besitzen in der Regel ein stark beschränktes Typsystem und erschweren Fehlermeldungen. Während interne automatisch auf die vollen Fähigkeiten der Hostsprache zurückgreifen können, muss man bei externen selbst dafür Sorge tragen, dass Zeileninformationen zu den Stellen gelangen, an denen sie mit hilfreichen, inhaltlichen Informationen angereichert werden können².

¹Diese Bachelorarbeit wurde in Zusammenarbeit mit der tarent solutions GmbH geschrieben.

²Es versteht sich, dass auch interne Sprachen nicht von Haus aus immer verständliche Fehlerinformationen liefern, sondern hier manuelles Nachbessern erforderlich werden kann.

2.2 Prototyp einer internen DSL

Wegen des wegfallenden Parsing-Aufwands und den besprochenen Features wurde zunächst auf eine zusätzliche Technologie verzichtet und eine prototypische interne DSL entwickelt, um abschätzen zu können, ob die mögliche Syntax den Anforderungen genügt. Falls sie als ungenügend beurteilt wird, wird eine externe DSL geschaffen. Da die *JVM* ([LY99]) das Einsatzgebiet des Validierungsprogramms darstellen wird und Java einen hohen syntaktischen Overhead und keine speziellen DSL-Fähigkeiten besitzt, wird Scala als Sprache für den Prototyp verwendet³. Scala bietet unter anderem mit impliziten Konvertierungen, faktisch überladbaren Operatoren und Semikolon-Interferenz viele DSL-freundliche Features⁴. Es ist zudem komplett Bytecode-kompatibel zu Java und erlaubt die Verwendung sämtlicher Java-Klassen.

Der Prototyp zeigte, ungeachtet der Syntax, die sich mit mehr Aufwand glatter gestalten ließe, ein unvorhergesehenes Problem, nämlich die mit acht Sekunden sehr lange Startzeit des Scala Interpreters. Umgehen könnte man dies, indem man die Beschreibungsklassen nach der Erstellung kompiliert und direkt den Bytecode einlesen lässt. Dieses Vorgehen hat drei nicht unerhebliche Nachteile. Zum einen muss man die Beschreibungen mit „import“-Anweisungen und Objektgerüsten versehen, was überflüssigen Ballast darstellt, und zum anderen ist die eingelesene Version nicht mehr menschenlesbar. Man kann zwar noch den Quelltext dem Archiv beilegen, muss sich aber dann darauf verlassen, dass er der aktuellen Version entspricht. Der dritte und schwerwiegendste Nachteil ist, dass jedes Projekt Scala als zusätzlichen Schritt in seine Buildkette aufnehmen muss, was den Beschreibungsprozess unnötig erschwert und die Akzeptanz mindert.

Als Konsequenz dessen wurde die interne DSL als Medium für den Anwender fallengelassen. Stattdessen werden nun die möglichen Technologien für eine externe DSL evaluiert.

2.3 Entwicklung einer externen DSL mit ANTLR

Bei der externen DSL steht man vor der Wahl, die Beschreibung entweder in einer vorhandenen Sprache zu formulieren oder eine eigene Sprache mit angepasster Struktur zu entwickeln. Es wurden die Dateiaustauschformate *XML*, *YAML* und *JSON* als populäre Vertreter bereits existierender Sprachen evaluiert. *XML* ist der wahrscheinlich bekannteste Vertreter dieser Kategorie und kann nicht nur mit sehr guter IDE-Unterstützung, sondern auch durch Vertrautheit punkten. Nichtsdestotrotz ist es eine sehr geschwätige Sprache und durch seine strikte, umständliche Syntax nicht für die Bearbeitung durch Menschen geschaffen. *YAML*⁵ ist eine Umsetzung von *XML* in einer kompakteren, freundlicheren Syntax, welche aber die Sprache ebenfalls durch seine strikte Struktur beschränkt. Auch *JSON* wird wegen seiner inflexiblen Struktur und dem Zwang, überall geschweifte Klammern und Anführungszeichen benutzen zu müssen, ausgeschlossen. Die Untauglich-

³Hier spielen auch Firmeninteressen mit hinein.

⁴Für näheres Interesse sei [Gho11] empfohlen.

⁵<http://www.yaml.org/>

keit der Sprachen wurde durch Verfassen von Beispielsbeschreibungen konstatiert. Durch das vorherige Erstellen des Domänenmodells konnte dabei sofort geprüft werden, ob es möglich war, alle Objekte mit den Beispielen zu befüllen.

Da die syntaktischen Einschränkungen existierender Sprache nicht in Kauf genommen werden konnten, blieb als letzte Alternative noch, eine eigene Sprache zu definieren und einen Parser für diese zu schreiben. Zu diesem Zweck existieren ausgereifte Werkzeuge, so dass man keinen Parser mühsam von Hand implementieren muss, sondern sogenannte Parser Generatoren verwenden kann. Mit einer selbst geschriebenen Grammatik ist man auf der sicheren Seite, was zukünftige Erweiterungen anbelangt und muss keine Sorge haben, von einer vorgegebenen Struktur eingeschränkt zu werden. Als wirklicher Nachteil dieser Lösung ist die fehlende IDE-Unterstützung anzuführen.

Die Wahl des Parsergenerators fiel wegen seiner weiten Verbreitung und der ausführlichen Dokumentation auf das ausgereifte ANTLR⁶. Zudem besitzt es eine IDE, die bei der Erstellung und dem Debugging von Grammatiken hilft.

3 Fazit

Als wesentliche Erkenntnisse aus der Abwägung zwischen internen und externen DSLs lässt sich festhalten, dass eine interne DSL zwar Vorteile durch die vertraute Hostsprache verspricht, aber in der Realität der Programmierstil doch so unterschiedlich ist, dass man sich neue Entwicklungsmuster erarbeiten und Regeln verletzen muss, um die Syntax den Vorstellungen anzupassen. Konkret gilt es auch zu berücksichtigen, dass Scala bisher nur mäßig für Einsatz als interpretierte Skriptsprache geeignet ist. Bei externen DSLs lohnt es sich mitunter, wie in diesem Fall, eine eigene Syntax zu kreieren, da der Mehraufwand durch ausgereifte Werkzeuge, wie z. B. Parser Generatoren, vernachlässigbar ist.

Literatur

- [Eva03] Eric Evans. *Domain-Driven Design: Tackling Complexity in the Heart of Software*. Addison-Wesley, Boston, MA, USA, 2003.
- [FP11] Martin Fowler und Rebecca Parsons. *Domain-Specific Languages*. Addison-Wesley, Boston, MA, USA, 2011.
- [Gho11] Debasish Ghosh. *DSLs in action*. Manning Publications Co., Greenwich, CT, USA, 2011.
- [LY99] Tim Lindholm und Frank Yellin. *The Java™ Virtual Machine Specification, Second Edition*. Prentice Hall, Upper Saddle River, NJ, USA, 1999.
- [vDKV00] Arie van Deursen, Paul Klint und Joost Visser. Domain-specific languages: an annotated bibliography. *SIGPLAN Not.*, 35:26–36, June 2000.

⁶<http://www.antlr.org/>

Migration eines web-basierten Projektmanagement- und Kollaborations-Portals

Mathias Gebbe

Hochschule Osnabrück

Fakultät Ingenieurwissenschaften und Informatik

mathias.gebbe@hs-osnabrueck.de

GI-Vertrauensdozent: Prof. Dr. Frank M. Thiesing

Betreuer der Arbeit: Dipl.Systemwiss. MSc. Bernhard E. Reiter, Sascha Wilde

Abstract: Die Intevation GmbH[1], ein unabhängiger Freie-Software-IT-Dienstleister mit dem Hauptsitz in Osnabrück, betreibt seit 2005 ein Projektmanagement- und Kollaboration-Portal *Wald*. In diesem Portal sind zurzeit 50 Freie-Software-Projekte mit insgesamt über 13000 Benutzern registriert. Das Portal wird im Internet auf der Webseite <http://wald.intevation.org> betrieben und erlaubt jedem, auch ohne Anmeldung, Software-Produkte herunterzuladen. *Wald* basiert auf GForge[2], eine Freie-Software bei der es sich um eine Abspaltung des ab Version 3 proprietären SourceForge[3] handelt. GForge stellt ein Portal zur Verfügung und integriert unter anderem Benutzerverwaltung, Issue-Tracking-System, Foren sowie verschiedene Freie-Software-Anwendungen wie z.B. Mailinglisten und Versionsverwaltung. Inzwischen ist aus GForge wiederum eine proprietäre Software entstanden. Um Verwirrung mit den Namen zu umgehen, wird die freie Version des ursprünglichen GForge seit 2009 als FusionForge weitergeführt[4]. Dieses Migration befasst sich mit dem Wechsel des veralteten angepassten GForge auf ein aktuelles individuell angepasstes FusionForge 5.1. Die Migration umfasst dabei die Herstellung aller Funktionalitäten und Anpassungen aus *Wald* in FusionForge und den Umzug aller Daten auf Datei- und Datenbankebene. Freie-Software bietet in diesem Anwendungsfall die Möglichkeit einer tiefen Inspektion und Manipulation der Daten auf verschiedensten Ebenen, die bei proprietärer Software fehlt. Ein wichtiger Fokus in diesem Projekt liegt in der Nachhaltigkeit der Migration, in der Kommunikation mit dem Entwickler-Team und dem Austausch mit der Freien-Software-Gemeinschaft.

1 Einleitung

Das Portal *Wald* dient zur Bereitstellung einer Infrastruktur für Freie Software Projekte, die aus Intevations-Projekten[1] (ob intern oder extern) entstehen. Es werden aber auch Projekte von externen Benutzern gehostet, deren behandelte Thematik Schnittpunkte mit der Intevation GmbH hat. *Wald* steht hier als eine Metapher für Projekte (Bäume), die zusammen stetig wachsen und im Verbund einen Wald ergeben (Leitspruch: *ever-growing free software*). Das Portal ist nicht nur zur firmeninternen Nutzung gedacht, sondern auch

dazu, die Freie-Software-Gemeinschaft[5] zu unterstützen. *Wald* basiert auf GForge in der Version 4.5.14-22, in dem individuelle Anpassungen vorgenommen wurden. FusionForge befindet sich aktuell (Stand 12/2011) in der Version 5.1.1-2. In Abbildung 1 ist der Zusammenhang zwischen den verschiedenen Versionen und *Wald* grafisch dargestellt. Die Migration ist notwendig, da GForge AS nicht als Freie Software weiterentwickelt wird und das aktuelle Portal veraltet ist. Als Firma, die Freie Software einsetzt, ist der Betrieb von proprietärer Software unerwünscht. Bei dem Portal handelt es sich um eine Anwendung, die aus dem Internet erreichbar ist, sie muss daher aktuellen Sicherheitsstandards gerecht werden. Dadurch erfordert das Portal Kompatibilität mit aktuellen Versionen der einzelnen Komponenten wie z.B. des Datenbank- und Webservers. Dies kann durch die aktuelle Installation von *Wald* nicht gewährleistet werden, da die Komponenten aus der alten Debian GNU/Linux- und GForge-Version nicht mehr unterstützt werden. Zusätzlich zur Migration soll eine Funktionserweiterung um ein Mercurial-Plugin (plattformunabhängiges, verteiltes Versionskontrollsystem) in *Wald* implementiert werden. Während des gesamten Migrationsprozesses und bei anschließender Unterstützung der Nutzer von *Wald* werden Fehler (Bugs) in FusionForge, sofern möglich, behoben und den Entwicklern gemeldet. Die eigenen Anpassungen an FusionForge werden als Nachbesserung (Patch) festgehalten und über die FusionForge Homepage[6] anderen Nutzern zur Verfügung gestellt. Die eigene angepasste Version von FusionForge wird als Debian-Pakete[7], zur Software Installation auf Debian-basierten Betriebssystemen, paketierte. Somit können im Falle einer neuen Entwicklerversion (Upstream-Version)[7] die eigenen Anpassungen (Patches) nachvollziehbar wieder integriert oder weggelassen werden, wenn die eigene Anpassungen in die Entwicklerversion übernommen wurde.

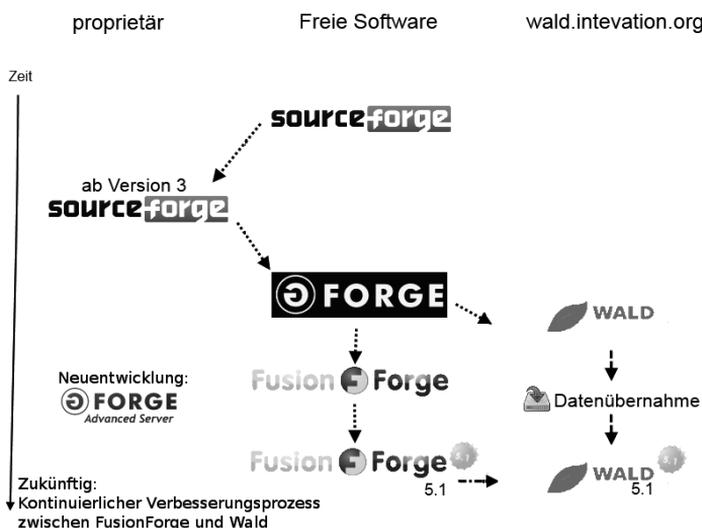


Abbildung 1: Zusammenhang zwischen SourceForge, GForge, FusionForge und *Wald*

2 Vorbereitung der Migration und Paketierung

Zur Vorbereitung wurde FusionForge 5.1 auf den funktionalen Gleichstand von *Wald* gebracht (z.B. Captcha-Eingabe bei Anmeldung) – hierzu waren Anpassungen notwendig. Diese Anpassungen lassen sich in Änderungen und Fehler (Bugs) unterteilen. Änderungen umfassen Erweiterungen und Abänderung von Funktionen in FusionForge. Die Fehlerbehebung bezieht sich auf *Bugs* in FusionForge, die während der Funktionstest aufgetreten sind. Diese Fehler wurden behoben (durch Anpassung der Quellen) und den Entwicklern gemeldet. Ziel ist nicht die Beseitigung aller Fehler, sondern die von den Anwendern genutzten Funktionalitäten von *Wald* in FusionForge zu gewährleisten. Darüber hinaus wurde das Corporate Design von *Wald* (Bilder, Logos, Texte) in die angepasste Installation integriert. Das Installationspaket wird aus den aktuellen Debian-Testing-Paketen[8] plus den eigenen Anpassungen (Patches) gebaut. Die Umgebung selbst, in der die neuen Debian-Pakete gebaut werden, ist in einem Mercurial-Repository abgebildet. Hierdurch können alle Änderungen gegenüber den Originalquellen detailliert nachvollzogen werden. Alle Änderungen und die aktuelle Versionsnummer werden in einem Änderungsprotokoll (Changelog) festgehalten. Die Versionsnummer setzt sich aus der aktuellen Paketversion (z.Z. 5.1.1-2) und der Version der Anpassungen (+wald01, +wald02 usw.) zusammen z.B. *fusionforge-full_5.1.1-2+wald01_all.deb*. Durch den Einsatz eigener Installationspakete ist die Wiederholbarkeit der Installation auf anderen Systemen gewährleistet.

3 Migration

Nach der Installation der eigenen Pakete auf dem Ziel-System, kann die Migration der Daten durchgeführt werden. Die Daten aus *Wald* umfassen Informationen aus Datenbank und Dateien die insgesamt ein Volumen von ca. 50 GiB haben. Damit die Migration auf die neue Version möglich wird, ist eine Aufbereitung, Anpassung und Transfer der Daten notwendig. Die Daten umfassen hierbei System-Daten, Benutzer-Daten und Projekt-Daten. Der Transfer dieser Daten findet direkt vom Quell-System statt und wird verschlüsselt auf das Ziel-System übertragen. Die Datenbank wird mithilfe eines Migrations-Skript auf die aktuelle FusionForge-Version angepasst und alle Dateien in das neue File-System transferiert und aufbereitet. Dies geschieht durch manuelle Anpassung anhand von vorher erstellten Anleitungen. Eine Automatisierung ist nur bedingt möglich und nicht notwendig, da diese spezifische Migration nur einmal stattfinden wird. Wenn das Ziel-System mit den transferierten Daten einwandfrei funktioniert, kann die Umschaltung des Produktiv-Systems vorbereitet werden. Alle Benutzer werden durch die Homepage und Mailinglisten über die anstehende Migration zum Zeitpunkt X informiert. Ab diesem Zeitpunkt wird *Wald* abgeschaltet, sodass keine Änderungen mehr stattfinden können. Nachdem alle geänderten und neuen Daten transferiert wurden, wird der DNS auf die IP-Adresse das neue Ziel-System umgeschaltet. Ab diesem Zeitpunkt findet die klassische Unterstützung der Anwender statt in der ggf. auftretende Fehler behoben werden. Ein Fall-Back Szenario ist im äussersten Notfall möglich – in erster Linie wird allerdings versucht alle auftretenden Fehler und Mängel unmittelbar zu beseitigen.

4 Fazit

Durch die Migration wird *Wald* auf einen aktuellen Stand der einzelnen Softwarekomponenten gebracht. Somit sind vor allem die Vorteile einer aktuellen und zusätzlichen Versionsverwaltung direkt für die Klienten nutzbar. Die Migration beseitigt viele Fehler und integriert neue Funktionalität für Benutzer und Administratoren im Portal. In dem Projekt wurden Vorteile von Freier Software nutzbar die sich durch Nachhaltigkeit und Flexibilität auszeichnen. Sie bietet die Möglichkeit, eine von vielen getestete, individuelle Lösung für die Migrationsproblematiken zu schaffen. Keine Daten aus dem Altsystem sind von der Analyse und Inspektion ausgeschlossen und können in jeglicher Form übernommen, manipuliert und aufbereitet werden. Durch die Anpassungen der Prozesskette zur Erstellung der Installationspakete und der jetzigen Nähe zur Entwickler-Version ist in Zukunft die Implementierung neuerer Versionen gesichert.

Literaturverzeichnis und Referenzen

- [1] Intevation GmbH
Neuer Graben 17
49074 Osnabrück
<http://www.intevation.org>
- [2] GForge - Collaborative Development Environment
<http://www.gforge.org>
(Abgerufen am 09.01.2012)
- [3] SourceForge - the largest Open Source applications and software directory - SourceForge.net
<http://www.sourceforge.net>
(Abgerufen am 09.01.2012)
- [4] GForge is now FusionForge
Erstellt von Roland Mas Date: 2009-01-25 20:06
gesichert mit Webcitation <http://www.webcitation.org/62NPYpcMf>
(Abgerufen am 11.10.2011)
- [5] Freie Software - Zwischen Privat- und Gemeineigentum
von Volker Grassmuck
Verlag: Bundeszentrale für politische Bildung Bonn, (2002) ISBN-13: 978-3893314324
- [6] FusionForge Tracker/Homepage Kommunikationsplattform um den Entwicklern Bugs, Patches oder Feature-requests für neue Version mitzuteilen
https://fusionforge.org/tracker/?group_id=6
- [7] Debian GNU/Linux 4 Anwenderhandbuch. Für Einsteiger, Umsteiger und Fortgeschrittene von Frank Ronneburg
Verlag: Addison-Wesley, München; Auflage: 1 (18. Oktober 2007)
ISBN-13: 978-3827325235
- [8] Debian GNU/LINUX 7.0 (Wheezy) Paketquellen
<http://packages.debian.org/wheezy/>
(Abgerufen am 09.01.2012)

Videounterstützte Ablaufverfolgung von Tests für Anwendungen mit grafischer Benutzeroberfläche

B.Sc. Helge Holzmann
research@helgeholzmann.de

Bachelorarbeit vom 19.08.2011
Leibniz Universität Hannover
Fachgebiet Software Engineering
Prüfer: Prof. Dr. Kurt Schneider
Betreuer: Dipl.-Math. Raphael Pham

Abstract: Dieser Artikel stellt die wesentlichen Konzepte zur videounterstützten Ablaufverfolgung von automatisierten GUI-Tests vor. Durch dieses Verfahren können Testabläufe rekonstruiert und die aufgetretenen Fehler lokalisiert werden. Die Konzepte wurden durch eine Implementierung in einem großen Unternehmen evaluiert. Zur Realisierung der Ablaufverfolgung wird dabei eine Zwischencode-Instrumentierung der .NET-Programme, die zur Testdurchführung dienen, mit Ansätzen der Aspektorientierten Programmierung durchgeführt. Bei der Bildschirmaufzeichnung wurde besonderer Wert auf eine hohe Abdeckung der sichtbaren Aktionen und eine effiziente Speicherung der Aufzeichnungen gelegt.

1 Einleitung

Anwendungen mit grafischer Benutzeroberfläche (GUI) stellen aufgrund ihrer komplexen Bedienmöglichkeiten eine besondere Herausforderung beim Testen dar. Insbesondere Regressionstests, die automatisiert ablaufen, sind häufig schwer zu realisieren. Durch spezielle Testrahmen sind jedoch auch dabei aufwendige Bedienungen durchführbar.

Mit dem Testen von Anwendungen geht auch der Umgang mit dem Fehlverhalten der zu testenden Software, des Prüflings [Sch07, S.25], einher. Schlägt ein Test fehl, so muss laut [MBTS04, S.157ff.] zuerst der entsprechende Fehler im Prüfling lokalisiert werden, um danach behoben werden zu können. [Sch07, S.90] differenziert dabei die Begriffe Failure (beobachteter Fehler) und Defect (Fehler im Code, der zu dem sichtbaren Problem geführt hat). Zur Lokalisierung des Defects im Prüfling ist es hilfreich erst einmal den Failure zu ermitteln. In vielen Fällen tritt dieser jedoch weit vor dem Abbruch des Tests auf, da er nicht sofort erkannt wird und erst dessen Folgen zum Fehlschlagen des Tests führen.

In Rahmen der Bachelorarbeit [Hol11] wurde deshalb ein Verfahren erarbeitet und implementiert, um den Ablauf von GUI-Tests videounterstützt zu protokollieren. Das Projekt wurde bei Capgemini¹ in Hannover durchgeführt, wo die Software bereits während

¹ <http://www.capgemini.de>

der Entwicklung regelmäßig in realen Testsituationen eingesetzt wurde. Sie zeichnet den Bildschirm während eines Tests in einem Video auf und synchronisiert dieses mit den Bediensritten in Form der Anweisungen zur Steuerung des Prüflings. Durch eine spezielle Viewer-Anwendung können die Tests nach der Aufzeichnung ähnlich wie mit einem Debugger schrittweise betrachtet und nachvollzogen werden. So kann ein fehlgeschlagener Test bis zu dem Zeitpunkt zurückverfolgt werden, an dem ein Failure erkennbar ist. Außerdem kann dieser auf den korrespondierenden Bedienschritt zurückgeführt werden.

2 Konzepte und Implementierung

2.1 Ablaufverfolgung

Eine Schwierigkeit bei der Implementierung war, dass der betrachtete Testrahmen die Steuerung des Prüflings mit eigenständigen .NET²-Programmen durchführt. Zur Ablaufverfolgung wurden diese mit sogenannten Tracing-Anweisungen instrumentiert, um die Positionen im Testablauf zu protokollieren. Somit war eine Manipulation des CIL³-Codes der Programme notwendig. Die dabei erarbeiteten Verfahren sind analog auf andere verwaltete Programmiersprachen übertragbar.

Bei der Ablaufverfolgung einer Software sind viele Ansätze der Aspektorientierten Programmierung (AOP) anwendbar. Dies ist ein Konzept zur Trennung der Hauptanliegen einer Software von nebensächlichen Aufgaben. In der AOP werden die verschiedenen Anliegen als Concerns bezeichnet. Dabei wird zwischen den zwei genannten Arten unterschieden. [Böh06, S.17] nennt diese fachliche und nicht fachliche Concerns. Als fachliche Concerns sind bei GUI-Tests die Anweisungen zur Bedienung des Prüflings zu betrachten. Nicht-fachliche Concerns ergeben sich durch die Ablaufverfolgung und bezeichnen die Tracing-Anweisungen zur Protokollierung des Testablaufs. Die Verknüpfung beider Concerns wird als Weaving bezeichnet [KLM⁺97, S.228ff.]. [KLM⁺97, S.228] unterscheiden hierbei zwischen Runtime-Weaving und Compile-Time-Weaving. [Wun05, S.65] differenziert Compile-Time-Weaving außerdem in Source- und Bytecode-Weaving. Zusätzlich erwähnt er das sogenannte Load-Time-Weaving, wobei das Weaving ebenfalls in den Byte- bzw. Zwischencode geschieht, jedoch erst, wenn dieser geladen wird. Mit einer Implementierung des Load-Time-Weavings wurde bei der Instrumentierung ein dynamisches Verhalten erreicht, wodurch vor jedem Testlauf entschieden werden kann, ob eine Ablaufverfolgung der Tests durchgeführt werden soll.

2.2 Bildschirmaufzeichnung

Die größte Schwierigkeit bei der Bildschirmaufzeichnung war es, eine hohe Abdeckung der Aktionen auf dem Bildschirm zu erreichen, ohne dabei identische Bilder zu speichern,

² Das Microsoft .NET-Framework

³ Common Intermediate Language, Zwischencode-Sprache der .NET-Plattform

auf denen der Bildschirm keine Veränderung aufweist. Es war also ein Konzept zu erarbeiten, um zu bestimmen, wann sich der Bildschirm verändert und somit aufgezeichnet werden muss. Dabei wurden zeitgetriebene, codegetriebene und ereignisgetriebene Ansätze zur Aufzeichnung sowie bereits existierende Softwarelösungen diskutiert.

Letztendlich führten die Recherchen zu VNC⁴-Programmen. Diese ermöglichen die Steuerung eines entfernten Computers über ein Netzwerk und übertragen dazu dessen Bildschirm an den fernsteuernden Rechner. Auch hierbei wird eine hohe Abdeckung angestrebt. Außerdem muss die Aufzeichnung sehr effizient geschehen, da die Bildschirmdaten möglichst in Echtzeit übertragen werden sollen. Um dies zu erreichen, werden in VNC-Programmen häufig sogenannte Mirror-Driver zur Ermittlung der Änderungen auf dem Bildschirm eingesetzt.

Durch den Einsatz eines solchen Treibers wurde ermöglicht, dass lediglich die Änderungen des Bildschirms gespeichert werden müssen, was eine deutliche Zeit- und Speicherersparnis gegenüber einer Kopie des gesamten Bildschirms darstellt. Außerdem ermöglichte die Speicherung dieser kleineren Segmente ein effizienteres Auffinden von mehrfach aufgezeichneten Änderungen. Speziell bei GUI-Tests tritt dies häufig auf, da die meiste Zeit auf der gleichen grafischen Oberfläche gearbeitet wird und sich dabei viele Situationen wiederholen. In dazu durchgeführten Versuchen hat es sich bei ca. 40 % der erfassten Segmente um Duplikate gehandelt. Durch den Einsatz einer Hash-Map wurde erreicht, dass bereits vorhandene Segmente schnell wiedergefunden und somit nicht erneut gespeichert werden müssen. Dazu wird ein Hash-Wert über die binären Bilddaten gebildet. Um dies zu beschleunigen, wurden die Segmente zuvor in das freie PNG-Format komprimiert. Auf diese Weise ist es möglich, weiterhin alle Aktionen auf dem Bildschirm zu erfassen und gleichzeitig den Speicherplatz deutlich zu reduzieren.

3 Zusammenfassung und Fazit

Die vorgestellten Konzepte ermöglichen die angestrebte videounterstützte Ablaufverfolgung von GUI-Tests. Die Ansätze der AOP bilden dafür die ideale Grundlage. Mit dem Mirror-Driver wurde eine Möglichkeit gefunden, nur die geänderten Segmente, statt des gesamten Bildschirms zu speichern. So ist es gelungen, die Aktionen auf dem Bildschirm mit einer sehr hohen Abdeckung bei einem trotzdem relativ niedrigen Speicherverbrauch aufzuzeichnen.

Die Grenzen des Systems liegen bei der Lokalisierung des Defects im Prüfling. Nachdem der Failure erkannt wurde, ist somit noch ein weiterer Schritt durchzuführen. In den Versuchen hat sich jedoch gezeigt, dass dieser durch eine Identifizierung des Failures auf dem Bildschirm mit dem korrespondierenden Bedienschnitt deutlich unterstützt wird.

Neben der beschriebenen Implementierung und dem Viewer zur Betrachtung der Aufzeichnungen wurde außerdem eine spezielle Messanwendung entwickelt. Mit dieser wurde ermittelt, dass die ausgeführten Bedienschritte bei der Ablaufverfolgung vollständig

⁴ Virtual Network Computing

überdeckt werden und bei der Bildschirmaufzeichnung eine Abdeckung von bis zu 30 Aktionen pro Sekunde möglich ist. Alle Beteiligten des Projektes zeigten sich mit diesen Ergebnissen sehr zufrieden und ein produktiver Einsatz der Software ist bereits geplant. Dabei kann evaluiert werden, inwiefern das Verfahren auch in anderen als den bereits betrachteten Testfällen erfolgreich nutzbar ist.

Literatur

- [Böh06] O. Böhm. *Aspektorientierte Programmierung mit AspectJ 5: Einsteigen in AspectJ und AOP*. Dpunkt, 2006.
- [Hol11] H. Holzmann. Videounterstützte Ablaufverfolgung von Tests für Anwendungen mit grafischer Benutzeroberfläche. Bachelorarbeit, Leibniz Universität Hannover, Fachgebiet Software Engineering, 2011.
- [KLM⁺97] G. Kiczales, J. Lamping, A. Mendhekar, C. Maeda, C. Lopes, J.M. Loingtier und J. Irwin. Aspect-oriented programming. *ECOOP'97Object-Oriented Programming*, Seiten 220–242, 1997.
- [MBTS04] G.J. Myers, T. Badgett, T.M. Thomas und C. Sandler. *The art of software testing*. John Wiley & Sons, 2004.
- [Sch07] K. Schneider. *Abenteuer Softwarequalität: Grundlagen und Verfahren für Qualitätssicherung und Qualitätsmanagement*. Dpunkt, 2007.
- [Wun05] L. Wunderlich. *AOP: Aspektorientierte Programmierung in der Praxis*. Entwickler.press, 2005.

Eine empirische Studie mit Java und Groovy über den Einfluss von statischen Typsystemen auf Entwicklereffizienz bei der Benutzung und Anpassung von Softwaresystemen

Sebastian Kleinschmager

Universität Duisburg-Essen
Institut für Informatik und Wirtschaftsinformatik
sebastian.kleinschmager@stud.uni-due.de

Art der Arbeit: Masterarbeit

Betreuer/in der Arbeit: Dr. Stefan Hanenberg

Abstract: Im Rahmen dieser Masterarbeit wurde ein empirisches Programmierexperiment konzipiert, durchgeführt und ausgewertet. Verglichen wurde die Entwicklungseffizienz der Teilnehmer bei Lösung verschiedener Aufgaben mit einer undokumentierten API in Abhängigkeit von der Nutzung entweder eines statischen oder eines dynamischen Typsystems. Die Ergebnisse sprechen grundsätzlich für das statische Typsystem.

1 Einleitung und Motivation

Statische Typsysteme sind schon seit einigen Jahren im Einsatz und sind auch Bestandteil einiger sehr verbreiteter Programmiersprachen (z.B. Java, C#, C++). Diese Form von Typsystemen soll dem Entwickler gegenüber einem dynamischen Typsystem eine Reihe von Vorteilen bringen, unter anderem auch die Effizienz steigern. Die Meinungen hierzu sind kontrovers, wie z.B. in [Ha10] nachgelesen werden kann. Leider gibt es trotz ihrer Relevanz nur wenige empirische Studien zu diesem Effekt statischer Typsysteme (z.B. [Ha10], [Ma11] oder [St11]). Daher wurde im Rahmen dieser Arbeit ein kontrolliertes Experiment konzipiert, welches anhand unterschiedlicher Aufgaben den Einfluss des Typsystems auf die Entwicklungseffizienz zu messen versuchte.

2 Experimentaufbau

Die Hauptfrage des Experiments war -wie bereits erwähnt- die nach der Leistung eines Entwicklers in Abhängigkeit des Typsystems welches er während der Programmierung nutzt. Dementsprechend ist die Haupthypothese:

- Hypothese 1: Ein Entwickler löst eine Aufgabe schneller, wenn er eine Programmiersprache mit statischem Typsystem nutzt.

Ferner wurden noch zwei weitere Hypothesen mit in das Experiment aufgenommen, diese sind:

- Hypothese 2-1: Je weiter das Auftreten eines Laufzeitfehlers von dem Punkt entfernt ist, an dem er seinen Ursprung hat, desto länger dauert es, den Fehler zu finden und zu beheben (in dynamischen typisierten Sprachen)
- Hypothese 2-2: Es dauert gleich lang einen semantischen Fehler zu finden, unabhängig davon ob ein dynamisches oder statisches Typsystem genutzt wird.

Nach Definition der Hypothesen wurden insgesamt 9 Aufgaben zur ihrer Überprüfung erstellt, die die Teilnehmer des Experiments lösen mussten. Jeder Teilnehmer musste alle Aufgaben sowohl in der statischen (vertreten durch Java) als auch in der dynamisch typisierten Programmiersprache (vertreten durch Groovy) lösen, wofür jeweils eine eigene API zugrunde lag. Dabei hatte die erste Gruppe der Teilnehmer zuerst den Groovy Teil zu lösen (Groovy Starter genannt), dann den Java Teil, und die zweite Gruppe löste die Aufgaben in umgekehrter Reihenfolge (Java Starter).

Um die Varianz in den Experimentergebnissen gering zu halten, wurden diverse Maßnahmen getroffen. So wurde die in beiden Sprachen strukturell gleiche Anwendung durch Umbenennung von Codeelementen in einer Sprache in eine komplett andere Domäne überführt. Ein weiterer wichtiger Faktor des API Designs war, dass sie vollkommen undokumentiert sein sollte. Das bedeutete, dass neben fehlenden Kommentaren keine Variable durch Ihren Namen ausdrücken durfte, von welchem Typ sie war. Dazu wurden statt der Klassennamen meist Synonyme oder Ausdrücke aus der Sprache der Applikationsdomäne genommen (z.B. wurde statt *LevelType* ein *LevelKind* eingesetzt).

Hauptindikator war die gemessene Zeit die die Teilnehmer für eine Aufgabe brauchten. Diese konnte berechnet werden, weil die genutzte Entwicklungsumgebung den Fortschritt der Teilnehmer erfasste und so eine automatische Auswertung der Daten und eine statistische Analyse der Ergebnisse ermöglichte.

Im Design der Aufgaben war explizit ein Lerneffekt eingeplant gewesen. Bei den Teilnehmern die mit Groovy starten sollten wurde vermutet, dass beim 2. Durchlauf mit Java der positive Effekt des statischen Typsystems und der Lerneffekt aus dem 1. Teil sich aufaddieren würden. Andersherum sollte bei den Java Startern der negative Einfluss des dynamischen Typsystems den Lerneffekt aus dem 1. Teil wieder ausgleichen.

3 Experimentauswertung

Insgesamt nahmen 36 Teilnehmer am Experiment teil, 33 der Datensätze waren nutzbar. Die Daten wurden nach Experimentdurchführung mit der Statistiksoftware SPSS und gängigen statistischen Methoden ausgewertet [Fi09]. Um bessere Ergebnisse zu erzielen, wurde zusätzlich noch geprüft, ob es einen messbaren Reihenfolgeeffekt gab (d.h. ob ein Einfluss der Startsprache auf die Ergebnisse messbar war). Da dieser per Test nachgewiesen werden konnte, mussten die beiden Gruppen (Groovy Starter und Java Starter) separat analysiert werden und konnten nicht in eine gemeinsame Analyse genommen werden.

Die beiden Abbildungen 1 und 2 zeigen die Ergebnisse der Auswertung in Sekunden für die beiden Gruppen, aufgeschlüsselt nach Programmiersprache und Aufgaben.

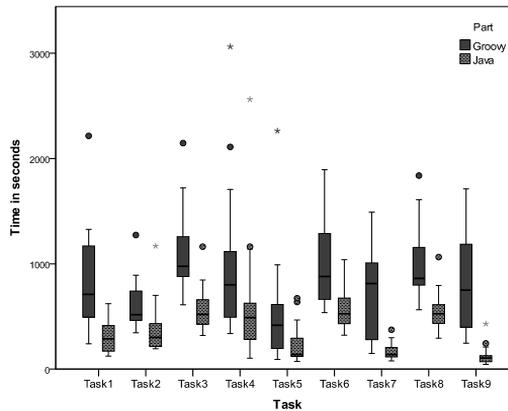


Abbildung 1: Boxplot der Ergebnisse für die Groovy Starter Gruppe

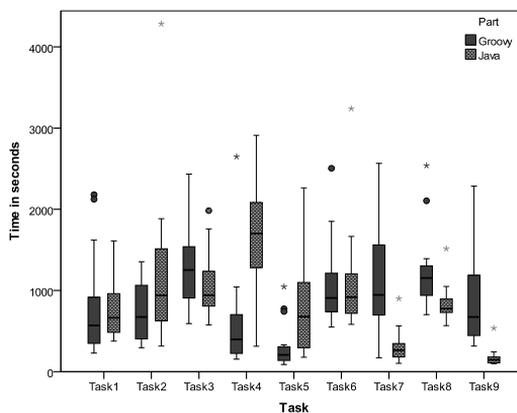


Abbildung 2: Boxplot der Ergebnisse für die Java Starter Gruppe

Relevant für die Haupthypothese waren die Aufgaben 1,2,3, 6 und 8, welche mit höherer Nummer auch aufwendiger wurden. Für die Groovy Starter Gruppe kann hierzu gesagt werden, dass alle Teilnehmer diese Aufgaben mit Java signifikant schneller lösten als mit Groovy. Etwas anders waren die Ergebnisse bei den Java Startern. Eine Signifikanz war hier nicht immer gegeben, was allerdings erwartet werden konnte, da sich besonders bei den weniger aufwendigeren Aufgaben der Lerneffekt aus dem 1. Teil und der negative Einfluss des dynamischen Typsystems auf die Effizienz grob ausglich. Die Ergebnisse beider Gruppen waren für die Haupthypothese 1 positiv zu bewerten. Grundsätzlich konnte bis auf eine Abweichung ein signifikant positiver Einfluss des statischen Typsystems auf die Lösung von Aufgaben gemessen werden. Dies galt sowohl für die Groovy Starter als auch für die Java Starter Gruppe. Neben den Auswertungen der Haupthypothese wurden noch diverse andere Analysen durchgeführt, u.a. für die beiden Nebenhypothesen 2-1 und 2-2. Beide mussten aufgrund von widersprüchlichen bzw. nicht signifikanten Daten verworfen werden. So dauerte z.B. das Lösen eines Fehlers, der eigentlich als schwieriger gekennzeichnet wurde, signifikant kürzer als das einer als leichter deklarierten Aufgabe. Für semantische Fehler wurde in einer Aufgabe ein signifikanter Unterschied, in einer anderen Aufgabe wiederum kein Unterschied zwischen beiden Programmiersprachen gemessen. Ein möglicher Lerneffekt der Teilnehmer oder unberücksichtigte Faktoren im Aufgabendesign könnte bei diesen beiden Hypothesen dazu geführt haben, dass die Ergebnisse unbrauchbar wurden.

4 Konklusion

Die Ergebnisse der Experimentauswertung sind für die Haupthypothese als zufriedenstellend zu bewerten. Es konnte ein signifikanter positiver Effekt von statischen Typsystemen auf die Zeit gefunden werden, die ein Entwickler benötigt um eine Aufgabe zu lösen. Allerdings kann das Ergebnis mit dem Hintergrund dass es erst sehr wenige Studien in diese Richtung gibt nur als ein kleiner Schritt gewertet werden. Zudem gibt es immer eine Reihe von Faktoren die die Validität der Ergebnisse eines Experiments gefährden können. Einige davon werden z.B. in [Ha10] diskutiert. Eine Wiederholung des Experiments in exakt dieser oder leicht abgewandelter Form könnte die Erkenntnisse erhärten und zu einem besseren Verständnis des Einflusses von Typsystemen auf die Programmierung führen.

Literaturverzeichnis

- [Fi09] Field, A. 2009. Discovering Statistics Using S PSS (Introducing Statistical Methods). Sage Publications Ltd.
- [Ha10] Hanenberg, S. 2010. An experiment about static and dynamic type systems: doubts about the positive impact of static type systems on development time. SIGPLAN Not 45, 22-35.
- [Ma11] Mayer, C. 2011. An empirical study of possible effects of static type systems on documentation - a controlled experiment with an undocumented application programming interface. Bachelor Thesis, University of Duisburg-Essen.

Konzeption und Realisierung eines Frameworks für transparente RPCs in PHP

Christian Lück

FH Aachen

Fachbereich Elektrotechnik und Informationstechnik

christian.lueck@alumni.fh-aachen.de

Art der Arbeit: Bachelor-Thesis

Betreuer: Prof. Dr. rer. nat. Heinrich Faßbender

Abstract: Im Backend moderner PHP-Web-Oberflächen wird häufig auf verschiedene Datenquellen zugegriffen, die aufgrund ihrer Verzögerung die Reaktionszeit der Oberfläche negativ beeinflussen. Diese Arbeit beschreibt eine mögliche Abhilfe durch einen parallelisierten Zugriff mittels RPCs. Neben der Analyse bestehender Lösungen und dem Aufzeigen ihrer Defizite, wird ein vollständig in PHP implementiertes Framework entworfen und auf Details der Realisierung eingegangen.

1 Motivation

Im Rahmen der Bachelor-Thesis wird ein tiefgehender Einblick in die Motivation, die Konzeption und die Realisierung eines neuen Frameworks für transparente RPCs in PHP geboten.

Die Firma Datus AG vertreibt mit der „indali“-Produktfamilie eine embedded Telefonanlage, die sich über den ebenfalls integrierten Apache-Webserver mittels einer Web-Oberfläche konfigurieren und verwalten lässt. Dabei wird PHP im Kontext eines (Apache-)Webserver ausgeführt und beantwortet eingehende HTTP-Requests. Da jeder Request von nur einem separaten PHP-Interpreter bearbeitet wird und PHP keine Möglichkeit bietet, parallele Handlungsstränge (Threads) auszuführen, können folglich auch die für die Beantwortung der Anfrage nötigen Befehle ausschließlich sequentiell abgearbeitet werden.

Ein typischer Request erfordert dabei häufig den Zugriff auf diverse (teils externe) Datenquellen, wie Datenbanken, Dateisysteme oder System-Dienste. Dabei wird ein Großteil der Bearbeitungszeit nicht durch die von PHP in Anspruch genommene Rechenzeit bestimmt, sondern durch das Warten auf die jeweilige Datenquelle (Wartezeiten summieren sich). In vielen Fällen ist eine sequentielle Abarbeitung jedoch nicht nötig, und kann durch einen parallelen Zugriff deutlich beschleunigt werden.

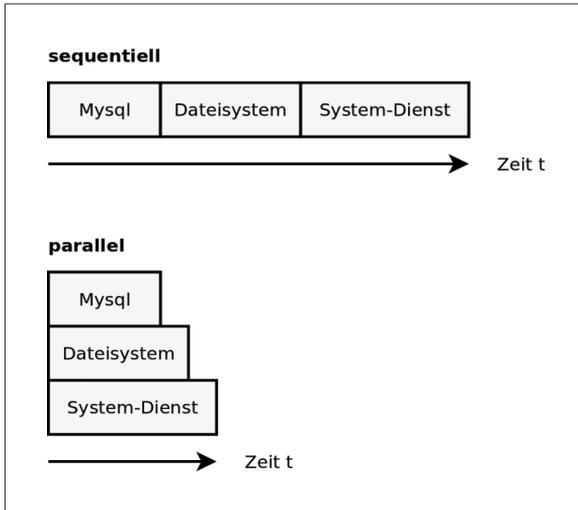


Abbildung 1: Ablaufdiagramm sequentielle vs. parallele Verarbeitung

2 Analyse

Um eine solche Parallelisierung der Teil-Operationen zu vereinfachen und thematisch einzuordnen, können diese ebenfalls wieder als Funktionen abgebildet werden. Der Begriff RPC (Remote Procedure Call) beschreibt in der Informatik ein Verfahren, mit dem Prozeduren (Funktionen bzw. Methoden) außerhalb des Prozess-internen Adressraums aufgerufen werden können. Hierbei werden RPCs an einen (lokalen) Netzwerkdienst übergeben. Über die Stream-Schnittstelle ist in PHP ein nicht-blockierender (asynchroner) Zugriff möglich, worüber die RPCs effektiv parallelisiert werden können.

Das RPC-Framework soll dafür eine vollständige Rahmenumgebung bereit stellen, um dem Entwickler eine einfache Schnittstelle (API) zu bieten. Das Framework koordiniert dabei intern den komplexen Ablauf des RPCs und verbirgt Implementierungs-Details.

Dabei sollen folgende Anforderungen erfüllt werden:

- Durch einen transparenten Ansatz soll eine einfache Integration in bestehende Systeme ermöglicht werden, d.h. RPCs sollen sich nicht von lokalen Funktionsaufrufen unterscheiden. Hierfür müssen sämtliche Sprachkonstrukte der objektorientierten Programmiersprache PHP unterstützt und durch das RPC-Framework bereitgestellt werden. Dazu zählen insbesondere die native, vollständig automatische Unterstützung der untypisierten und beliebig verschachtelbaren Datentypen als Funktions-Argumente und -Return-Werte, sowie Exceptions und Ausgaben auf dem Standard-Output.
- Dezentraler Ansatz mit Unterstützung externer RPC-Clients
- Bidirektionale Kommunikation

Tatsächlich existieren bereits unzählige Verfahren und Bibliotheken um RPCs in PHP auszuführen (wie beispielsweise SOAP, XML-RPC, Gearman, etc.). Wie in der Thesis ausführlich dargestellt, erfüllt jedoch keine dieser Lösungen die obigen Anforderungen. Insbesondere integrieren sich diese meist schlecht in eine reine PHP-Umgebung und erfordern hierdurch eine komplizierte und damit aufwändige Anpassung der Umgebung.

3 Ergebnisse

Aufgrund der Mängel bestehender Lösungen wird ein neues Framework geschaffen, welches den Ansporn hat, die besten Konzepte der bestehenden Lösungen zu bündeln. Im weiteren Verlauf der Arbeit wird das hierfür nötige Kommunikations-Protokoll sowie die RPC-Schnittstellen (API) definiert und die dabei nötige Projekt- und Klassenstruktur dokumentiert.

Das Ergebnis ist ein in rund 4000 SLOC und in 30 Klassen vollständig in PHP implementiertes, Event-getriebenes Framework, das durch seinen modularen Aufbau optimal auf die Anforderungen angepasst wurde und sich leicht einsetzen und erweitern lässt. Das Framework wurde unter der MIT-Lizenz veröffentlicht [URL: Worker] und befindet sich aktuell im produktiven Einsatz. Zusätzlich wird auf eine über die Bachelor-Thesis hinausgehende geplante Erweiterung des RPC-Frameworks eingegangen.

Literaturverzeichnis

[URL: Datus]

http://www.datus.com/?option=com_content&view=article&id=101

Unternehmensvorstellung Datus AG

[URL: Indali]

http://www.datus.com/?option=com_content&view=article&id=48

Produktbeschreibung indali OBX

[URL: Usability]

<http://www.useit.com/alertbox/response-times.html>

Usability Engineering: Website Response Times

Jakob Nielsen, 2010-06-21

[URL: Comet]

<http://cometdaily.com/2007/10/19/comet-and-push-technology/>

Comet and Push Technology

Alessandro Alinone, 2007-10-19

[URL: Worker]

<https://github.com/clue/Worker>

Worker - Projekt-Homepage

[URL: dnode]

<https://github.com/substack/dnode>

dnode – Freestyle RPC for node.js

[URL: EventDriven]

<http://c2.com/cgi/wiki?EventDrivenProgramming>

Event Driven Programming

Entwicklung und Untersuchung einer Softwarearchitektur für medizinische Anwendungen am Beispiel von Hüftgelenken

Benjamin Saul

Martin-Luther-Universität Halle-Wittenberg
Institut für Informatik
benjamin.saul@student.uni-halle.de

Abstract: Das Ziel in dieser Bachelorarbeit war es, Ansätze einer Software im Bereich von medizinischen Untersuchungen vorzustellen und zu analysieren. Dazu erfolgt eine Herleitung der nichtfunktionalen Anforderungen an die Architektur. Zu diesen gehört unter anderem die leichte Austauschbarkeit von Algorithmen. Der Zweck der Anwendung ist es, aus gegebenen 3D-Modellen des Hüft- und des Oberschenkelknochens die Bewegungsfreiheit zu berechnen. Außerdem sollen die Knochen verändert werden können, um mögliche Operationsergebnisse zu untersuchen.

Betreuer: Prof. Dr. Wolf Zimmermann
Art der Arbeit: Bachelorarbeit

1 Einleitung

Softwaregestützte medizinische Untersuchungen ermöglichen eine Vielzahl von Erkenntnissen. Durch bildgebende Verfahren wie die Computertomographie (CT) ist es möglich, vollständige Aufnahmen des Körpers zu erstellen. Dabei können auf den Bildern die Knochen erkannt und segmentiert werden. Aus diesen Segmentationsregionen können schließlich durch entsprechende Algorithmen frei bewegliche 3D-Modelle erstellt werden. An diesen können dann verschiedene Eigenschaften gemessen und berechnet werden.

Eine dieser Eigenschaften ist die Bewegungsfreiheit. Im hier betrachteten Beispiel wird von 3D-Modellen des Hüft- und des Oberschenkelknochens ausgegangen (Abbildung 1). Nun ist es Aufgabe der Software, zu berechnen, wie weit der Oberschenkel gedreht werden kann, ohne mit den anderen Knochen zu kollidieren. Die medizinische Zielsetzung hierbei ist es, statistische Untersuchungen an Hüftgelenken durchzuführen, um Beziehungen zwischen eingeschränkter Bewegungsfreiheit und Knochenablagerungen zu erkennen [KLTM⁺07]. Die Berechnung erfolgt in verschiedenen Stufen, bei denen es oft eine Auswahl an Algorithmen gibt, die z. T. verschiedene Ergebnisse erzielen. Am Beispiel der Berechnung des Rotationszentrums, also dem Punkt, um den der Oberschenkel gedreht wird, soll gezeigt werden, welche Anforderungen dadurch an die Architektur gestellt werden.

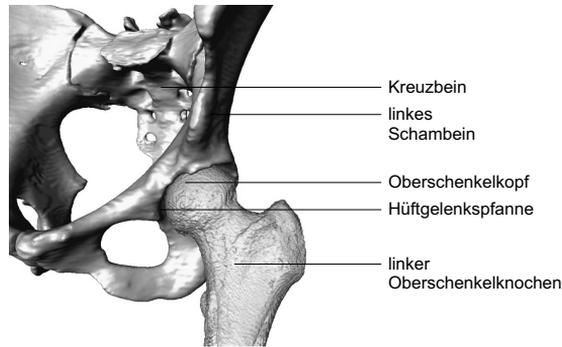


Abbildung 1: An der Rotation des Oberschenkels beteiligte Knochen

2 Anforderungen an die Architektur

Ausgangspunkt der Untersuchungen sind wie oben erwähnt die 3D-Modelle der Hüft- und Oberschenkelknochen. Diese wurden mit einer Visualisierungssoftware aus den segmentierten Regionen erstellt. Der erste Schritt in der Berechnung der Bewegungsfreiheit besteht darin, das Rotationszentrum zu ermitteln. Dazu stehen verschiedene algorithmische Ansätze zur Verfügung. Einige werden beispielsweise vorgestellt in [TKLM⁺07] und [BHH06]. Beide Algorithmen liefern im Allgemeinen verschiedene Ergebnisse, da diesen verschiedene Modellannahmen zugrunde liegen. Beim ersten Verfahren wird ausgehend vom Oberschenkelkopf eine Kugel angenähert, deren Zentrum als das Rotationszentrum angenommen wird. Beim Zweiten wird ausgehend von anatomischen Landmarken auf dem Hüftknochen das Rotationszentrum berechnet. Dies zeigt deutlich, dass für verschiedene Algorithmen verschiedene Eingabedaten vorliegen müssen.

Da viele Berechnungen auf Modellannahmen beruhen sind Abweichungen von den echten Werten nicht vermeidbar. Das bedeutet gleichermaßen, dass einige Algorithmen präzisere Ergebnisse liefern als andere. Um nun bei den Untersuchungen nicht auf fest implementierte, eventuell ungenaue Algorithmen angewiesen zu sein, ist es wichtig, dass die Architektur ein einfaches Auswechseln von Algorithmen erlaubt. Im Idealfall sollte es dem Benutzer der Software möglich sein, zu wählen, welchen Algorithmus man anwendet.

Um die Ergebnisse der Berechnung dem Nutzer zu veranschaulichen bietet sich eine ausführliche Visualisierung der einzelnen Werte an. Hierzu kann man als Entwickler auf verschiedene Visualisierungsbibliotheken zurückgreifen [HJ04]. Diese liefern gute Ergebnisse und verfügen z. T. über eine Vielzahl an Funktionen. Die Bibliotheken sind zumeist mit einer Pipes-and-Filter-Architektur umgesetzt. Ob diese auch für die hier entwickelte Anwendung geeignet ist, wird im nächsten Abschnitt diskutiert. Zunächst wird jedoch auf die Umsetzung des Einsatzes von verschiedenen Algorithmen eingegangen.

3 Umsetzung

Insgesamt lässt sich die Architektur in drei Schichten einteilen: Oberfläche, Fachdomäne und Infrastruktur. Die Aufgabenbereiche der einzelnen Schichten sind in Tabelle 1 zusammengetragen. Die hier besprochenen Bestandteile befinden sich alle in der Fachdomäne, dem eigentlichen Hauptbestandteil des Programmes. Sie ist auch als Bibliothek auffassbar, mit deren Hilfe die Oberfläche umgesetzt werden kann. Durch entsprechende sinnvolle Aufteilung der Komponenten in der Fachdomäne ist es möglich, jeder wichtigen Komponente ein eigenes Widget, also Interaktionsfenster, zuzuordnen. So wurde hier beispielsweise eine Gelenkkomponente entwickelt, die ihre verschiedenen Parameter als Schnittstelle anbietet. Diese können dann vom Nutzer interaktiv verändert werden, um entsprechende Zusammenhänge zu verdeutlichen. Beispielsweise kann man das Drehzentrum manuell setzen und so dessen Einfluss direkt nachvollziehen.

Schicht	Zuständig für
Benutzeroberfläche	Visualisierung der Ergebnisse Benutzerinteraktion
Fachdomäne	Verwalten der Arbeitsobjekte Bereitstellen von Bearbeitungsoperationen
Infrastruktur	Dateiverwaltung Integration von Fremdsystemen

Tabelle 1: Überblick über die Zuständigkeiten der einzelnen Schichten

Die Umsetzung der einzelnen Komponenten orientiert sich stark an den aus [GHJV94] bekannten Entwurfsmustern. Für die Gewährleistung der Austauschbarkeit der Algorithmen bietet sich das Strategie-Muster an. Dabei werden die unterschiedlichen Algorithmen durch Unterklassen einer abstrakten Algorithmenklasse implementiert. Zur Laufzeit wird dann der berechnenden Komponente ein entsprechender Algorithmus zugeteilt. Wichtig ist hierbei, dass die Algorithmen unterschiedliche Eingangsdaten benötigen. Unter Umständen liegen diese für einige Ansätze gar nicht vor, da z. B. die anatomischen Landmarken erst gesetzt werden müssen.

Es wäre umständlich zu fordern, dass alle Eingabedaten für alle Algorithmen vorliegen bzw. bekannt sind. Daher wurde bei der Umsetzung ein dem Mediator-Muster ähnliches Entwurfsmuster angewandt. Prinzipiell werden die benötigten Daten von einer übergeordneten Instanz erfragt, die alle Daten einsehen kann. Sind diese vorhanden, so können sie geladen und verarbeitet werden. Falls nicht, so kann eine Meldung an den Nutzer generiert werden. Dies gibt ihm die Möglichkeit, die erforderlichen Daten selbst nachzutragen.

Diese Nutzerinteraktion ist für medizinische Untersuchungen unverzichtbar, da nur so ein Verständnis für die Ergebnisse erzeugt werden kann. Alle Berechnungen müssen einsichtig sein. Der Nutzer muss also wissen, wie ein Ergebnis zustande gekommen ist. Dies ist auch z. T. Aufgabe der Architektur, da ein unverständliches Programm nicht nachvollzogen und auch nicht weiterentwickelt werden kann. Aus diesem Grund wurde hier auch nicht auf bereits bestehende Frameworks eingegangen. Diese ermöglichen zwar ein schnelles Entwickeln von neuartigen Algorithmen, jedoch besteht die Gefahr des Übersichtsverlusts.

4 Zusammenfassung und Ausblick

In diesem Artikel wurden einige Ergebnisse der Bachelorarbeit vorgestellt. Ziel war es, eine Architektur für medizinische Anwendungen zu konzipieren. Das Prinzip der Anwendung wurde erklärt und die wichtigsten Anforderungen an die Architektur geklärt. Dazu gehört unter anderem die Austauschbarkeit der Algorithmen, da deren Ergebnisse bei der Forschung untereinander verglichen werden müssen. Die Anforderung wurde mit dem Strategie-Entwurfsmuster unter Hilfe eines Mediators umgesetzt. Insgesamt ist die Architektur der Anwendung in Schichten geteilt, um zwischen Darstellung und zugrunde liegenden Berechnungen zu unterscheiden. Die Komponenten sind den entsprechenden Schichten zugeteilt, um die Wartung und Weiterentwicklung zu vereinfachen. Entstanden ist eine Software, die nach einiger weiterer Entwicklung für medizinische Untersuchungen verwendet werden kann. Das heißt, es werden neue Algorithmen implementiert und vorhandene verbessert. Außerdem werden weitere Funktionalitäten hinzugefügt.

Wie oben kurz erwähnt sind die Ergebnisse durch verschiedene Modellannahmen zustande gekommen. Da es meist nicht feststeht, ob die Annahme der Realität nahe kommt, werden reale Versuche zu Vergleichszwecken durchgeführt. Da diese aufwändig und teuer sind, ist es in Zukunft unabdingbar, dass wirklich nur korrekt umgesetzte Modelle geprüft werden. Dazu ist es erforderlich, bereits im Programm Tests durchzuführen, die die berechneten Werte überprüfen. Ein weiteres wichtiges Forschungsgebiet ist die Datenverwaltung. Bei den Untersuchungen werden ständig neue Daten ermittelt, z. B. CT-Aufnahmen, Segmentationsregionen, 3D-Modelle, Bewegungsfreigrade. Diese müssen alle verwaltet und vor allem gut zugreifbar sein. Außerdem ist eine automatische Statistikerstellung denkbar.

Insgesamt ist das Gebiet der medizinischen Untersuchungen ein dankbares Thema für verschiedene Bereiche der Informatik, wie beispielsweise Software-Engineering, Bildverarbeitung, Computergrafik und Datenbanken. Daher wird es auch in Zukunft Neu- und Weiterentwicklungen geben.

Literatur

- [BHH06] U. Boudriot, J. Hilgert und F. Hinrichs. Determination of the rotational center of the hip. *Archives of orthopaedic and trauma surgery*, 126(6):417–420, 2006.
- [GHJV94] Erich Gamma, Richard Helm, Ralph Johnson und John M. Vlissides. *Design Patterns: Elements of Reusable Object-Oriented Software*. Addison-Wesley Professional, 1. Auflage, 11 1994.
- [HJ04] Charles D. Hansen und Chris R. Johnson. *Visualization Handbook*. Academic Press, 1. Auflage, 12 2004.
- [KLTM⁺07] M. Kubiak-Langer, M. Tannast, SB Murphy, KA Siebenrock und F. Langlotz. Range of motion in anterior femoroacetabular impingement. *Clinical orthopaedics and related research*, 458:117, 2007.
- [TKLM⁺07] M. Tannast, M. Kubiak-Langer, SB Murphy, TM Ecker, M. Puls, F. Langlotz und KA Siebenrock. Computer-assisted Simulation of Femoro-acetabular Impingement Surgery. *Navigation and MIS in Orthopedic Surgery*, Seiten 440–447, 2007.

Numerische Optimierung eines Ölsaugkanals mithilfe adjungierter Topologieoptimierung

Markus Towara*

LuFG Informatik 12: Software and Tools for Computational Engineering, RWTH Aachen,
in Zusammenarbeit mit der Volkswagen AG, Werk Kassel - Getriebeentwicklung

1 Einleitung

Ich habe mich im Rahmen meiner Diplomarbeit mit der automatischen Formoptimierung eines Ölsaugkanals im Inneren der Mechatronikeinheit eines Doppelkupplungsgetriebes [AG08] beschäftigt. Dazu wurde eine bestehende Geometrie modifiziert um die erzielbare Verbesserung im Druckverlust bestimmen zu können. Die Optimierung wurde mit dem Open-Source Strömungslöser OpenFOAM durchgeführt. Als Optimierungsverfahren wurde die adjungierte Topologieoptimierung gewählt, deren Grundlagen nachfolgend kurz erläutert werden.

2 Grundlagen

Laminare instationäre und inkompressible Strömungen Newtonscher Fluide können durch die Navier-Stokes-Gleichungen beschrieben werden. Dieses System partieller Differentialgleichungen kann geeignet diskretisiert und numerisch gelöst werden.

$$\mathbf{v} \cdot \nabla \mathbf{v} = \nu \nabla^2 \mathbf{v} - \frac{1}{\rho} \nabla p + \mathbf{b} \quad \text{Impulserhaltung} \quad (1)$$

$$\nabla \cdot \mathbf{v} = 0 \quad \text{Massenerhaltung} \quad (2)$$

Für die Optimierung wird ein künstlicher Strömungswiderstand $-\alpha \mathbf{v}$ in die Impulsgleichung (1) eingeführt. Dieser entspricht physikalisch einer Porosität, d.h. eine hohe Porosität (= niedriges α) entspricht geringem Strömungswiderstand und umgekehrt.

$$\mathbf{v} \cdot \nabla \mathbf{v} = \nu \nabla^2 \mathbf{v} - \frac{1}{\rho} \nabla p + \mathbf{b} - \alpha \mathbf{v} \quad (3)$$

Mit diesem Parameter α können nun Zellen im diskretisierten Rechengebiet bestraft und somit praktisch entfernt werden, wenn davon ausgegangen wird, dass sie eine kontraproduktive Auswirkung auf die Strömung haben. Somit kann die Geometrieform variiert werden ohne in jedem Optimierungsschritt die Vernetzung anpassen zu müssen.

*towara@stce.rwth-aachen.de

Um feststellen zu können welche Zellen bestraft werden sollen muss zuerst eine geeignete Kostenfunktion J definiert werden (z.B. der Druckverlust zwischen Eingang und Ausgang des Kanals). Anschließend werden Ableitungsinformationen dieser Kostenfunktion nach dem Strömungswiderstandsparameter α benötigt. Da α für jede einzelne Zelle definiert ist wird auch für jede Zelle i eine partielle Ableitung $\partial J/\partial\alpha_i$ benötigt. Dies ist für reale Probleme mit klassischen Ableitungsbestimmungsmethoden wie finiten Differenzen nicht mehr in vertretbarer Zeit berechenbar, da mit diesen Verfahren für jede Ableitung das komplette Differentialgleichungssystem (3+2) gelöst werden muss. Einen Ausweg bietet hier die adjungierte Ableitungsbestimmung, welche es ermöglicht die Sensitivität der Eingangsparameter (hier die Widerstandswerte α) bezüglich der Ausgangsparameter (hier die skalare Kostenfunktion) zu bestimmen. Da die Kostenfunktion eine skalare Größe liefert ist es möglich alle Sensitivitäten mit nur einer zusätzlichen Auswertung des DGL-Systems (3+2) zu bestimmen.

Die Sensitivität kann aus dem Skalarprodukt des Geschwindigkeitsfeldes \mathbf{v} und der neu eingeführten adjungierten Geschwindigkeit \mathbf{u} bestimmt werden.

$$\frac{\partial J}{\partial\alpha_i} = (v_i \cdot u_i)V_i \quad (4)$$

Hierbei bezeichnet V_i das Volumen der Zelle i .

Zur Bestimmung der adjungierten Geschwindigkeit \mathbf{u} werden aus den Impuls- und Massenerhaltungsgleichungen mittels Variationsrechnung die adjungierten Impuls- und Massenerhaltungsgleichungen hergeleitet [Oth08]. Sie lauten:

$$-2\mathbf{D}(\mathbf{u})\mathbf{v} = \nu\nabla^2\mathbf{u} - \nabla q - \alpha\mathbf{u} \quad (5)$$

$$\nabla \cdot \mathbf{u} = 0 \quad (6)$$

mit $\mathbf{D}(\mathbf{u}) = \frac{1}{2} (\nabla\mathbf{u} + (\nabla\mathbf{u})^T)$.

3 Implementierung in OpenFOAM

Sowohl die primalen (3+2) als auch die adjungierten (5+6) Differentialgleichungen werden in OpenFOAM mit finiten Volumen diskretisiert und iterativ gelöst. Dabei wird nicht die Konvergenz der einzelnen Gleichungen abgewartet sondern die Gleichungen werden jeweils wechselnd gelöst. Somit existieren ab dem ersten Iterationsschritt Schätzungen für alle zu bestimmenden Größen und die Optimierung kann sofort mit den vorläufigen Werten begonnen werden (vgl. Abb. 1).

Die erweiterte Funktionalität der Berechnung der Adjungierten wird durch eine Erweiterung des Löser `simpleFoam` namens `adjointSimpleFoam` realisiert. Dieser intern entwickelte Löser wurde dankenswerterweise von der Volkswagen AG zur Verfügung gestellt.

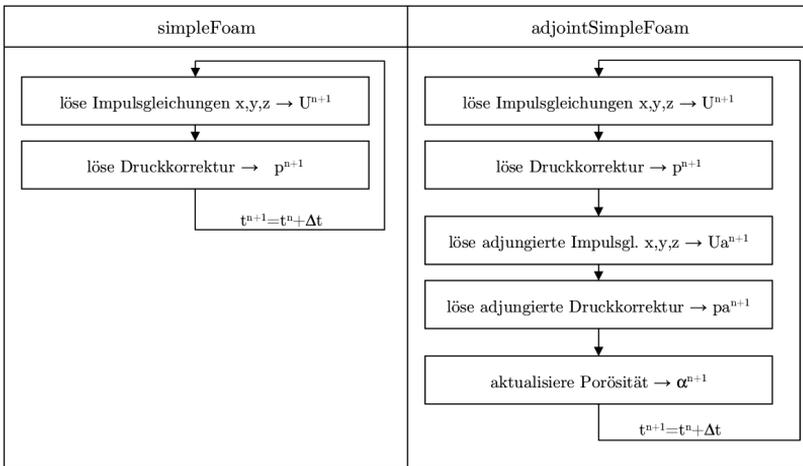


Abbildung 1: Ablaufdiagramm simpleFoam und adjointSimpleFoam

4 Validierung

Um die Qualität und das Verhalten des Optimierers beurteilen zu können wurde ein Vergleich mit veröffentlichten Ergebnissen[OOB06] durchgeführt. Betrachtet wird ein Kanal mit jeweils zwei Ein- und Auslässen. Je nach Einlassgeschwindigkeit können sich zwei verschiedene Lösungsformen herausbilden. Es entsteht entweder eine gerade oder eine gebogene Verbindung, wobei der Übergang zwischen beiden Lösungen bei einer Reynoldszahl von etwa $Re = 90$ auftritt. Bei den Rechnungen zeigte sich eine sehr gute Übereinstimmung (vgl. Abb. 2).

5 Ergebnisse Topologieoptimierung

Die Topologieoptimierung wird nun auf das reale Problem des Ölsaugkanals angewendet (s. Abb. 3). Es wird ein Ersatzbauraum definiert, welcher größer als das ursprüngliche Volumen ist um zusätzlichen Freiraum für die Optimierung zu schaffen. Anschließend wird der Optimierung ein Anteil dieses Bauraums vorgegeben welchen die optimierte Lösung nutzen darf. Im dargestellten Fall wurden nur 50% des ursprünglichen Volumens genutzt. Trotz der geringeren Volumennutzung konnte eine Druckeinsparung von knapp 25% erreicht werden.

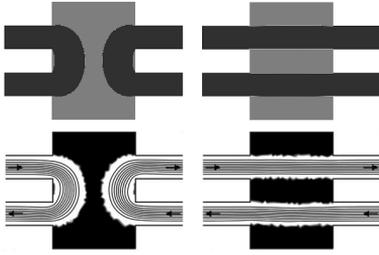


Abbildung 2: Vergleich zwischen mit `adjointSimpleFoam` erzielten (oben) und veröffentlichten Geometrien (unten). Links $Re = 20$, rechts $Re = 200$

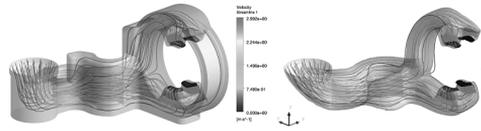


Abbildung 3: Ursprünglicher Kanal (links) und optimierte Geometrie (rechts)

Literatur

- [AG08] Volkswagen AG. DSG - die intelligente Automatik von Volkswagen. http://www.volkswagenag.com/content/vwcorp/info_center/de/themes/2008/01/the_7speed_dsg.html, 2008. Pressemitteilung, 21.01.2008. Abgerufen am 10.02.2012.
- [OOB06] Laurits Hjøgaard Olesen, Fridolin Okkels und Henrik Bruus. A high-level programming-language implementation of topology optimization applied to steady-state Navier-Stokes flow. *International Journal for Numerical Methods in Engineering*, 65(7):975–1001, 2006.
- [Oth08] C. Othmer. A continuous adjoint formulation for the computation of topological and surface sensitivities of ducted flows. *International Journal for Numerical Methods in Fluids*, 58(8):861–877, 2008.

Durchführung einer Usability-Evaluation am Beispiel des Warenwirtschaftssystems "CounTable"

Alexander Gottschall

alexander.gottschall@inf.hs-anhalt.de

Abstract: Die benutzerfreundliche Gestaltung von Anwendungssystemen stellt heutzutage einen nicht zu unterschätzenden Aspekt der Anwendungsentwicklung dar. Um die Benutzerfreundlichkeit einer Software zu ermitteln existieren verschiedene Evaluationsverfahren. Der nachfolgende Text dokumentiert die Durchführung eines solchen Verfahrens, einer Usability-Evaluation, in allen Schritten von der Problemanalyse über die Auswahl der Evaluationsmethoden hin zur Vorbereitung und Durchführung sowie der ausführlichen Darlegung der ermittelten Ergebnisse.

1 Einleitung

Die Disziplin der Mensch-Computer-Interaktion beschäftigt sich mit der benutzergerechten Gestaltung von interaktiven Systemen. Durch die zunehmende Komplexität von Anwendungen, verbunden mit den wachsenden Ansprüchen des Anwenders, gewann diese Thematik in den letzten Jahren stark an Bedeutung, auch wenn sie von vielen Anwendungsentwicklern nach wie vor unterschätzt wird [CHC01]. Dabei wird nicht nur auf visuelle Schnittstellen eingegangen, sondern auch auf umliegende Faktoren. So wird z.B. eine gut geschriebene Software keine gute Gebrauchstauglichkeit hervorbringen, wenn sie entgegen ihrem eigentlichen Verwendungszweck verwendet wird [SB06]. Auch stellt die MCI ihren Anwendern (Anwendungsentwickler, Webentwickler, Designer, Usability-Experten, etc.) diverse Werkzeuge zur Verfügung, welche helfen, ein System optimal zu gestalten. Dazu gehören das Wissen über psychologische Vorgänge, wie das Lernen oder die Wahrnehmung, und Prozessbeschreibungen, welche helfen, die Gebrauchstauglichkeit während des Entwicklungsprozesses - und auch danach - zu ermitteln und gegebenenfalls zu verbessern [NB95]. Eine formale Definition der Mensch-Computer-Interaktion liegt in Form der international gültigen Norm DIN EN ISO 9241 vor.

Im Rahmen der hier vorgestellten Bachelorarbeit wurde eine Usability-Evaluierung durchgeführt. Es wurde eine Software evaluiert, die durch den Autor und einen Kommilitonen für ein mittelständisches Catering-Unternehmen im Rahmen einer nebenstudentischen Tätigkeit entwickelt wurde. Bei der Anwendung handelt es sich um ein Warenwirtschaftssystem mit Namen "CounTable", welches für einen Großteil der innerbetrieblichen Datenverarbeitungsvorgänge genutzt werden sollte. Während der Entwicklung wurden die Möglichkeiten des Usability-Engineering jedoch nicht wahrgenommen, sodass mit einem vermehrten Auftreten von Usability-Problemen zu rechnen war. Dies war insbesondere

problematisch, da die Zielgruppe hauptsächlich aus Köchen bestand, welche nur über begrenzte PC-Kenntnisse verfügten. Hinzu kam, dass die Repräsentanten der Zielgruppe aus terminlichen und logistischen Gründen nicht für eine Evaluation zur Verfügung standen. Diese Aspekte mussten für das weitere Vorgehen berücksichtigt werden. Die Zielstellung gab vor, dass durch die Evaluation herausgefunden werden sollte, wo die Hürden der selbstständigen Einarbeitung in die Anwendung für den Anwender liegen.

Um eine fundierte Auswahl der Evaluationsmethoden treffen zu können, wurden zunächst die fünf etabliertesten ermittelt und beschrieben [MC02]. Die Evaluation mit Hilfe von **Gestaltungsrichtlinien** bedient sich einer Menge von Richtlinien und Prinzipien, welche, wenn eingehalten, zu einer hohen Gebrauchstauglichkeit führen sollen. Bei der **Heuristischen Evaluation** untersuchen Usability-Experten das zu evaluierende System anhand von Heuristiken. Diese repräsentieren Eigenschaften, welche ein interaktives System bei der Interaktion mit dem Nutzer aufweisen soll. Während des **Cognitive Walkthroughs** versetzen sich Usability-Experten in einen hypothetischen Anwender und analysieren jeden konkreten Handlungsschritt. Bei **Usability-Tests** werden Vertreter der Zielgruppe angewiesen, vorgegebene Aufgaben zu erfüllen. Ihr Vorgehen wird dabei dokumentiert und anschließend analysiert. Für dieses Verfahren wird ein Usability-Labor benötigt. **Fragebögen** dienen der subjektiven Meinungserfassung. Dabei beantworten Test-Probanden vorgefertigte Fragen meist im Anschluss an ein anderes Evaluationsverfahren. Die Wahl der Evaluationsmethoden fiel auf eine Kombination aus Usability-Test und Fragebogen. Diese Methode versprach die aussagekräftigsten Ergebnisse [Nie93]. Da jedoch nicht auf tatsächliche Repräsentanten der Zielgruppe zurückgegriffen werden konnte, musste die Evaluation mit Test-Probanden aus dem studentischen Umfeld durchgeführt werden.

2 Vorbereitung und Durchführung

Ein Usability-Test bedarf einer umfangreichen Vorbereitung. Zunächst wurden die Test-szenarien ausgearbeitet. Sechs Aufgaben wurden erstellt, welche den Großteil der täglich anfallenden Aufgaben abdeckten. Bei deren Formulierung wurde darauf geachtet, dass sie auch für Probanden verständlich waren, welche nicht mit der Firmenterminologie und den internen Arbeitsabläufen vertraut waren. Zusätzlich wurde ein Glossar zur Verwendung während der Evaluation erstellt. Das für die Evaluation benötigte Labor wurde von der Hochschule Anhalt zur Verfügung gestellt und musste ebenfalls präpariert werden. Insgesamt fünf Kameras wurden so ausgerichtet, dass die Probanden optimal während der Testphase beobachtet werden konnten. Die zu testende Anwendung wurde auf dem Test-Rechner installiert und die Beobachtungssoftware so konfiguriert, dass sie alle relevanten Aktionen dokumentierte und an einen Client im angrenzenden Beobachtungsraum streamte. Es wurden zwei verschiedene Fragebögen vorbereitet. Einer zur individuellen Bewertung der Aufgaben und ein abschließender, mittels welchem der subjektive Eindruck über das gesamte Programm erfasst werden sollte. Die Akquise der Test-Probanden erfolgte im studentischen Umfeld und war nach zwei Wochen abgeschlossen. Insgesamt erklärten sich sieben Personen bereit, als Test-Kandidaten zu fungieren.

Die Evaluationen wurden jeweils in drei Schritten durchgeführt. Der erste Teil, die Vor-

besprechung, war nötig, um den Probanden zu vermitteln, wie die Evaluation ablaufen würde, was von ihnen verlangt wurde und welche Hilfestellungen sie zu erwarten hatten. Zusätzlich wurden einige notwendige Informationen vermittelt, welche für das Verständnis der zu evaluierenden Software hilfreich waren. Dies ist von besonderer Bedeutung, da die Probanden nicht aus dem betrieblichen Umfeld der Zielgruppe kamen. Der zweite Teil stellte den eigentlichen Evaluationsteil des Tests dar, welchen die Probanden, unter Anleitung eines Testleiters, zu absolvieren hatten. Die Testkandidaten führten nacheinander die ihnen vorgegebenen Aufgaben aus und bewerteten diese anschließend mit Hilfe der Fragebögen. Während der Durchführung wurden sie gefilmt und vom Testleiter angehalten, ihre momentanen Gedankengänge zu äußern. Der Evaluationsleiter im angrenzenden Beobachtungsraum überwachte die Kameras und fertigte Notizen an. Abschließend galt es im dritten Schritt noch, eine Nachbesprechung durchzuführen, um auf interessante Aspekte des Tests näher einzugehen. Die Nachbesprechung ist mitunter von entscheidender Bedeutung. Sowohl die Probanden als auch die Usability-Experten verfügen über noch frische Eindrücke der gerade durchgeführten Evaluation.

3 Auswertung

In der Auswertung fließen alle ermittelten Informationen zusammen. Aus den aufgezeichneten Audio- und Videodaten, den Gesprächen mit den Probanden und den angefertigten Notizen wurde eine Liste mit insgesamt 41 Usability-Problemen erstellt. Neun wurden als „bedeutend“ eingestuft und sollten unmittelbar behoben werden. Vier mussten gar als „Usability-Katastrophe“ kategorisiert werden, weshalb ihnen bei der Behebung die höchste Priorität zuzuordnen ist. Zwar scheinen diese Usability-Probleme auf den ersten Blick hin nicht sehr bedeutend zu sein, doch kann deren Ignorierung schwerwiegende Folgen nach sich ziehen. Für jedes dieser Probleme wurde eine Übersicht der folgenden Daten angefertigt: Ort/Auftreten, Testaufgabe, Problembeschreibung, mögliche Ursachen, Gewichtung, Begründung der Gewichtung, Screenshot und Lösungsvorschlag.

Durch die Auswertung der Fragebögen sowie der sensorischen Daten (z.B. durchschnittlich vergangene Zeit zwischen zwei Eingaben) konnte trotz der ermittelten 41 Usability-Probleme ein positives Feedback der Anwender hinsichtlich der Anwendung ermittelt werden. Zudem konnte gezeigt werden, dass sich zukünftige Anwender sehr schnell in die Handhabung und Funktionsweise des Programms einarbeiten können, sofern sie mit den internen Arbeitsabläufen des Unternehmens und der verwendeten Firmenterminologie vertraut sind. Trotzdem wird empfohlen, eine einführende Schulung anzubieten, welche dem Anwender genau vermittelt, wie die Anwendung zu handhaben ist. Zwar konnten fast alle Probanden die Aufgabenstellungen vollständig erfüllen, jedoch waren sie sich nie hundertprozentig sicher, ob sie alles richtig gemacht haben. Die Ergebnisse der im Rahmen dieser Arbeit durchgeführten Evaluation zeigen, dass die Nichtbeachtung von Usability-Kriterien während der Entwicklung im Nachhinein einen nicht zu unterschätzenden Mehraufwand nach sich ziehen kann. Die Einbeziehung von tatsächlichen Repräsentanten der anvisierten Zielgruppe während des Entwicklungsprozesses kann die Menge der nachträglich notwendigen Korrekturen erheblich reduzieren.

Literatur

- [CHC01] M. Czerwinski, E. Horvitz und E. Cutrel. Subjective Duration Assessment: A New Metric for HCI? [Website], 2001. <http://research.microsoft.com/en-us/um/people/cutrell/ihtmhcisda.pdf>. Zugriff am 28.07.2011.
- [MC02] Rosson M. und J. Carrol. *Usability Engineering*. Academic Press, San Diego, 2002.
- [NB95] J. Nielsen und G. P. Bergman. Independent Iterative Design: A Method that didn't Work. *Advances in Human Computer Interaction*, 1995.
- [Nie93] J. Nielsen. *Usability-Engineering*. AP Professional Ltd., London, 1993.
- [SB06] F. Sarodnick und H. Brau. *Methoden der Usability Evaluation*. Verlag Hans Huber, Bern, 2006.

Low-Cost NUI

Malte Jannasch, Dennis Ludl, Claudia Vöhringer

Hochschule Reutlingen
Medien- und Kommunikationsinformatik Master
Malte_Simon.Jannasch@Student.Reutlingen-University.DE
Dennis.Ludl@Student.Reutlingen-University.DE
Claudia.Voehringer@Student.Reutlingen-University.DE

Art der Arbeit: Masterprojekt im ersten und zweiten Mastersemester
Betreuer/in der Arbeit: Prof. Dr. Gabriela Tullius, Prof. Dr. Uwe Kloos

Abstract: Im Virtual Reality Laboratory der Hochschule Reutlingen werden neue Interaktionsarten entwickelt und evaluiert. Natural User Interfaces (NUI) bilden dabei einen Schwerpunkt, der den Studierenden die Möglichkeit eröffnet, sich vertiefend mit neuen Interaktionsformen auseinanderzusetzen. Sowohl die entwickelten touchbasierten Systeme mit horizontaler und vertikaler Orientierung als auch die berührungslose Interaktion eröffnen vielfältige Anwendungsmöglichkeiten in unterschiedlichen Anwendungsfeldern. Dabei spielt die Usability dieser Interaktionsverfahren eine zentrale Rolle. Hierzu wurden Nutzertests entwickelt, die eine systematische Untersuchung der Einsetzbarkeit dieser NUI ermöglichen.

1 Einleitung

Im Virtual Reality Laboratory (VRLab) des Masterstudiengangs Medien- und Kommunikationsinformatik an der Hochschule Reutlingen forschen und entwickeln Studenten Lösungen im Themenumfeld virtueller Umgebungen und innovativer und natürlicher Interaktionsmöglichkeiten (NUI natural user interfaces) zwischen Mensch und Maschine. Weitere Schwerpunkte sind die stereoskopische Projektion von medialen und dreidimensionalen Inhalten sowie das Tracking zur Steuerung virtueller Welten. Dabei wird insbesondere Wert auf Low-Cost Ansätze und Lösungen gelegt [He11]. Zu den Natural User Interfaces (NUI) zählen Multitouch Systeme (MTS) sowie freihändig gestenbasierte Systeme, die auf Basis von bildverarbeitenden Algorithmen bestimmte Strukturen im Raum erkennen können. NUI werden meist als intuitive und vertraute Interaktionsformen beschrieben. Es wird oft übersehen, dass auch die gestenbasierte Interaktion zunächst erlernt werden muss und für die Gestaltung bestehende Richtlinien erweitert werden müssen [NN10]. Dazu leistet die vorliegende Arbeit einen Beitrag, indem drei Low-Cost NUI Systeme vorgestellt werden, mit denen repräsentative Daten zur Nutzung erhoben werden können.

2 Multitouch Systeme

Ein Schwerpunkt der MTS Arbeiten ist die Definition neuer Gesten oder die Entwicklung neuer Szenarien [Käl1]. Wenige Arbeiten beschäftigen sich mit der Evaluierung von MTS [So11]. Ziel der vorliegenden Arbeit ist die Untersuchung der Usability unterschiedlicher Low-Cost MTS. Dafür wurde innerhalb des VRlabs zwei MTS entwickelt: horizontal (Tisch) und vertikal (Cube), um unterschiedliche Auswirkungen der Bedienung testen zu können.

2.1 Technische Umsetzung der MTS

Die Umsetzung der Multitouch (MT) Funktionalität wurde unter Anwendung des Rear Diffused Illumination Ansatzes realisiert [Nu09]. Der Ansatz basiert auf der Low-Cost Technik wie von [Ha05] beschrieben. Dabei wird die Interaktionsfläche von der Rückseite mit Infrarot LEDs ausgeleuchtet. Eine Berührung führt zur Änderung des Reflexionsverhaltens der Scheibe. Dies kann mittels einer Kamera detektiert und lokalisiert werden. Durch Einsatz von Computer Vision Verfahren werden die Kameradaten in Trackingdaten umgerechnet und mit Hilfe des TUIO Protokolls [Ka05] an die verschiedensten MT Anwendungen weitergeleitet.



Abbildung 1: Multitouch Tisch

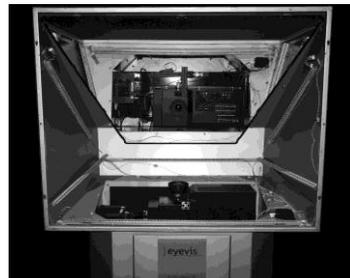


Abbildung 2: Multitouch Cube (Innen)

Der selbst konstruierte MT Tisch verfügt über eine horizontal ausgerichtete Bedienfläche, auf der das Bild eines LED Projektors über einen im Boden des Tisches angebrachten Spiegel dargestellt wird. Für die Touch Oberfläche wird halbtransparentes Acrylglas verwendet. Die Interaktionsfläche wird mit Hilfe einer Playstation 3 Eye Cam erfasst. Durch die Entwicklungsarbeit der Studenten kann der MT Tisch stabil betrieben werden (Abbildung 1). Der MT Cube basiert auf einem Rückprojektionscube der Firma eyevis GmbH, welcher dem VRlab zur Verfügung gestellt wurde. Aus dem als Ausgabegerät konzipierten Großbildsystem wurde von den Studierenden unter Verwendung von Low-Cost Komponenten ein weiteres MT Gerät, dieses Mal mit vertikaler Interaktionsfläche, konstruiert. Eine Herausforderung war hierbei, dass das Projektionssystem, das über einen schräg zur Projektionsfläche (1,1m x 1,38m) ausgerichteten Spiegel und einen DLP-Projektor realisiert wurde, nicht verändert werden konnte. Daher musste die Kamera aufgrund der Gehäusebeschaffenheit idealerweise an derselben Position wie der Projektor positioniert werden.

Diese und weitere technische Herausforderungen, wie die gleichmäßige Ausleuchtung, die Druckempfindlichkeit der großen Scheibe, konnten ebenfalls erfolgreich gelöst werden. Der innere Aufbau des MT Cube kann Abbildung 2 entnommen werden.

2.2 Nutzertests mit dem MT Cube

Zur Entwicklung von Gestaltungsrichtlinien für MTS wurde zunächst ein Nutzertest entwickelt und mit dem MT Cube durchgeführt. Bei der Interaktion mit dem MT Cube existieren eine Reihe von Unterschieden zu herkömmlichen MT Geräten. Insbesondere die Größe und die vertikale Ausrichtung der Benutzungsoberfläche fallen auf. Auf geeignete Bedienkonzepte für die Entwicklung einer benutzungsfreundlichen Anwendung kann auf Grund fehlender Erfahrungen mit dem Gerät nicht zurückgegriffen werden [NN10]. Faktoren wie die Ermüdung des Nutzers während der Interaktion, die Größe und Positionierung von aktiven und nicht aktiven Elementen sowie die teilweise Verdeckung der Inhalte durch den Nutzer müssen bei der Entwicklung von Anwendungen für den MT Cube zusätzlich beachtet werden, um eine möglichst gute User Experience zu erzielen. Für die Testanwendung wurde anhand des Überblicks über MT Frameworks von Kammer et al. [KFW10] MT4J ausgewählt. MT4J ist ein Java Framework zur Entwicklung von 2D und 3D MT Anwendungen. Für den Test wurden 18 Teilnehmer rekrutiert. Davon war die Mehrzahl Rechtshänder. Es wurden unterschiedliche Aufgaben zu Thesen (beispielsweise bzgl. der Positionierung von Interaktionsobjekten) hinsichtlich der Bedienung des Cubes gestellt bzw. bekannte Modelle wie Fitts Law überprüft. Gegenwärtig werden die Ergebnisse der Nutzertests ausgewertet. Erste Aussagen dieser Tests sind, dass die meisten Teilnehmer eine Navigationsleiste im unteren Bereich als negativ empfanden, obwohl sie ähnlich effizient wie an anderen Stellen (oben, rechts, links) bedient wurde. Interessant war auch die Beobachtung, dass fast alle Teilnehmer den Cube ausschließlich mit einer Hand bedienen, egal ob sich die Interaktionsobjekte im rechten oder linken Bereich von ihnen befanden. Die weitere statistische Auswertung der Test wird zur Entwicklung von Richtlinien für die Gestaltung von MT Cube Systemen führen. Darauf aufbauend können gestenbasierte Anwendungen entwickelt werden, welche dem Ideal der NUI näher kommen.

3 NUI auf Basis der Kinect

Mit der Einführung der Microsoft Kinect ergeben sich neue Möglichkeiten einer kostengünstigen Erfassung freihändiger Interaktionsformen. Dabei werden Körperbewegungen des Nutzers anhand eines Infrarot-Tiefensensors sowie einer RGB-Farbkamera erfasst und mittels Bildverarbeitungsverfahren selektiert [En11]. Grundlage der Erfassung sind einfache Bewegungen und Gesten. Für die Erkennung komplexer oder zusammenhängender Gesten mit der Kinect, die dann dem Anspruch NUI genügen können, existieren derzeit nur wenige Ansätze. Auf dieser Grundlage wird ein Ansatz verfolgt, Gesten unter Verwendung eines künstlichen neuronalen Netzes zu erkennen. Die Basis hierzu bildet ein Feedforward Netzwerk, das mit Hilfe des Backpropagation Algorithmus trainiert wird.

Für die Umsetzung des neuronalen Netzes wird das .NET Encog Framework 3.0 [Ha11] verwendet. Es wurde eine Software entwickelt die es ermöglicht Trainings- sowie Testdaten aufzunehmen und auszuwerten. Die Daten liegen hierbei zunächst als Liste von 3D-Koordinaten vor. Anschließend werden jeweils die Distanzen von x, y und z Koordinaten zweier benachbarter 3D-Punkte gemessen und in einem Vektor gespeichert, der als Eingabe für das neuronale Netz dient. Die Ausgabe sieht ein Neuron für jede Geste vor, die im Idealfall die Werte 0 für nicht erkannt bzw. 1 für erkannt annehmen. In einem ersten Test des neuronalen Netzes wurden 26 unterschiedliche komplexe und zusammenhängende Gesten verwendet. Die Buchstaben A-Z dienten hierbei als Vorlage. Die Erkennungsrate war in ersten Tests überwiegend gut. Gegenwärtig wird mit dem Ziel der Validierung des Verfahrens ein repräsentativer Nutzertest vorbereitet.

4 Ausblick

Die weiteren Arbeiten zielen auf die Formulierung von Gestaltungsempfehlungen für die Entwicklung von Anwendungen für NUIs ab, so dass die Interaktion nicht nur natural heißt, sondern auch natürlich empfunden wird. Weiterhin wird an Konzepten gearbeitet, die einen gemeinsamen Einsatz von vertikalen und horizontalen Multitouch Systemen ermöglichen, um so die Vorteile beide Interaktionsformen zu kombinieren.

Literaturverzeichnis

- [En11] Engineering and Technology Magazine: The Teardown: The Kinect for Xbox 360. Engineering & Technology, Vol.6, Nr.3, 2011; S.94-95.
- [Ha11] Haeton Research, Inc.: Encog Java and DotNet Neural Network Framework - <http://www.heatonresearch.com/encog> Stand 12.11.2011.
- [Ha05] Han, J.Y.: Low-cost multi-touch sensing through frustrated total internal reflection. In UIST '05: Proceedings of the 18th annual ACM symposium on User interface software and technology (New York, NY, USA, 2005), ACM, pp. 115–118.
- [He11] Herrmann, E. et al.: Low Cost Tracking: A student's approach to a tracking system. Accepted Poster at 2011, Joint Virtual Reality Conference 20-21 September, Nottingham UK EuroVR-EGVE, 2011.
- [Ka05] Kaltenbrunner, M. et al.: TUIO: A protocol for table-top tangible user interfaces. In: Proc. of the 6th International Workshop on Gesture in Human-Computer Interaction and Simulation, GW, 2005.
- [Kä11] Käser, D. et al.: FingerGlass: Efficient Multiscale Interaction on Multitouch Screens. In: ACM Human Factors in Computing Systems (CHI), 2011. pp. 1601-1610.
- [KFW10] Kammer, D.; Freitag, G.; Wacker, M.: Taxonomy and Overview of Multi-touch Frameworks: Architecture, Scope and Features. In: Symposium on Engineering Interactive Computing Systems; Workshop; Engineering Patterns for Multi-Touch Interfaces, ACM, Berlin, 2010.
- [NN10] Norman, D. A.; Nielsen, J.: Gestural Interfaces: A Step Backwards In Usability. In: Interactions, Vol.17, Nr.5, ACM, New York, NY, USA, 2010; pp .46-49.
- [Nu09] NUI Group Authors: Multitouch Technologies, NUI Group, Community Release, 2009.
- [So11] Soro, A. et al.: Evaluation of User Gestures in Multi-touch Interaction: a Case Study in Pair-programming. In Proceedings of ICMI 2011, Alicante, Spain: pp. 161-168

Entwicklung eines iPhone Uni-Navigators

Stefanie Müller, Joan Tomás Pape

Universität Hamburg
Mensch-Computer-Interaktion
stefanie-mueller86@gmx.de
joanpape@gmx.de

Art der Arbeit: Bachelorarbeit

Betreuer der Arbeit: Prof. Dr. Horst Oberquelle

Abstract: Im Zuge dieser Arbeit, wurde eine iPhone-Anwendung („App“) mit dem Namen „Uni-Navigator“ entwickelt, welche sich in zwei Teile gliedert. Im ersten Teil ermöglicht diese eine benutzerfreundliche Berechnung von Routen zu Universitätsgebäuden, wobei als Transportmittel sowohl Auto als auch öffentliche Verkehrsmittel angegeben werden können. Im zweiten Teil der App wird eine Funktionalität geboten, bei der Informationen zu mit dem Smartphone anvisierten Universitätsgebäuden angezeigt werden können, um dadurch das Zurechtfinden auf dem Campus zu erleichtern.

1 Motivation

Die Gebäude der Universität Hamburg sind über die ganze Stadt verteilt. Deshalb verbringen Studierende des Fachbereichs Informatik, vor allem zu Beginn des Studiums, viel Zeit damit, zwischen den Teilen der Universität umherzufahren und herauszufinden, in welchen Gebäuden die Veranstaltungen stattfinden. Da die Zeit zwischen zwei Veranstaltungen ohnehin knapp bemessen ist, sollte vermieden werden, dass die Fahrzeit zusätzlich durch suboptimale Routen verlängert wird. Zwar ist es Smartphone-Besitzern möglich, verschiedene Webseiten für deren Routenplanung und Informationsbeschaffung zu benutzen, jedoch ist dies auf einem Smartphone häufig ein zeitintensives und umständliches Unterfangen.

Anknüpfend an diese Problematik wurde der *iPhone Uni-Navigator* entwickelt. Hierbei handelt es sich um eine Applikation, welche zwei grundlegende Funktionalitäten bietet. Das erste Modul, genannt „Navi“, bietet die Möglichkeit einer mobilen Navigation zu Orten, welche zur Universität Hamburg gehören. Demgegenüber kann das zweite Modul mit dem iPhone anvisierte Gebäude der Universität Hamburg erkennen und Informationen zu diesen ausgeben. Dieses Modul wird im Folgenden mit „Scan“ bezeichnet. Der in dieser Arbeit entwickelte Prototyp umfasst nur diejenigen Gebäude, Orte und Informationen, die für Studierende der Informatik relevant sind.

2 Das Navi

Nachdem das Navi (s. Abb. 1) aus dem Hauptmenü der Anwendung heraus gestartet wurde, wird zunächst automatisch die aktuelle geographische Position des Geräts über den internen Location Manager festgestellt [BA10]. Anschließend werden diese Daten mittels Reverse Geocoding an Google Maps gesendet und dort in Adressdaten umgewandelt, welche dann automatisch als Startpunkt der Route eingetragen werden. Es ist dem Benutzer aber auch möglich, eine andere Adresse als Startpunkt einzugeben. Zum Verifizieren dieser wird die Eingabe wiederum an den Geocoding Webservice von Google Maps gesendet, um dort in eine geographische Position umgerechnet zu werden. Gelingt dies, wird die Eingabe des Benutzers akzeptiert und als Punkt auf einer Karte angezeigt. Der Benutzer kann die Adresse nun als Startadresse übernehmen und zum Hauptmenü der Navi-Komponente zurückkehren. Alternativ ist es auch möglich, als Startadresse einen der eingespeicherten Orte zu wählen. Mit einem Druck auf den Reset-Button wird wieder der Standort des Geräts als Startadresse ausgewählt.



Abbildung 1: Das Hauptmenü des Navi

Das Ziel der Route wird aus einer Tabelle ausgewählt, in der alle in der internen Datenbank eingespeicherten Orte angezeigt werden. Die Zielauswahl verfügt außerdem über eine Suchfunktion. Zum Vereinfachen der Suche wurden zu allen Orten zusätzlich Tags eingespeichert, welche beim Suchen mit einbezogen werden. So wird das Informatikum nicht nur über die direkte Bezeichnung, sondern auch über „Ikum“ oder „Haus F“ gefunden.

2.1 Suche über Google Maps

Wenn der Benutzer im Navi-Interface als Transportmittel „Auto“ ausgewählt hat, wird die Route über Google Maps berechnet. Dafür wird die Start- und Zieladresse an den Directions Webservice von Google Maps gesendet, woraufhin dieser die Route in Form eines XML-Dokuments zurückgibt. In diesem Dokument wird die Route in sogenannte *Legs* und *Steps* gegliedert [Goo11]. Da *Legs* nur bei Routen über zu passierende Zwischenpunkte auf dem Weg relevant sind, liegt der Fokus hier auf den *Steps*. Diese bilden die einzelnen Schritte der Route ab. Jeder *Step* hat einen Anfangs- und einen Endpunkt, welche über geographische Koordinaten festgelegt sind. Außerdem steht in den *Steps* deren Beschreibung in HTML-Text, sowie die Dauer und die in diesem Schritt zurückgelegte Entfernung. Nachdem das XML-Dokument „geparsed“ und die Beschreibungen von den HTML-Tags bereinigt wurden, wird die Route in einer Tabelle ausgegeben. Zusätzlich werden die einzelnen Schritte in Form von Punkten auf einer Karte angezeigt.

2.2 Suche über Geofox

Sollte der Benutzer für die Routen-Ermittlung öffentliche Verkehrsmittel nutzen, findet die Routenanfrage über die Webservices von Geofox.de statt. Geofox ist ein System des Hamburger Berater Teams, welches von der Hamburger Hochbahn AG eingesetzt wird, um Verbindungsempfehlungen innerhalb von Hamburg zu berechnen [HBT11]. Da dieses allerdings nicht frei verfügbar ist und eine Nutzung nur für den Rahmen dieser Arbeit gestattet wurde, wirkt der Webserver des Fachbereichs Informatik als Proxy zwischen dem iPhone- und dem Geofox-Server. So kann die IP-Adresse des Servers für Geofox freigeschaltet werden.

Die Requests an Geofox verwenden das SOAP-Protokoll. Demnach ist hierbei nicht nur die Antwort auf die Anfrage, sondern auch die Anfrage selbst ein XML-Dokument. Die Erstellung eines XML-Dokuments wurde mit Hilfe von Klassen realisiert, welche mittels der Webseite *www.sudzc.com* erstellt wurden. Die Route wird auf dem Smartphone wie die der Google-Anfrage in Form einer Tabelle aus Schritten angegeben. Weiterhin ermöglicht es Geofox, die geographischen Koordinaten der Haltestellen zu berechnen. So ist es möglich, die Fußwege von und zu den Haltestellen zu berechnen. Dies geschieht erneut über Google Maps und wird in derselben Form wie die Autorouten dargestellt.

3 Der Scan

Die zweite Komponente ist der Gebäude-Scan. Hierbei wird das Smartphone ähnlich wie ein Kompass in Richtung eines der Universitätsgebäude gerichtet, woraufhin der Name dieses Gebäudes oder dieser Einrichtung angezeigt wird. Der Benutzer hat dann die Möglichkeit, sich Informationen zu dem Gebäude anzeigen zu lassen.

Damit die Gebäude erkannt werden können, wurden in der internen Datenbank alle Gebäude zusammen mit ihren geographischen Positionen eingespeichert. Außerdem ist jedes Gebäude einem Ort zugewiesen. Zum Beispiel steht das Gebäude „Audimax“ auf dem Ort „Hauptcampus“. Das Erkennen der Gebäude erfolgt in drei Schritten und wird jedes Mal erneut ausgeführt, wenn sich der Winkel des digitalen Kompasses des iPhones um mindestens 10° verändert oder sich die Position des Smartphones um mindestens 1 Meter verändert.

1. Zunächst wird über den Location Manager des iPhones die geographische Position des Geräts festgestellt. Diese wird dann verwendet, um alle Orte auszuschließen, die sich nicht in der näheren Umgebung befinden. Die Gebäude, die zu den gefundenen Orten gehören, werden dann in ein Array gespeichert.
2. Im zweiten Schritt werden alle Gebäude aus dem Array entfernt, die mehr als 40 Meter von der Position des Smartphones entfernt sind.

3. Abschließend werden die Gebäude aus dem Array entfernt, bei denen die Richtung zwischen dem iPhone und den Koordinaten des Gebäudes nicht der Richtung entspricht, in die das iPhone gerade zeigt. Dabei wird eine Toleranz von $+ / - 40^\circ$ akzeptiert.

Das Array wird nun in einer Tabelle aufgelistet. Hier kann der Benutzer auf die Einträge tippen, um Informationen zu diesen zu erhalten. Alternativ ist es möglich, sich die eingespeicherten Bilder aller Gebäude des aktuellen Ortes anzeigen zu lassen. Die Informationen können dann über das Antippen der Bilder angezeigt werden.

4 Datenbank

Ein weiteres Feature der App ist die Datenbank, in welcher der Benutzer direkt nach den eingespeicherten Informationen suchen kann. Hierbei kann man die Suche auf Bilder, Orte, Gebäude oder Lehrende beschränken. Das Antippen eines Eintrags führt zu einer Detailansicht (s. Abb. 2). Abhängig vom Typ des Eintrags ist diese unterschiedlich gegliedert, da sie verschiedene Informationen anzeigt.

5 Auswertung

Der Uni-Navigator könnte bezüglich des Entwicklungsstands für Studierende der Informatik an der Universität Hamburg bereits eingesetzt werden. Die Genauigkeit des Scan ist jedoch verbesserungswürdig und ein Interface zur Eingabe der Datenbankdaten würde die Erweiterbarkeit der App unterstützen. Durch das Einhalten der im Vorfeld festgelegten Design-Richtlinien bietet die Applikation ein nahezu konsistentes Interface. Die Verbreitung der App ist jedoch nicht möglich, da die Nutzung des Geofox-Servers nur für diese Bachelorarbeit legitimiert wurde. Auch wäre es notwendig, eine Genehmigung für die Verwendung von Google Maps zu beantragen.

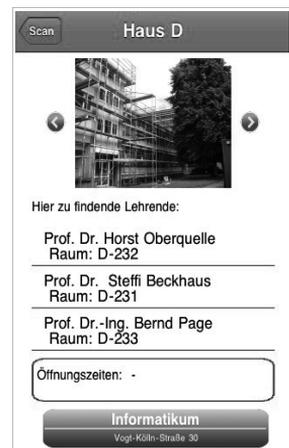


Abbildung 2: Die Detailansicht

Literaturverzeichnis

- [BA10] Dudney, Bill, Chris Adamson. Entwickeln mit dem iPhone SDK. Köln. O'Reilly Verlag. 2010.1.Auflage. Seite 488
- [HBT11] HBT Hamburger Berater Team GmbH. Geofox: HBT Hamburger Berater Team GmbH <http://www.hbt.de/geofox.html> (zuletzt besucht am 19.12.2011)
- [Goo11] Google. (2011). Google Directions API – Google Maps API-Webdienste <http://code.google.com/intl/de/apis/maps/documentation/directions/> (zuletzt besucht am 03.01.2012)

A Pattern Evolution Process - From Ideas to Patterns

René Reiners

rene.reiners@fit.fraunhofer.de

Abstract: Current pattern languages are derived after long experience by an expert group of pattern authors. This may lead to a decelerated reuse of available design knowledge. In this work, we propose an early involvement of many stakeholders within a project to early contribute their ideas as pattern candidates. Over time, these ideas are developed towards validated patterns or refined in future iterations. Our approach intends to integrate available application design knowledge directly from the beginning of the development process.

1 Application Design Patterns

Design patterns are used to describe best practices and effective design solutions, and in order to capture and share design knowledge with other people facing similar problems in a related context. The solution proposed by a design pattern should be generic rather than specific, such that it can be implemented in numerous different ways. The benefit of using patterns is that they communicate insights into common design problems and reflect solutions that a community of experts has developed over time. A given pattern is not always the optimal solution in every case, but tends to work in practice and supports user acceptance for the system.

Design patterns originated as an architectural concept introduced in the seminal book "A Pattern Language" written by Christopher Alexander [Ale77]. Seeing how design patterns could also be useful in other contexts, the work of Alexander inspired other engineering disciplines to adopt the concept. In particular, Gamma et al. extended the notion of patterns into the domain of software engineering, and constructed twenty-three classic software design patterns (cf [GHJV94]). Since then, the concept of design patterns also became essential in the domain of human-computer-interaction (HCI), where patterns are commonly used to describe and preserve solutions to recurring UI design problems.

Tidwell describes the influence of patterns in user interface design. Here, patterns are described as "best-practice within a given design domain" for improving objects, like user interfaces [Tid05]. As Alexander and Tidwell mention in their work, each implementation of the same pattern differs somehow in its characteristics although it comes from the same origin. Patterns should be seen as starting points that describe problem solutions - and not as fixed design rules.

Borchers states that patterns should be readable by many people, also non-domain experts [Bor01]. From his point of view, especially patterns in HCI need to bridge the gap between

users with conceptual knowledge and understanding of the problem domain and software engineers who are deeply involved in the technical development.

A similar approach is introduced by Schümmer and Lukosch [SL07]. They, however, introduce more technical details within the structure of their pattern language meaning the more detailed a pattern describes a certain solution, the more technical this description becomes.

In addition to working solutions, the description of *anti-patterns* is also a valid information source for application and interface designers. They document surprisingly failing approaches that turn out to be ineffective or counter-productive in practice [RAZ11]. Other collections, e.g., in UI design, focus on pointing out repetitions of design flaws [Joh07]. Here, concepts that have intruded many designs but actually lead to rejection are discussed and the reasons for the design failures are explained.

2 The Design Pattern Evolution Process

Traditionally, pattern collections are published after a longer period of development time extracting well-studied essences from design experience. This is mostly done by a small, closed group of design experts as described in [RAZ11]. Besides the continuous derivation of patterns, we want to involve many stakeholders to participate in the pattern evolution process. Our aim is to document and develop application design knowledge from the very beginning of the development process and continuously develop the findings further. Thus, we define the pattern evolution process for a given pattern collection as follows:

Each newly submitted pattern undergoes a first semi-automated quality check process meaning that the system first ensures that all required pattern fields are actually filled. The submission is then internally forwarded to a group of pattern reviewers who decide whether the pattern can be published in the library or if certain formulations need to be changed again by the author. With this quality check, we want to avoid flooding the pattern library with incomplete contributions.

As soon as a pattern candidate is published in the pattern library, every registered user can provide feedback to the pattern or its formulation. It is also possible to support or refute the pattern statement by providing more references in favor of the pattern or against it. This way, the pattern maturity changes over time (see below).

To reflect the liveliness and bottom-up approach of the patterns in the design pattern library, two additional fields are added to the known pattern structure: *state* and *origin*. We introduce the notion of a pattern's state that is used to track the development of the pattern over time. Our current implementation provides the following fields:

- *just created* patterns were recently submitted as a non-validated idea.
 - *under consideration* means that the pattern looks promising but needs further validation.
 - *pattern candidate* states that the pattern is close to being approved.
 - *approved* finalizes the pattern review process and settles the pattern as a design pattern.
- However, this does not exclude the possibility to open the pattern for discussion in the future again due to new findings.

Currently, we have not yet defined a measure for the state of a pattern's maturity but consider the number of successful applications of a pattern as used by Grill et al. [GB08].

Already proven existing patterns from other collections that can be integrated into the project domain without any changes can directly reach the *approved* state. Adapted patterns and newly integrated ones must initially be put in the state *under consideration* and be validated as a pattern or rejected over time. A pattern idea that surprisingly fails during validation might turn out as being an anti-pattern which is also to be documented in order to avoid future repetitions of similar design flaws. Patterns can originate both from the project itself and from external sources. We distinguish three different categories:

(i) *Derived from project*: The pattern derives directly from the work within the project. Patterns in this category will automatically be assigned the state *under consideration*. The pattern will be reviewed, perhaps be re-worked and finally be validated through an approval process by a validator.

(ii) *Adapted to project*: The pattern originates from external sources, but has been adapted for use in the context of the project.

(iii) *External*: The pattern exists in other pattern collections (e.g. a standard UI pattern) and is implemented in the current project's products and services. Regardless of originating from the project or external pattern collections, all patterns must be justified by references in literature or other proofs of success such as applied working solutions in research, processes or products.

Since we want to involve a large group of pattern authors we need to keep the access to the design pattern library easy. Therefore, any visitor can browse the design patterns, and contribute and suggest new ones by contacting the library administrators. Registered users can directly submit patterns and comment on existing patterns as well as the library structure. After submission, a user automatically becomes the pattern responsible and therewith responsible for updating or refining the pattern in question. Note that the pattern responsible is not necessarily the originator of the pattern.

A user's reputation also changes over time; this happens by submitting more patterns and by reviewing other submissions. At a certain level, the user achieves an expert's rank and co-decides with others about structural changes within in library, pattern merges or pattern divisions. The influence on the pattern library evolution grows according to the user's reputation. In order to realize such a reputation system for contributors, we think about integrating the *CollabReview* platform developed by Prause [Pra11].

3 Results and Future Work

In this work, we described our concept of involving *many* stakeholders into the design pattern creation and evolution process and to adapt already available knowledge and findings from the project's domain as early as possible. The pattern-evolution process allows for continuously evaluating the design knowledge gathered during the project's lifetime and makes patterns as well as pattern ideas available during the whole development process. Successful findings or surprisingly failing results will be formulated as (anti-)patterns.

This way, pattern collections can be formulated collaboratively without needing to wait for a closed author group that shares its well-evaluated design knowledge after a longer period of time. The approach is currently being implemented in the BRIDGE Design Pattern Library¹. Currently, three *external patterns under consideration* from other collections are integrated, one pattern is *adapted* from another collection and four patterns were *derived from the project*. One of them, describing a standard color schema corresponding to different levels of danger, is *approved*. Our next steps will address evaluations on the users' behavior concerning contributing patterns and reviewing existing ones. In terms of growth and structural development we will monitor the pattern library evolution and assess the current set of process rules and pattern states to be integrated into upcoming versions.

Acknowledgements

This work was supported by the European Commission within the FP7-Security Programme (SEC-2010.4.2-1) project BRIDGE, grant agreement no. 261817.

References

- [Ale77] Christopher Alexander. *A Pattern Language: Towns, Buildings, Construction*. Oxford University Press, New York, New York, USA, 1977.
- [Bor01] Jan Borchers. *A Pattern Approach to Interaction Design*. John Wiley & Sons, 1st edition, 2001.
- [GB08] Thomas Grill and Margit Blauhut. Design Patterns Applied in a User Interface Design (UID) Process for Safety Critical Environments (SCEs). In Andreas Holzinger, editor, *HCI and Usability for Education and Work*, volume 5298 of *Lecture Notes in Computer Science*, pages 459–474. Springer Berlin / Heidelberg, 2008.
- [GHJV94] Erich Gamma, Richard Helm, Ralph E. Johnson, and John Vlissides. *Design Patterns. Elements of Reusable Object-Oriented Software*. Addison-Wesley Longman, Amsterdam, 1 edition, 1994.
- [Joh07] Jeff Johnson. GUI bloopers 2.0: Common User Interface Design Don'ts and DOS. *Ubiquity*, 2(October), 2007.
- [Pra11] Christian R Prause. Reputation-based self-management of software process artifact quality in consortium research projects. In *Proceedings of the 19th ACM SIGSOFT symposium and the 13th European conference on Foundations of software engineering, ESEC/FSE '11*, pages 380–383, New York, NY, USA, 2011. ACM.
- [RAZ11] René Reiners, Irina Astrova, and Alfred Zimmermann. Introducing new Pattern Language Concepts and an Extended Pattern Structure for Ubiquitous Computing Application Design Support. In Fritz Laux, René Reiners, and Alfred Zimmermann, editors, *PATTERNS 2011, Third International Conferences on Pervasive Patterns and Applications*, number c, pages 61–66, Rome, September 2011. XPS - Xpert Publishing Services.
- [SL07] Till; Schümmer and Stephan Lukosch. *Patterns for Computer-Mediated Interaction*. John Wiley & Sons, Chistester, West Sussex, England, 2007.
- [Tid05] Jenifer Tidwell. *Designing Interfaces*. O'Reilly Media, 1 edition, 2005.

¹<http://patterns.bridge-sec.eu>

Informationsqualität in Unternehmenswikis – Eine internationale empirische Studie zu Kriterien der Informationsqualität

Nadine Hellweg

Universität Paderborn
Didaktik der Informatik
Nadine-Hellweg@gmx.de

Art der Arbeit: Bachelorarbeit

Betreuer der Arbeit: Wolfgang Reinhardt

Abstract: Gemeinsame Projektarbeit und Echtzeit-Kommunikation wird zunehmend wichtiger für das Wissensmanagement in Unternehmen [GP09]. Deshalb werden Betriebswikis vermehrt in Unternehmensstrukturen integriert, denn sie stellen eine leicht zu bedienende Plattform für kooperatives Arbeiten an Texten dar. Eine Studie aus dem Jahr 2009 zeigt, welche Kriterien die Informationsqualität der Betriebswikis beeinflussen können (vgl. [F09], S. 60). Basierend auf einer empirischen Studie wurde gezeigt, dass die Kriterien Aktualität, Vollständigkeit, Glaubwürdigkeit, Relevanz und Verständlichkeit relevant sind. In einer weiteren Studie sollte nun die Beständigkeit des Wiki-Sets an international tätigen Großunternehmen überprüft werden. Für die direkte Vergleichbarkeit wurde der bereits bestehende Fragebogen der vorherigen Studie mit allen Fragen beibehalten. Die Indikatoren und Kriterien aus den beiden Studien wurden danach in Beziehung gebracht und miteinander verglichen. Im direkten Vergleich zeigte sich, dass Glaubwürdigkeit, Vollständigkeit und auch Verständlichkeit in ihrer Bedeutung für die Mitarbeiter an Wichtigkeit gewonnen haben. Die Bedeutung von Relevanz für die Bewertung der Informationsqualität blieb im Mittel nahezu gleich. Lediglich das Kriterium Aktualität verlor an Wichtigkeit für die bewertenden Personen.

1 Das Wiki-Set

Unternehmenswikis (auch Corporate oder Enterprise Wikis genannt) gibt es bereits seit Mitte der 1990er Jahre. Diese Wissensmanagementsysteme dienen seitdem dem kollaborativen Arbeiten und dem Teilen von Mitarbeiterwissen. Ähnlich wie bei Wikipedia können Mitarbeiter des Unternehmens ihr Wissen in Artikeln festhalten und es damit zum festen Bestandteil der Projektgruppe oder sogar des gesamten Unternehmens machen. Die Informationsqualität solcher Artikel ist daher von großer Bedeutung, weil hochwertigere Beiträge ein effizienteres Arbeiten mit dem Mitarbeiterwissen erleichtern.

Die Informationsqualität wurde in der Bachelorarbeit innerhalb einer empirischen Studie untersucht, indem fünf verschiedene Kriterien und dazugehörige Indikatoren von den Mitarbeitern international tätiger (Groß-) Unternehmen bewertet wurden. Dieses Wiki-Set besteht aus folgenden Kriterien:

- (1) Aktualität: tatsächliche Eigenschaften des beschriebenen Objektes werden zeitnah abgebildet
- (2) Vollständigkeit: Informationen dürfen nicht fehlen
- (3) Glaubwürdigkeit: Informationen weisen einen hohen Qualitätsstandard auf
- (4) Relevanz: notwendige Informationen werden für den Leser geliefert
- (5) Verständlichkeit: Informationen werden vom Leser verstanden und können eingesetzt werden

Ziele der Studie waren zum einen die Überprüfung, ob das Wiki-Set in den teilgenommenen Unternehmen weiterhin Bestand für die Informationsqualität hat. Zum anderen sollte untersucht werden, in wie weit die Ergebnisse konsistent zur vorherigen Studie waren. Bei Abweichung sollten mögliche Erklärungsansätze herausgearbeitet werden. Zuletzt war auch von Interesse, in welchem Ausmaß die Mitarbeiter einen Nutzen in der Qualitätsfeststellung von Artikeln in Unternehmenswikis sehen.

2 Auswertung der empirischen Studie

An der Studie haben insgesamt 91 Personen aus verschiedenen Unternehmen teilgenommen. Sowohl das derzeitige als auch das zukünftige Nutzungsverhalten der Teilnehmer ist positiv ausgefallen: mehr als die Hälfte nutzt ihr Wiki und veröffentlicht regelmäßig Beiträge. Fast alle Personen gaben Interesse an Kommentaren innerhalb einer Bewertung der Informationsqualität an. Auch die Bereitschaft zum Bewerten ist bei nahezu allen vorhanden. Als Auswertung für die Dauer einer Bewertung kann man fünf Klicks nennen, die innerhalb von ca. 30 Sekunden bearbeitet werden können. Im Folgenden werden die einzelnen Kriterien innerhalb der beiden Studien miteinander verglichen.

2.1 Aktualität

Tabelle 1 zeigt, dass das Kriterium Aktualität 'sehr wichtig' für die Bewertung der Informationsqualität ist. Im Vergleich zu 2009 verliert es 3,45% und fällt auf nunmehr 53,85%. Die Bewertungen 'wichtig' und 'eher unwichtig' bleiben relativ konstant. Nur die letzte Bedeutung 'ziemlich unwichtig' erhält 3,3% an Stimmen dazu. Man kann folglich sagen, dass es eine generelle Verschiebung der Bewertungen des Kriteriums Aktualität in Richtung weniger bedeutend gibt. Ein möglicher Grund für diese Tatsache ist, dass in großen Unternehmen auch Artikel existieren, die nicht an Aktualität verlieren. Auch im Hinblick auf international ausgerichtete Betriebe ist es so, dass viele Artikel Basisthemen beschreiben, die sich nur selten verändern. Daher ist die Aktualität bei diesen Beiträgen nicht von Bedeutung, da sie zeitlos sind und in Zukunft auch weiterhin ihre Gültigkeit besitzen.

Bedeutung	Häufigkeit 2009	Prozent 2009	Häufigkeit 2011	Prozent 2011
sehr wichtig	118	57,3	49	53,85
wichtig	71	34,5	31	34,07
eher unwichtig	17	8,3	8	8,79
zieml. unwichtig	0	0,0	3	3,30

Tabelle 1: Vergleich des Kriteriums Aktualität

2.2 Vollständigkeit

Beim Vergleich des Kriteriums Vollständigkeit lässt sich eine starke Veränderung zugunsten der Kategorie 'sehr wichtig' ausmachen (siehe Tabelle 2). Eine deutliche Verschiebung von 21,4% auf 34,07% stellt sich bei den Teilnehmern der großen und internationalen Unternehmen ein, was ein Plus von 12,67% bedeutet. Diese Tatsache lässt sich mit der Ursache begründen, dass international ansässige Unternehmen global agieren. D. h., sie treffen an Märkten auf mehr Konkurrenz als ein Unternehmen, welches primär deutschlandweit sein Geschäftsfeld betreibt. Internationale Unternehmen müssen einen sehr eingespielten Geschäftsablauf garantieren. Daher ist die Vollständigkeit von Wiki-Artikeln für diese Unternehmen besonders wichtig, um reaktionsschnell und fehlerfrei eine führende strategische Marktposition erhalten zu können.

Bedeutung	Häufigkeit 2009	Prozent 2009	Häufigkeit 2011	Prozent 2011
sehr wichtig	44	21,4	31	34,07
wichtig	97	47,1	44	48,35
eher unwichtig	63	30,6	16	17,58
zieml. unwichtig	2	1,0	0	0

Tabelle 2: Vergleich des Kriteriums Vollständigkeit

2.3 Glaubwürdigkeit

Die Glaubwürdigkeit eines Artikels ist für die befragten Personen neben der Verständlichkeit das wichtigste Kriterium innerhalb der neuen Umfrage. Besonders deutlich zu sehen ist die Steigerung der Kategorie 'sehr wichtig' um 9,34% (Tabelle 3). Auffällig ist die gleichzeitige Abnahme der Kategorie 'wichtig' um 9,94%. Die Verschiebung ist fast ausschließlich auf diese beiden genannten Werte bezogen. Die letzten beiden Bewertungen 'eher unwichtig' und 'ziemlich unwichtig' weisen kaum Veränderungen auf. Dieses Verschieben der Wertigkeit im direkten Vergleich zur Vorstudie zeigt, dass die Bedeutung des Kriteriums Glaubwürdigkeit für die Informationsqualität an Bedeutung zugenommen hat.

Bedeutung	Häufigkeit 2009	Prozent 2009	Häufigkeit 2011	Prozent 2011
sehr wichtig	103	50,0	54	59,34
wichtig	93	45,1	32	35,16
eher unwichtig	9	4,4	5	5,49
zieml. unwichtig	1	0,5	0	0

Tabelle 3: Vergleich des Kriteriums Glaubwürdigkeit

2.4 Relevanz

Die Bewertungen der Relevanz eines Artikels verteilen sich auf nahezu gleich auf die Kategorien 'sehr wichtig' mit 46,15% und 'wichtig' mit 45,05%. Hier ist ein deutlicher Unterschied zur Vorstudie zu sehen wie in Tabelle 4 dargestellt wird. Die höchste Bewertungsstufe erhält ein Plus von 2,95 Prozentpunkten, dafür sinkt die zweite Kategorie um 5,95%. Nur eine geringe Abweichung der letzten beiden Bewertungskategorien ist ersichtlich. Sie erhielten zusammengenommen insgesamt einen Zuwachs von 2,99%. Generell folgt, dass die Bewertung 'wichtig' deutlich seltener angekreuzt wurde und sich dafür auf alle an deren verteilt. Diese Verschiebung kann mit der eher subjektiv aufzufassenden Einstellung für die Relevanz als Bewertungskriterium für die Informationsqualität in Corporate Wikis erklärt werden.

Bedeutung	Häufigkeit 2009	Prozent 2009	Häufigkeit 2011	Prozent 2011
sehr wichtig	89	43,2	42	46,15
wichtig	105	51,0	41	45,05
eher unwichtig	12	5,8	7	7,69
zieml. unwichtig	0	0,0	1	1,10

Tabelle 4: Vergleich des Kriteriums Relevanz

2.5 Verständlichkeit

Auch beim Kriterium Verständlichkeit zeigt sich ein deutlicher Zugewinn in der Bewertung als 'sehr wichtig' für die Bewertung der Informationsqualität in Unternehmenswikis. Der Zugewinn kommt primär von der Kategorie 'wichtig', die dadurch 7,85% im Vergleich zur Vorstudie verliert. Die Bedeutung der Verständlichkeit (Tabelle 5) wird demnach zunehmend wichtiger. Ein Artikel muss gut lesbar geschrieben werden, damit man beim Lesen keine unnötige Zeit verliert. Sollte ein Text unverständlich sein, so kann er so gar als unbrauchbar betrachtet werden, obwohl wichtige Informationen womöglich enthalten sind.

Bedeutung	Häufigkeit 2009	Prozent 2009	Häufigkeit 2011	Prozent 2011
sehr wichtig	80	38,8	44	48,35
wichtig	118	57,3	45	49,45
eher unwichtig	8	3,9	2	2,20
ziemlich	0	0,0	0	0,0

Tabelle 5: Vergleich des Kriteriums Verständlichkeit

Literaturverzeichnis

- [F09] Friberg, T.: Bewertung der Informationsqualität in Unternehmenswikis – empirische Studie zu Kriterien der Informationsqualität, Diplomarbeit, Paderborn 2009
- [GP09] Gartner-Prognose <http://metaphorous.com/2009/11/04/gartner-wiki-und-die-starken-manager-siegen-bis-2013/>, abgerufen am 30.08.2011

Porting a Functional Query Language to the Parallel Data Flow Framework Stratosphere

Moritz Kaufmann, Matthias Ringwald

{moritz.kaufmann, matthias.ringwald}@campus.tu-berlin.de

Fachgebiet Datenbanksysteme und Informationsmanagement (Prof. Markl), TU-Berlin

Abstract: In recent years, Map/Reduce frameworks have become very popular for large-scale data processing. However, the definition of complex analytical tasks using the Map/Reduce programming model is often very cumbersome. To overcome this problem, higher-level languages have been designed that compile queries into Map/Reduce jobs. For the popular Map/Reduce framework Hadoop, a couple of higher-level languages emerged among them Pig, Hive, and JAQL. In our work we ported JAQL to the parallel execution framework Stratosphere. Stratosphere provides the parallel programming model PACT which is a generalization and extension of the Map/Reduce programming model. In contrast to Map/Reduce, PACT offers additional parallelization primitives and more flexible data flow compositions. We show that the additional capabilities can be successfully integrated in the runtime of JAQL. It was possible to port the language with nearly no loss of functionality so that the same queries can be executed independently of the underlying platform. With a set of example workloads we compare the performance of running the language on Stratosphere to running it on the Hadoop framework.

1 Introduction

In recent years parallel execution frameworks such as Map/Reduce and Dryad/LINQ gained a lot of attraction and popularity. This is fueled by the demand for web-scale data analytics which require processing terabytes to petabytes of data. A very influential framework is Map/Reduce which was published by Google [DG08] and is now adopted in several open-source implementations like Apache Hadoop. Its basic premise is that most workloads can be run in a massively parallel fashion using just the two primitives *map* and *reduce*.

Many parallel execution platforms provide high-level abstraction layers. These enable developers to write parallel programs in an easier and faster way by e.g. using a SQL-like syntax such as Hive or a functional language like JAQL.

The DFG funded collaborative research unit Stratosphere is conducting basic research in the area of parallel data analytics¹. Part of this effort is the PACT programming model [BEH⁺10]. It generalizes and extends the Map/Reduce programming model. Most important, additional second-order functions were introduced in order to extend expressiveness and enable better optimization capabilities. This also allows expressing many problems in a much easier way compared to Map/Reduce [AEH⁺11].

¹The Stratosphere system is open-source and available under <http://www.stratosphere.eu>

Related to the research on finding a programming model which allows efficient parallel execution is providing an abstraction layer to make it easily usable. One important contribution to the original Map/Reduce model is the JAQL [BEG⁺11] language which was developed by IBM and is also used in their Big Data Analytics products. It allows expressing queries in a functional way and automatically translates them into Map/Reduce jobs. The goal of this work is to provide such a layer for PACT too, by porting JAQL to it. The main reasons for this choice were its powerful syntax (turing-completeness [STL11]), its support for schema-full and schema-less data and increasing popularity of this language.

2 Realization

Originally, the JAQL rewriter was designed to convert JAQL queries into a sequence of Map/Reduce jobs which can be executed by the Apache Hadoop framework. More precisely the query is transformed in such a way, that it is separated into subqueries which can be executed as first-order functions inside the *map* and *reduce* functions using a JAQL interpreter provided by the runtime.

In order to port JAQL to Stratosphere, the following requirements were identified: integrating support for the additional second-order functions in the rewriter, handling the job submission, and a runtime environment to execute JAQL code inside the second-order functions of Stratosphere.

Rewriter The JAQL query language consists of a set of expressions that operate on data streams. Examples include *filter*, *group*, *join*, *sort* and *transform*. The rewriter converts declarative queries consisting of these expressions into first-order functions that serve as arguments for the second-order functions *map* and *reduce*. Thereby, the task of parallelization is handled by the Map/Reduce framework.

Internally, the rewriter works on a tree representation of the query. This tree is transformed in consecutive compilation phases, where each phase contains rules that eliminate or simplify expressions or restructure parts of the program. The rewriting rules include general optimizations such as predicate push down as well as more specific tasks like generation of the first-order functions for Hadoop (ToMapReduce). To enable the JAQL rewriter to handle multiple inputs, we defined a generic function expression for the second-order functions *match*, *cross* and *cogroup* which are available in PACT. Furthermore, we implemented a rewriter rule that generates a corresponding expression from the JAQL join expressions, which have been translated into reduce-side joins originally. Lastly, the ToMapReduce rule was redesigned to be aware of all second-order functions of PACT.

Job Submission A limitation of the Map/Reduce programming model is that jobs can only consist of a single *map* and a (optional) *reduce* step. Therefore, complex JAQL queries must be compiled into a sequence of Map/Reduce jobs that are consecutively submitted to the Hadoop framework for execution. Thereby, a job reads the output of its predecessor as input. In order to enable optimizations across multiple processing steps,

Stratosphere uses a data flow graph representation which allows to compile even complex queries into a single PACT job. JAQLs original behavior was changed, so that subqueries are not issued instantly, but are queued until their results are required. Prior to submission, the queued subqueries are connected based on their inputs and outputs and only one job containing the whole data flow is issued against the Stratosphere execution framework.

Runtime For each second-order function a generic first-order function was developed which can be customized by a configuration. The JAQL code that should be run inside a task is serialized into a text representation and transferred inside the configuration. Prior to execution, the user-function instantiates the JAQL interpreter and loads the configuration.

3 Performance evaluation

The performance evaluation was done with data and queries from the TPC-H benchmark [Cou11]. Three queries were translated from the SQL syntax to the JAQL syntax and the data was transformed into the JSON format. During this process, 1-to-N relations were included as nested data to be able to test JAQLs nested data feature resulting in 26 GB of data. We ran the queries on both execution frameworks, Hadoop (v0.20.2) and Stratosphere (pre-release). The nodes were equipped with a single-core CPU, 2 GB RAM and Linux. A JVM heap size of 1 GB was configured. For the tests with Hadoop we used one mapper and one reducer per node. The degree of parallelism for Stratosphere was set according to the number of nodes used which corresponds to one subtask per node.

Figure 1 shows the runtimes of three queries for an increasing number of processing nodes. Both approaches show the same scaling behavior. The Hadoop runtimes beat Stratosphere’s in two out of three queries. However, the gap between both frameworks is more narrow for degrees for parallelism >2 . In contrast to the queries 5 and 7, the query 3 which runs faster on Stratosphere does not include any joins (because of the prejoined 1-to-N relations). Both, Hadoop and Stratosphere internally use the same JAQL interpreter and the performance difference exists also for the same internal strategies. This indicates weaknesses of the sort and join implementation in the benchmarked pre-release version of Stratosphere. However, this is not a conceptual problem of Stratosphere and fixed by optimizations and implementation improvements in the current releases.

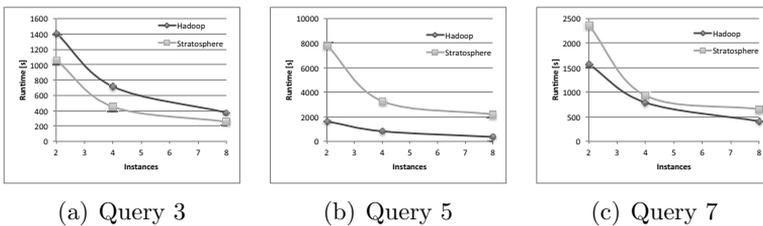


Figure 1: Results of the performance evaluation

4 Conclusion & Future Work

In conclusion, the project to port JAQL to the Stratosphere stack was successful. We implemented an interface for Stratosphere, which enables users to easily switch between frameworks without the need to change their JAQL queries. This lays an important foundation for interoperability and comparability of these frameworks. We successfully utilized the additional second-order functions which offer a higher flexibility in the execution strategy. Therefore, optimizations are possible, also during runtime and thereby improved response times.

Stratosphere can support all features of JAQL - due to the limited timeframe of the project, support for operators with more than two inputs and outer joins are not yet implemented. Future work also includes extraction of semantics and annotations of tasks with output contracts [BEH⁺10]. Output contracts enable the Stratosphere optimizer to choose more efficient execution plans.

The number of high-level abstractions for Hadoop like JAQL, Pig, and Hive show that there is demand for such abstractions and future efforts in this area for the Stratosphere stack seem advisable.

We thank our mentors Stephan Ewen, Fabian Hüske, and Daniel Warneke for their support, as well as the remaining project team for their contributions to the project.

References

- [AEH⁺11] Alexander Alexandrov, Stephan Ewen, Max Heimel, Fabian Hueske, Odej Kao, Volker Markl, Erik Nijkamp, and Daniel Warneke. MapReduce and PACT - Comparing Data Parallel Programming Models. In *Proceedings of the 14th Conference on Database Systems for Business, Technology, and Web (BTW)*, BTW 2011, pages 25–44, Bonn, Germany, 2011. GI.
- [BEG⁺11] Kevin S. Beyer, Vuk Ercegovic, Rainer Gemulla, Andrey Balmin, Mohamed Y. Eltabakh, Carl-Christian Kanne, Fatma zcan, and Eugene J. Shekita. Jaql: A Scripting Language for Large Scale Semistructured Data Analysis. *PVLDB*, pages 1272–1283, 2011.
- [BEH⁺10] Dominic Battré, Stephan Ewen, Fabian Hueske, Odej Kao, Volker Markl, and Daniel Warneke. Nephel/PACTs: a programming model and execution framework for web-scale analytical processing. In *Proceedings of the 1st ACM symposium on Cloud computing*, SoCC '10, pages 119–130, New York, NY, USA, 2010. ACM.
- [Cou11] Transaction Processing Performance Council. TPC Benchmark H. <http://www.tpc.org/tpch/>, February 2011.
- [DG08] Jeffrey Dean and Sanjay Ghemawat. MapReduce: simplified data processing on large clusters. *Commun. ACM*, 51:107–113, January 2008.
- [STL11] Robert J. Stewart, Phil W. Trinder, and Hans-Wolfgang Loidl. Comparing high level mapreduce query languages. In *Proceedings of the 9th international conference on Advanced parallel processing technologies*, APPT'11, pages 58–72, Berlin, Heidelberg, 2011. Springer-Verlag.

Outlier Mining Based on Subspace Analysis in High Dimensional Databases

Yvonne Mülle

yvonne.mueller@student.kit.edu

Karlsruhe Institute of Technology (KIT)

published as Bachelor thesis

Advisers: Prof. Dr.-Ing. Klemens Böhm klemens.boehm@kit.edu
Dr. Emmanuel Müller emmanuel.mueller@kit.edu
Patricia Iglesias Sánchez patricia.iglesias@kit.edu

1 Motivation

Nowadays, a large amount of data can be collected with little effort and it can be cheaply stored. To extract new, potentially useful and unexpected knowledge from that much data, automated data mining processes are required. Outlier detection is such a data mining task. It is needed for finding objects in data which show a rare and unexpected behaviour. According to Hawkins, an outlier is “an observation which deviates so much from other observations as to arouse suspicions that it was generated by a different mechanism” [Haw80].

Traditional outlier mining approaches that detect outliers in the full-dimensional space such as LOF [BKNS00] and ABOF [KSZ08] no longer provide useful results when the dimensionality of the data is very high. This is due to some outliers only deviating in a certain subset of attributes. To cope with this problem, outlier detection in subspaces, i.e. in subsets of the attributes, is required. But considering subspaces leads to new challenges: How should significant subspaces be found out of the exponential number of existing subspaces. Let us consider the area of market research. If a manager of a store is interested in the buyer behaviour for food, the attributes vegetables, fruit or cheese are interesting whereas the buyer behaviour for hair care is of no interest. Two existing approaches are SOD [KKSZ09] which only considers one subspace per object and OUTRES [MSS11] which individually selects relevant subspaces per object. However, this problem was already well-researched in the context of subspace clustering [PHL04, KKZ09], so it would be desirable to be able to benefit from these results. The combination of subspace clustering and outlier mining was first introduced by OutRank [MASS08, Mül10].

Extending the idea of OutRank [MASS08, Mül10], our outlier ranking functions are based on subspace analysis: subspace clustering approaches [SZ04, MAG⁺09, AWY⁺99] perform the extraction of subspaces whereas scoring functions compare the subspaces found with respect to the object’s behaviour in them. A drawback of OutRank is that it uses a

very simple outlier scoring function, i.e. the number of objects per cluster and the dimensionality of the subspaces, so that hidden outliers can still be missed. The outlier ranking functions proposed in the following section perform outlier mining based on the idea of OutRank but we propose more powerful measures and therefore we enhance scoring functions to perform their task.

2 Outlier Mining in Subspaces

The whole outlier ranking process of this approach is as follows: A subspace clustering algorithm preprocesses the raw data, then the scoring function processes the subspace clustering result and calculates a positive ranking value for each object. The ranking values are sorted and the objects that are at the beginning of the ranking result are the most likely outliers.

The scoring functions compare the subspaces found by the clustering to determine the degree of outlierness for each object. Analysing the different kinds of subspace clustering results and therefore the information that is provided by such results is the only way to define appropriate measures for computing the outlier degree of an object. We formalized two properties of subspaces: the similarity between subspaces and the similarity between clusters contained.

One of the outlier ranking functions proposed – the Subspace Similarity Function – uses the similarity between subspaces in order to calculate the degree of outlierness per object. The more dimensions two subspaces share, the more similar these subspaces are. With the subspace similarity, it can be checked if an object behaves in the same way in similar subspaces. If this is not the case, it is an indicator that the object is an outlier. More precisely, if an object is clustered in two subspaces, this should never contribute to the object being considered an outlier, regardless of the similarity of the subspaces. Thus, this object should get a low outlier score. But if an object is clustered in one subspace but not in another one, the similarity of the subspaces is a valuable measure for calculating the degree of outlierness of this object. Considering the case that the two subspaces are dissimilar or share only a small amount of the same dimensions, it is likely that the object is not again clustered. Therefore, while this object should be considered as an outlier, it should not get the highest outlier score. An object should get the highest outlier score if the two subspaces are similar to a great extent. As the behaviour of the object should be the same in similar subspaces, the object shows an unexpected behaviour. This indicates that the object is an outlier and not only considered an outlier due to the dissimilarity of the subspaces. Please note, that the meaning of the score in the Subspace Similarity Function is inverse to the outlier score: a low score here implies a high outlier score.

The definition of the Subspace Similarity Function is exemplarily shown for the two outlier ranking functions in Definition 1.

Definition 1. *Subspace Similarity Function*

Given a subspace clustering result $R = \{(C_1, S_1), \dots, (C_m, S_n)\}$ with a cluster C_i in its corresponding subspace S_j . For each object o that is clustered, the following score is

calculated:

$$score(o) = \sum_{o \in (C, S)} \text{mean} \left(\sum_{S^*} \text{similarity_function}(o, S, S^*) \right)$$

with

$$\text{similarity_function}(o, S, S^*) = \begin{cases} \text{dist}(S, S^*) & \text{if } \nexists C^* \text{ in } S^* : o \in (C^*, S^*) \\ 1 & \text{otherwise} \end{cases}$$

The ranking score is only calculated for objects that are part of at least one cluster. If an object is never clustered in the clustering result this indicates that the subspace clustering algorithm did not find any objects from which this object does not deviate to a great extent. This means, that the clustering algorithm marks these objects as outliers. As the outlier ranking is based on the subspace clustering algorithm, it adopts the decision of considering objects to be outliers and assigns the highest outlier score to these objects. For all other objects which are clustered the ranking function must decide for itself to which extent an object is considered an outlier.

3 Evaluation & Conclusion

In comparison to other outlier mining techniques (LOF, ABOF, OutRank, SOD and OUT-RES), our ranking functions show more precise outlier detection behaviour: fewer false positives and more true positives. This was confirmed by creating and analysing ROC plots [HK06] for several data sets. The effect of clustering result characteristics on the quality of the outlier ranking result was thoroughly evaluated: Both of our functions compute high-quality ranking results when using PROCLUS. The Subspace Similarity Function additionally achieved good results when used in combination with RESCU. The experiments on real world data proved that the merits of our functions are not merely theoretical, but are also valid for practical use cases.

In summary, the clear separation between clustering and outlier detection allows for improvements in the former to directly improve the quality of the latter as well. We presented a general solution of the outlier ranking function for various subspace clustering approaches and therefore formalized several properties of subspaces which are used in the Subspace Similarity Function – one representative solution.

Outlier ranking based on subspace clustering is a very young research area: only OutRank and the outlier ranking functions proposed in this thesis use this approach. Future research on this method of outlier mining in subspaces is promising. It is simple to introduce new scoring functions and improvements in subspace clustering can automatically lead to improvements in outlier mining.

References

- [AWY⁺99] Charu C. Aggarwal, Joel L. Wolf, Philip S. Yu, Cecilia Procopiuc, and Jong Soo Park. Fast algorithms for projected clustering. In *Proceedings of the ACM SIGMOD International Conference on Management of Data*, pages 61–72, 1999.
- [BKNS00] Markus M. Breunig, Hans-Peter Kriegel, Raymond T. Ng, and Jörg Sander. LOF: identifying density-based local outliers. *SIGMOD Record*, 29:93–104, 2000.
- [Haw80] D. Hawkins. *Identification of outliers*. Chapman and Hall New York, 1980.
- [HK06] Jiawei Han and Micheline Kamber. *Data Mining: Concepts and Techniques*. Morgan Kaufmann, 2006.
- [KKSZ09] Hans-Peter Kriegel, Peer Kröger, Erich Schubert, and Arthur Zimek. Outlier Detection in Axis-Parallel Subspaces of High Dimensional Data. In *Proceedings of the 13th Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining*, pages 831–838, 2009.
- [KKZ09] H.P. Kriegel, P. Kröger, and A. Zimek. Clustering high-dimensional data: A survey on subspace clustering, pattern-based clustering, and correlation clustering. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 3(1):1–58, 2009.
- [KSZ08] Hans-Peter Kriegel, Matthias Schubert, and Arthur Zimek. Angle-based outlier detection in high-dimensional data. In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 444–452, 2008.
- [MAG⁺09] Emmanuel Müller, Ira Assent, Stephan Günnemann, Ralph Krieger, and Thomas Seidl. Relevant Subspace Clustering: Mining the Most Interesting Non-Redundant Concepts in High Dimensional Data. In *Proceedings of the IEEE International Conference on Data Mining (ICDM)*, pages 377–386, 2009.
- [MASS08] Emmanuel Müller, Ira Assent, Uwe Steinhausen, and Thomas Seidl. OutRank: ranking outliers in high dimensional data. In *Proceedings of the 2nd International Workshop on Ranking in Databases (DBRank)*, pages 600–603, 2008.
- [MSS11] Emmanuel Müller, Matthias Schiffer, and Thomas Seidl. Statistical Selection of Relevant Subspace Projections for Outlier Ranking. In *Proceedings of the IEEE 27th International Conference on Data Engineering (ICDE)*, pages 434–445, 2011.
- [Mül10] Emmanuel Müller. *Efficient Knowledge Discovery in Subspaces of High Dimensional Databases*. PhD thesis, RWTH Aachen University, 2010.
- [PHL04] Lance Parsons, Ehtesham Haque, and Huan Liu. Subspace clustering for high dimensional data: a review. *SIGKDD Explorations Newsletter*, 6:90–105, 2004.
- [SZ04] Karlton Sequeira and Mohammed Zaki. SCHISM: A New Approach for Interesting Subspace Mining. In *Proceedings of the 4th IEEE International Conference on Data Mining (ICDM)*, pages 186–193, 2004.

alpha-OffSync: An Offline-Capable Synchronization Approach for Distributed Document-Oriented Process Management in Healthcare

Andreas M. Wahl and Christoph P. Neumann
Friedrich-Alexander University, Erlangen-Nuremberg, Germany

Abstract: The α -OffSync project offers a synchronization concept for α -Flow, an electronic process support in heterogeneous inter-institutional scenarios in healthcare. A distributed case file is provided by α -Flow to represent workflow schemas as documents which are shared coequally to content documents. α -OffSync allows the detection and resolution of global concurrent modification conflicts on process artifacts based on locally available information by using logical timestamps. Further mechanisms are included to ease the management and dynamic extension of the group of participating actors.

1 Introduction & Background

The research project α -Flow, as described in [NL09], [NL10], [NL12], provides an approach to electronic process support in healthcare following the case handling paradigm [vdAWG05] while focusing on distributed environments and document-oriented systems integration. The traditional paper-based interaction paradigm in medicine is picked up and further extended. Both the often initially unknown amount of participating actors and the dynamic order of treatment steps in healthcare processes are considered to offer benefit to the cooperating physicians.

Relevant medical documents are gathered in a distributed electronic case file called α -Doc. Each α -Doc represents one α -Episode in which the actors collaborate in the treatment of a specific patient. An α -Doc consists of multiple α -Cards which are the units of validation, organizational accountability and subject to synchronization. Changes in the treatment process are reflected by the creation or modification of α -Cards. Each α -Card is decomposed into a descriptor with process-relevant status attributes [NSWL11] and a payload carrying information about medical results.

Unlike passive paper-based case files, an α -Doc offers a variety of additional active properties [TN11]. The α -Doc, as an active document [HM00], embeds all facilities for direct interaction with itself and the embedded pieces of information. It does not depend on another application, but is a light-weight application itself. Among others, these active properties include automatic synchronization with the remaining actors and administration of the group of collaborating participants. Actors only need a file copy of the α -Doc in order to participate. The α -Doc replicas represent nodes in an overlay network for synchronization.

α -Docs are files on the participant's desktop and active only when opened. In terms of networking, all nodes have an offline characteristic and usually no two peers are online at the same time. The α -OffSync objective is to provide generic concepts for synchronization that is offline-capable such that locally conducted synchronization operations provide global consistency across all physically distributed but logically centralized replicas. It also covers technical integration of these concepts into the existing system architecture of α -Flow.

2 Methods & Objectives

The communication channel is supposed to use store-and-forward to deliver messages once recipients are reachable. Non-FIFO behavior of the channel is tolerated as messages may be delayed in transit or arrive out-of-order.

Due to these preconditions, common synchronization mechanisms, based on mutual exclusion or on other techniques that require reaching a global consensus between participants, are not sufficient for the α -Flow system. A suitable protocol must detect global conflicts, but for reconciliation a local decision must suffice: Further communication is not possible because all other nodes are offline. A decision about a valid version must be derived instantly otherwise the local human actor is blocked. Local reconciliation must ensure global consistency among all nodes.

The synchronization adopts lists of logical timestamps, inspired by vector clocks [Fid88] and version vectors [PJPR⁺83]. Further synchronization approaches such as Independent Updates or Timestamped Anti-Entropy had been evaluated; a survey is best gained by [Len97].

3 Result

An *adaptive version clock* (AVC) has been implemented as a new data structure for dynamic lists of logic clocks and an offline-capable *OffSync protocol* for synchronization of dynamic nodes in a totally partitioned network (cf. [Wah11]). Each AVC holds timestamps in an associative array, which contains one key for each locally known node in the overlay network. Associated counters indicate how often which actors have modified an α -Card from which network nodes. Each modification increments the counter of the change originator. Applying logic clocks is necessary to establish an order on change events of an artifact. This order is based on element-wise comparing the AVCs received from other nodes with the locally known AVCs.

The OffSync protocol requires a version control system (VCS) that supports artifact-specific logical timelines together with history manipulation capabilities. If the comparison indicates that an incoming version properly succeeds the latest local

one, it is classified as a valid version. Gaps in an artifact timeline can be computed from AVCs and indicate the existence of further versions being delayed in transit. In both cases the incoming version is persisted at the appropriate position in the VCS history.

Concurrent modifications are global conflicts. Thus, the OffSync protocol queries the local version repository and computes the last globally non-conflicting version (LGNCV). All versions between the LGNCV and the conflict-causing one are invalidated; this means they are moved to a conflict branch within the VCS for data provenance purposes. The LGNCV is selected as the reconciled version that is inserted at the latest position of the timeline; its AVC is then set to the element-wise maximum of the locally known timestamps. Global consistency is achieved by local consistence reconciliation without further communication.

To enable dynamic change of actors, nodes and AVCs (now in terms of cardinality), a join protocol forms a second part of the OffSync protocol. New α -Docs introduce themselves to their peers and automatically synchronize by mutually exchanging information about locally known α -Cards based on their AVCs. New nodes are incrementally updated to the latest state. Each participant gains knowledge about newly joined actors.

The outlined concepts have been integrated into the existing prototypical system architecture of the α -Flow project in form of the α -OffSync subsystem. SMTP and IMAP are preliminarily used for store-and-forward communication. In the future, they are to be substituted by other protocols being more reliable in terms of guaranteed delivery, such as XMPP. Hydra [Had11] is used as VCS implementation.

4 Discussion

Unique to the α -OffSync approach is the ability to establish a shared view on the process state among all participating actors in totally partitioned networks where no guaranteed assumptions about the reachability of any network nodes can be made. Global conflicts can be detected and reconciled without additional online message exchange for determining a globally valid version. α -OffSync facilitates the management of dynamic groups of participants by minimizing the effort for inviting new actors and joining an ongoing treatment episode to provide the necessary flexibility for heterogeneous processes in healthcare.

Acknowledgements

Prof. Richard Lenz, who supervises the α -Flow project, provided invaluable support. Many useful comments and his open minded personality made him a great contributor to α -OffSync.

References

- [Fid88] C. J. Fidge. Timestamps in message-passing systems that preserve the partial ordering. In *Proc of the 11th Australian Computer Science Conf*, volume 10, pages 56–66, 1988.
- [Had11] Scott A. Hady. Logical Unit Oriented Version Control System with Version Validity Support. Diploma thesis, University of Erlangen, 2011.
- [HM00] E. Heinrich and H.A. Maurer. Active documents: Concept, implementation and applications. *Journal of Universal Computer Science*, 6(12):1197–1202, 2000.
- [Len97] Richard Lenz. *Adaptive Datenreplikation in verteilten Systemen*. Teubner, Stuttgart, 1997. PhD dissertation.
- [NL09] Christoph P. Neumann and Richard Lenz. alpha-Flow: A Document-based Approach to Inter-Institutional Process Support in Healthcare. In *Proc of the 3rd Int'l Workshop on Process-oriented Information Systems in Healthcare*, September 2009.
- [NL10] Christoph P. Neumann and Richard Lenz. The alpha-Flow Use-Case of Breast Cancer Treatment – Modeling Inter-Institutional Healthcare Workflows by Active Documents. In *Proc of the 8th Int'l Workshop on Agent-based Computing for Enterprise Collaboration (ACEC)*, June 2010.
- [NL12] Christoph P. Neumann and Richard Lenz. The alpha-Flow Approach to Inter-Institutional Process Support in Healthcare. *International Journal of Knowledge-Based Organizations (IJKBO)*, 2, 2012. Accepted for publication.
- [NSWL11] Christoph P. Neumann, Peter K. Schwab, Andreas M. Wahl, and Richard Lenz. alpha-Adaptive: Evolutionary Workflow Metadata in Distributed Document-Oriented Process Management. In *Proc of the 4th Int'l Workshop on Process-oriented Information Systems in Healthcare*, August 2011.
- [PJPR⁺83] D. S. Parker Jr, G. J. Popek, G. Rudisin, A. Stoughton, B. J. Walker, E. Walton, J. M. Chow, D. Edwards, S. Kiser, and C. Kline. Detection of mutual inconsistency in distributed systems. *IEEE TSE*, 9(3):240–247, 1983.
- [TN11] Aneliya Todorova and Christoph P. Neumann. alpha-Props: A Rule-Based Approach to ‘Active Properties’ for Document-Oriented Process Support in Inter-Institutional Environments. In Ludger Porada, editor, *Lecture Notes in Informatics (LNI) Seminars 10 / Informatiktage 2011*. Gesellschaft für Informatik e.V. (GI), March 2011.
- [vdAWG05] W. M. P. van der Aalst, M. Weske, and D. Grünbauer. Case handling: a new paradigm for business process support. *Data & Knowledge Engineering*, 53(2):129–162, 2005.
- [Wah11] Andreas Maximilian Wahl. alpha-OffSync: Verteilte Datensynchronisation in Form von IMAP-basiertem Mailtransfer. Bachelor thesis, University of Erlangen, 2011.

Ein evolutionärer Algorithmus auf einem FPGA zur gleichmäßigen Netzauslastung

Projektgruppe SWARM, Universität Siegen

J.-C. Barczak, S. Hardt, J. Schmidt, K.-U. Müller, B. Brachthäuser,
M. Mielke, A. Grünewald, R. Brück

Abstract: Intelligente Systeme zur besseren Energieausnutzung werden in den kommenden Jahren zunehmend wichtiger. Es geht nicht nur darum die Energie weniger umweltbelastend zu erzeugen, sondern auch den zur Verfügung stehenden Strom besser auszunutzen. Eine studentische Arbeitsgruppe der Universität Siegen hat ein Konzept entwickelt, welches eine gleichmäßigere Auslastung des Stromnetzes ermöglicht. Das Hauptaugenmerk liegt hierbei auf der Vermeidung von Spitzenlasten, da diese durch Kraftwerke, die mittels regenerativer Energiequellen arbeiten, nur bedingt abgefangen werden können. Die Umsetzung dieses Konzepts erfolgte durch Implementierung eines evolutionären Algorithmus auf einem FPGA.

1 Einleitung

Die Stromerzeugung aus erneuerbaren Energiequellen hat in der Vergangenheit weltweit stark zugenommen und wird auch in Zukunft steigen [1]. So betrug beispielsweise die weltweit installierte Leistung an Windenergie im Jahr 2000 17,4 MW, 2010 waren es bereits 196,3 MW [2]. Der Anteil der Stromerzeugung aus erneuerbaren Energien am Bruttostromverbrauch soll bis 2050 in Deutschland über 80% betragen [3]. Außerdem wird eine steigende Zulassung von Elektrofahrzeugen bis zum Jahre 2020 erwartet [4]. Diese werden hauptsächlich während der Arbeitszeit und abends geladen und erzeugen so Lastspitzen im Stromnetz [5], die vom Stromlieferanten abgefangen werden müssen. Kraftwerke, die mit erneuerbaren Ressourcen arbeiten, können Energie nur bedingt auf Abruf bereitstellen. Deshalb werden Lastspitzen heute noch durch konventionelle Kraftwerke gedeckt. Können solche Lastspitzen jedoch verringert bzw. vermieden werden, kann auch der eingespeiste Strom verringert werden. Deshalb hat eine studentische Arbeitsgruppe (Projektgruppe SWARM) der Universität Siegen ein Konzept entwickelt, mit dem eine gleichmäßigere Netzauslastung erreicht werden kann. Die Ladesteuerung von mehreren Verbrauchern soll hierbei nicht durch eine zentrale Steuereinheit geschehen, sondern dezentral durch die Verbraucher selber. Die Verbraucher werden zu einem Netzwerk zusammengeschaltet, in welchem sie untereinander ihre Ladezeitpunkte aushandeln. Dadurch ergibt sich eine hohe Ausfallsicherheit des Systems und es sind keine zusätzlichen Steuergeräte für den Netzwerk-Aufbau notwendig. Der Netzbetreiber gibt hierzu die maximale Leistung des Stromnetzes bekannt, die abgerufen werden kann. Jeder Teilnehmer im Netz bekommt anfangs bestimmte Initialwerte: Deadline, Ladedauer, Verbrauch, Schaltbarkeit

sowie eine eindeutige ID. Die Deadline gibt an bis zu welchem Zeitpunkt der Ladevorgang abgeschlossen sein soll, während die Ladedauer angibt, wie lange dieser dauert. Durch den Wert von Schaltbarkeit kann eingestellt werden, ob der Ladevorgang unterbrochen werden darf. Die Problemstellung ist, dass die Verbraucher vor einem gewissen Zeitpunkt geladen werden müssen. Des weiteren darf eine gegebene maximale Netzauslastung nicht überschritten werden. Dieses Problem lässt sich grundlegend auf das Rucksackproblem zurückführen [6], welches NP-vollständig ist. Für die Lösung dieses Problems ist ein evolutionärer Algorithmus [7] besonders gut geeignet. Die Berechnung der Ladezeitpunkte wird durch einen solchen erreicht. Aufgrund der dezentralen Steuerung muss eine ständige Kommunikation der Teilnehmer untereinander möglich sein. Jedem Teilnehmer wird dabei eine eindeutige ID zugewiesen. Die Möglichkeit zur Kommunikation, sowie den evolutionären Algorithmus hat die Projektgruppe auf einem FPGA realisiert. In Kapitel 2 wird zunächst die Funktionsweise evolutionärer Algorithmen im Allgemeinen und die des implementierten Algorithmus erklärt. Es folgt ein Überblick über den Aufbau des Designs (Kapitel 3) und dessen Simulation (Kapitel 4). Das letzte Kapitel beschreibt den aktuellen Stand und gibt einen Ausblick auf die weitere Arbeit.

2 Evolutionärer Algorithmus

Ein evolutionärer Algorithmus ist ein Optimierungsverfahren, welches auf dem Vorbild der biologischen Evolution beruht. Dabei vererben einzelne Individuen (eine Lösung) einer Population (eine Lösungsmöglichkeit) ihre Gene, falls sie eine vorherige Selektion überstehen. Zudem verändert Mutation vorhandene Individuen, wodurch neue Populationen entstehen. Der entwickelte evolutionäre Algorithmus führt zu Beginn einer Berechnung eine Initialisierung durch. Hierbei werden zufällig Verbraucher aus dem Netzwerk ausgewählt und in einer sogenannten Bestenliste gespeichert. Die Bestenlisten werden dann durch eine mathematische Funktion bewertet, der daraus resultierende Wert (Fitness) spiegelt die Qualität der Bestenliste wieder. Die Bestenliste mit der höchsten Fitness wird im Netzwerk verteilt. Die von einem Teilnehmer empfangenen Bestenlisten werden mit den eigenen erstellten verglichen, wobei nur die fünf mit dem höchsten Fitnesswert gespeichert werden, alle anderen werden verworfen. Nach dieser Initialisierung wird ein Zyklus von 50 Berechnungen durchgeführt. In diesem Zyklus werden die gespeicherten Bestenlisten mutiert, dabei können neue Verbraucher hinzugefügt und vorhandene gelöscht werden, wobei Verbraucher nur dann gelöscht werden können, wenn diese auch schaltbar sind. Durch diese Mutation entstehen neue Kombinationen von Verbrauchern (mutierte Listen). Zu diesen mutierten Listen wird jeweils eine Fitness berechnet, wobei die am besten bewertete mutierte Liste an das Netzwerk verteilt wird. Nach der Verteilung der eigenen besten mutierten Liste wird gewartet bis die mutierten Listen aller Teilnehmer im Netzwerk empfangen wurden. Nach den 50 Berechnungen hat jeder Teilnehmer die am besten bewertete Liste von Verbrauchern empfangen. Anhand dieser Liste wird entschieden, welche Verbraucher vor dem nächsten Berechnungsvorgang ihren Ladevorgang beginnen.

3 Module

Die Module des Designs (siehe Abb.1) sind in die drei Gruppen Algorithmus, Kommunikation und Steuerung unterteilt. Der Kern der Implementierung ist der evolutionäre Algorithmus, bestehend aus dem Mutationsmodul sowie dem Fitnessmodul. Das Modul Mutation variiert die Elemente der von den anderen Teilnehmern gesendeten Bestenlisten und legt die Listen anschließend im Speicher (Memory) ab. Das Fitnessmodul berechnet zu den mutierten Listen die entsprechende Fitness. Die Kommunikation kann auf zwei Arten geschehen: zum einen über ein Two-Wire-Interface (TWI), zum anderen über einen SPI-Bus. Das TWI ist an den I^2C -Bus [8] angelehnt und besitzt eine Kollisionserkennung. Es dient als Kommunikationsschnittstelle zwischen Verbraucher und dem Netzwerk. Der SPI-Bus stellt einerseits eine Möglichkeit dar, die Kommunikation auf ein anderes System zu erweitern (z.B. PowerLAN oder Drahtlossysteme), andererseits werden darüber die Initialdaten mittels Debug Controller in den Speicher geschrieben. Durch die Möglichkeit eines direkten Zugriffs auf den Speicher können mutierte Listen und berechnete Fitnesswerte gelesen und so eventuelle Fehler erkannt werden. Die Auswahl der Kommunikationsart geschieht über einen Multiplexer, welcher durch einen externen Eingang angesprochen wird. Die Kommunikation wird durch den Communication Controller gesteuert. Dieser startet auch die Mutation und entscheidet, welche Daten im Netzwerk verteilt werden. Für ein definiertes Zeitsignal hat der Communication Controller einen Timer-Eingang. Zu jedem Timer-Signal wird ein On/Off-Ausgang gesetzt. Dieser gibt an, ob der angeschlossene Verbraucher geladen wird. Der Memory Controller übernimmt die Ansteuerung des Speichers, sowie die Sortierung der Bestenlisten nach Fitnesswert.

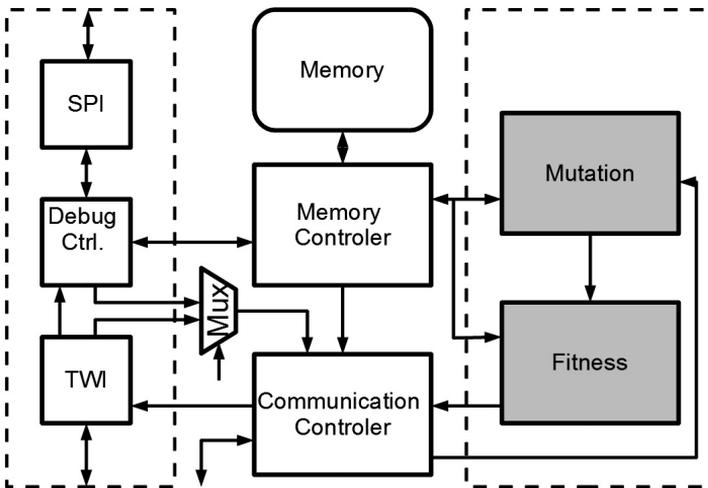


Abbildung 1: Modulübersicht

4 Simulation

Der evolutionäre Algorithmus wurde mit Hilfe eines Java Programms getestet. Dabei wurden mehrere tausend Simulationen mit zufällig generierten Initialwerten von Verbrauchern durchgeführt. Als Referenzalgorithmus wurde ein Algorithmus gewählt, der die Verbraucher mit nahender Deadline zuerst anschaltet, solange bis die maximale Netzauslastung erreicht ist. Bei den Simulationen wurde festgestellt, dass der evolutionäre Algorithmus im Durchschnitt eine geringere Zeit braucht, bis alle Verbraucher vollständig geladen sind, und dabei eine höhere Netzauslastung hat. Zudem überschreiten wenige Verbraucher die geforderte Deadline.

5 Aktueller Stand und Ausblick

Neben dieser Simulation des Algorithmus wurde das gesamte Verhalten erfolgreich verifiziert, es wurde unter anderem ein komplettes Netzwerk mit 16 Verbrauchern simuliert. Des Weiteren wurde ein Netzwerk aus fünf FPGAs aufgebaut und erfolgreich getestet. Durch einen zusätzlichen FPGA ist es möglich, die Kommunikation auf dem Bus zu überwachen und so die Korrektheit der übertragenen Daten zu bestätigen.

Das Einsatzgebiet Ladezeitenregelung, insbesondere von Elektrofahrzeugen, muss nicht das einzige bleiben. Sinnvoll wäre auch die Implementierung in Haushaltsgeräten wie zum Beispiel Waschmaschinen. Der nächste Schritt ist der Entwurf und die Fertigung eines ASIC (application-specific integrated circuit) in einer 350 nm Technologie mit vier Metalllagen bei austriamicrosystems. Durch die Realisierung des Projekts als ASIC wäre eine Massenfertigung möglich, wodurch die Kosten pro Chip minimiert würden.

Literatur

- [1] Bull, S.R.: Renewable energy today and tomorrow, Proceedings of the IEEE. 2001; S. 1216-1226.
- [2] WWEA Head Office: World Wind Energy Report 2010. 2011.
- [3] BMWi: Energiekonzept für eine umweltschonende, zuverlässige und bezahlbare Energieversorgung. Berlin, 2010.
- [4] Dr. Rolf Bulander: Antriebsstrangvielfalt und Elektrifizierung. Bosch, 2010; S. 15.
- [5] VDE: VDE-Studie Elektrofahrzeuge: Bedeutung, Stand der Technik und Handlungsbedarf. 2010; S.209.
- [6] Karsten Weicker: Evolutionäre Algorithmen. Wiesbaden, 2007; S. 128.
- [7] Oliver Kramer: Computational Intelligence. Springer-Verlag, Berlin, Heidelberg, 2009.
- [8] NPX Semiconductors: I^2C -bus specification and user manual. 2007.

Concurrent Programming for Scalable Web Architectures

Benjamin Erb
Ulm University
benjamin.erb@uni-ulm.de

Abstract: Web architectures are an important asset for various large-scale sites, such as social networks or ecommerce sites. Being able to handle huge numbers of users concurrently is essential, thus scalability is one of the most important properties of these architectures. Multi-core CPUs, distributed backend architectures and new web technologies are forcing us to reconsider approaches for concurrent programming in order to fulfill scalability demands. While focusing on different stages of scalable web architectures, we provide a survey of competing concurrency approaches and point to their adequate usages.

1 Introduction

The Internet and the World Wide Web are the backbone of our modern society and economy. Despite its background as hypermedia system, the web is progressively embracing interactive, collaborative and soft real-time features. Modern web applications are already replacing native mobile and desktop applications. The ubiquity of the web bears new challenges for web-based applications and their underlying architectures. Handling and processing huge numbers of concurrent requests is common for social platforms and large enterprise applications. Distributed backend systems are often deployed to meet these scalability requirements and the hardware is utilized to full extent. This architectural trend and the shift towards multi-core processors reemphasizes the question of how to handle and program concurrency. We will have a look on programming concurrency in different stages of web architectures – connection handling, application logic and backend persistence.

2 Scalable Web Architectures

Web architectures are server-side architectures that provide applications (like web sites or web services) over HTTP. Traditional web architectures have evolved from web servers for static content to more complex setups also serving dynamic content. Tiered architectures have emerged that separated connection/request handling, business logic and backend data persistence. Caches, reverse proxies and load balancers are often added to the architecture to improve performance and scalability. In the context of web applications, scalability is a crucial feature in multiple domains. The most important one is coping with huge

numbers of parallel requests/connections and still maintaining reasonable throughput, low latency and steady service quality. Collaborative and real-time web applications often use techniques like *long polling* or protocols such as the upcoming *WebSocket* standard related to *HTML5*. Both require huge numbers of open connections in parallel, though many of them are mostly idle. Unlike regular requests that are easily parallelizable, these new techniques introduce new forms of interaction and coordination between requests.

Scalability can be achieved using two basic strategies – scaling up (more resources per node) and scaling out (more nodes per system). It is common practice to focus on horizontal scale by deploying lots of low cost commodity systems. While there are many possible opportunities to improve scalability of a web application and its underlying architecture, we focus solely on concurrency.

2.1 Highly Concurrent Web Servers

The challenge of developing a scalable web server is the right trade-off between I/O utilization, scheduling and resource allocation per connection. A scalable web server must be able to handle as many concurrent connections as possible while still maintaining operational with graceful degradation. There are two main design decisions to make:

Activity model Servers either associate a dedicated activity (i.e. process/thread) to an incoming connection/request, or they use a single flow of execution and an event-based model that multiplexes I/O events of multiple connections/requests.

I/O model I/O operations can be both synchronous or asynchronous and blocking or non-blocking, providing different semantics and performance impacts.

The thread-per-connection model combined with synchronous, blocking I/O has been used predominantly for servers, while there have also been several event-based servers, mostly using asynchronous, non-blocking I/O [PDZ99]. There are different opinions on the programming models of these two approaches [vBCB03, Ous96]. But event-based servers have proven to provide slightly better performance in case of many concurrent connections compared to regular threads [PBH⁺07]. This is mainly caused by the lack of context switching, stack management and preemptive scheduling. There are also some efforts to combine both models [WCB01, LZ07, AHT⁺02], and to improve thread libraries to compensate for their disadvantages [vBCZ⁺03].

2.2 Concurrency in Application Logic

For our considerations, we assume that a dedicated application server handles application logic. This decouples the application server from the web server that handles connections. Application logic often combines CPU-bound and I/O bound operations, such as validation functions, computations and template rendering, respectively database queries, file access

or communication with backend services. For minimal delay, these operations need to be executed as concurrently as possible. They still require proper coordination, though.

The most common form of concurrent programming is based on **threads, locks and shared state**. It is available in most languages used for web programming, including *Java*, *Perl* or *Python*. Being a low-level approach, it allows explicit control and fine-grained tuning for good performance. But it is also said to be difficult and error-prone [Ous96] due to the risk of livelocks, deadlocks or race conditions and composability issues. **Software transactional memory** [HM93, HMJH08] isolates operations on shared states similar to database systems. It allows lock-free programming and the programming language *Clojure* has recently demonstrated its advantages. The **actor model** [Agh90] represents a completely different approach that isolates mutability of state. Actors are separate, single-threaded entities that communicate via immutable, asynchronous and guaranteed messaging. Thus, actors encapsulate state and provide a programming model that embraces message-based distributed computing. Popular languages using this model are especially *Erlang* and recently *Scala*. A related approach are asynchronous, **event-driven architectures** [Hoh06]. The most popular platform following this principle is currently *node.js*. Its single-threaded model prevents locking and concurrency issues and provides high performance for I/O bound operations. However, it takes extra efforts to scale CPU-bound applications and the event-based design results in an inversion of control.

These approaches can be found in different programming languages and frameworks used for web application development, and we have only covered the most important here. There is no silver bullet, so appropriate concurrency approaches must be chosen depending on the actual requirements and challenges of web applications.

2.3 Scalable & Concurrent Backend Storage

Historically, the database layer has been the first tier of a web architecture that had to take into account concurrency challenges due simultaneous requests. This remains unchanged, however new challenges have arisen due to the massive scale of data to be stored and the huge amount of simultaneous queries to be processed. As a result, distributed databases, non-relational database models and weakened consistency models such as *eventual consistency* [GL02] have been rediscovered and have gained increasing popularity.

3 Conclusion

We have seen that concurrency is crucial for scalability, which in turn is inherent for large-scale architectures. Event-based, non-blocking connection handling provides a good performance in case of many concurrent connections. The downside of lock-based concurrency has reinforced interest in alternative approaches, that can be used for application programming. Relational, transaction-oriented database system are getting complemented by alternative systems that favor weaker consistencies, but provide better scalability.

References

- [Agh90] Gul Agha. Concurrent object-oriented programming. *Commun. ACM*, 33:125–141, September 1990.
- [AHT⁺02] Atul Adya, Jon Howell, Marvin Theimer, William J. Bolosky, and John R. Douceur. Cooperative Task Management Without Manual Stack Management. In *Proceedings of the General Track of the annual conference on USENIX Annual Technical Conference*, pages 289–302, Berkeley, CA, USA, 2002. USENIX Association.
- [GL02] Seth Gilbert and Nancy Lynch. Brewer’s conjecture and the feasibility of consistent, available, partition-tolerant web services. *SIGACT News*, 33:51–59, June 2002.
- [HM93] Maurice Herlihy and J. Eliot B. Moss. Transactional memory: architectural support for lock-free data structures. In *Proceedings of the 20th annual international symposium on computer architecture*, ISCA ’93, pages 289–300, New York, NY, USA, 1993. ACM.
- [HMJH08] Tim Harris, Simon Marlow, Simon Peyton Jones, and Maurice Herlihy. Composable memory transactions. *Commun. ACM*, 51:91–100, August 2008.
- [Hoh06] Gregor Hohpe. Programming Without a Call Stack à Event-driven Architectures. Technical report, eaipatterns.com, 2006.
- [LZ07] Peng Li and Steve Zdancewic. Combining events and threads for scalable network services implementation and evaluation of monadic, application-level concurrency primitives. In *Proceedings of the 2007 ACM SIGPLAN conference on Programming language design and implementation*, PLDI ’07, pages 189–199, New York, NY, USA, 2007. ACM.
- [Ous96] John Ousterhout. Why Threads are a Bad Idea (for most purposes). In *USENIX Winter Technical Conference*, January 1996.
- [PBH⁺07] David Parigi, Tim Brecht, Ashif Harji, Peter Buhr, Amol Shukla, and David R. Cheriton. Comparing the performance of web server architectures. In *Proceedings of the 2nd ACM SIGOPS/EuroSys European Conference on Computer Systems 2007*, EuroSys ’07, pages 231–243, New York, NY, USA, 2007. ACM.
- [PDZ99] Vivek S. Pai, Peter Druschel, and Willy Zwaenepoel. Flash: an efficient and portable web server. In *Proceedings of the annual conference on USENIX Annual Technical Conference*, pages 15–15, Berkeley, CA, USA, 1999. USENIX Association.
- [vBCB03] Rob von Behren, Jeremy Condit, and Eric Brewer. Why events are a bad idea (for high-concurrency servers). In *Proceedings of the 9th conference on Hot Topics in Operating Systems - Volume 9*, pages 4–4, Berkeley, CA, USA, 2003. USENIX Association.
- [vBCZ⁺03] Rob von Behren, Jeremy Condit, Feng Zhou, George C. Necula, and Eric Brewer. Capriccio: scalable threads for internet services. In *Proceedings of the nineteenth ACM symposium on Operating systems principles*, SOSP ’03, pages 268–281, New York, NY, USA, 2003. ACM.
- [WCB01] Matt Welsh, David Culler, and Eric Brewer. SEDA: an architecture for well-conditioned, scalable internet services. In *Proceedings of the eighteenth ACM symposium on Operating systems principles*, SOSP ’01, pages 230–243, New York, NY, USA, 2001. ACM.

Realisierung eines VoIP-Testnetzwerkes mit einem Network Impairment Generator

Kevin Ernst

Hochschule Darmstadt
Fachbereich Informatik
Kevin-Ernst@web.de

Art der Arbeit: Projekt
Betreuer/in der Arbeit: Prof. Dr. Michael Massoth

Abstract: Dieses Projekt beschäftigt sich mit dem Monitoring von Voice over IP Dienstleistungen. Beim Monitoring von solchen IP basierten Diensten unterscheidet man grundsätzlich intrusive und nicht intrusive Methoden, allgemein werden diese auch als aktives und passives Monitoring bezeichnet. Beim passiven Monitoring wird bestehender Verkehr auf einer Plattform überwacht. Beim aktiven Monitoring wird im Gegensatz zum passiven Monitoring eigener Verkehr erzeugt und bemessen. Ziel dieses Projektes ist die Erstellung eines idealen Testnetzwerkes, sowie die Integration eines Network Impairment Generators, basierend auf einer Open Source Lösung. Dieser soll in der Lage sein innerhalb des Testnetzwerkes spezifische Störungen auf IP Ebene zu generieren, wie zum Beispiel Packet Loss (Paketverlust), Delay (Laufzeitverzögerung) und Jitter (Taktschwankungen). Mit Hilfe des Network Impairment Generators ist es einerseits möglich die Auswirkungen solcher Störungen in IP Netzen zu analysieren, andererseits kann das System auch als Referenzumgebung für bereits vorhandene Monitoringsysteme genutzt werden.

1 Grundlagen

1.1 Voice over IP

Der Begriff Voice over IP (kurz VoIP) bezeichnet das Telefonieren über Netzwerk, wie zum Beispiel über das Internet. Die Nutzung von VoIP hat in Deutschland in den letzten Jahren stark zugenommen. Im Vergleich zum Jahr 2009, wo immerhin bereits 8,7 Millionen Nutzer regelmäßig VoIP Dienste nutzten, gab es im Jahr 2010 eine Steigerung von 18% auf 10,3 Millionen Nutzer. [Bit10] Für das Jahr 2011 wird wiederum mit einem Anstieg dieser Nutzerzahlen gerechnet, man schätzt hierbei die regelmäßige Nutzerzahl für das Jahr 2011 auf 11,2 Millionen. [Bit11] Diese Steigerung lässt sich vor allem durch die sehr niedrigen Gesprächskosten und durch eine steigende Nutzung von kostenlosen Instant Messengern mit VoIP Funktion, wie beispielsweise Skype erklären.

1.2 Session Initiation Protocol

Das Session Initiation Protocol (kurz SIP) ist ein auf der Anwendungsschicht liegendes Kontrollprotokoll, dessen Hauptaufgabe die Erzeugung, Modifizierung, sowie das Beenden von Verbindungen zu einem oder mehreren Gesprächsteilnehmern ist. SIP basiert auf dem meist genutzten Protokoll im Internet, dem Hypertext Transfer Protocol (kurz HTTP). Zur Kommunikation nutzt SIP Requests (Anfragen) und Responses (Antworten). In der folgenden Grafik lässt sich erkennen wie SIP diese Requests und Responses nutzt um eine Verbindung aufzubauen: [Jj07] [Ngs02]

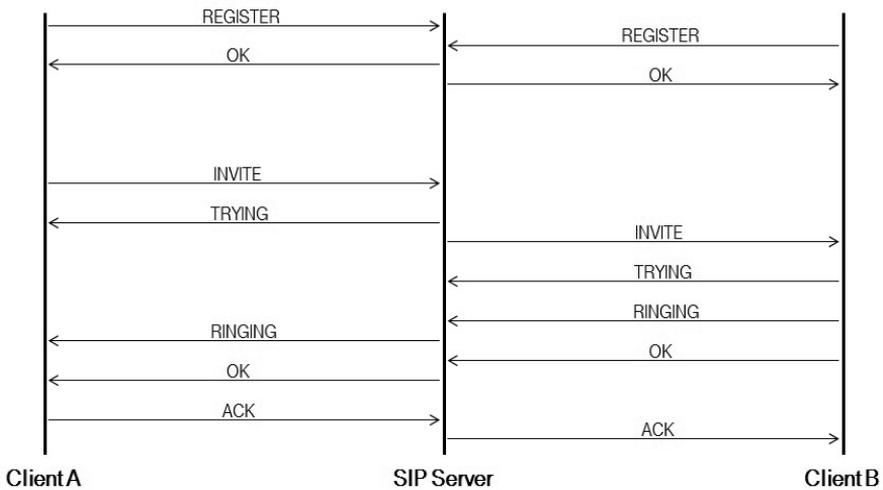


Abbildung 1: Aufbau eines VoIP Calls

1.2 Real-Time Transport Protocol

Im Gegensatz zu SIP, welches nur für den Verbindungsauf und -abbau zuständig ist, übernimmt das Real-Time Transport Protocol (kurz RTP) den Austausch von Sprach- oder Videodaten. RTP ist ein ebenfalls auf der Anwendungsschicht liegendes Protokoll, dessen Hauptaufgabe die Übermittlung von Multimediestreams ist. Da solche Multimediestreams echtzeitintensiv sind, nutzt RTP eine Übertragung via UDP und besitzt daher keine Fehlererkennung oder -behebung. [Bdt09] [Ngr03]

2 Konzeption & Realisierung

2.1 Testnetzwerk

Um ein funktionsfähiges Testnetzwerk aufzubauen benötigt man mindestens zwei SIP Clients und einen SIP Server. Hierbei macht es keinen Unterschied, ob es sich bei den SIP Clients um IP Phones oder entsprechende Softwarelösungen handelt. Aus Kostengründen wurde hier der kostenlose SIP Client Jitsi in der Version 1.0 verwendet. [Jit12] Die Wahl des SIP Servers fiel auf den Open Source SIP Server Kamailio in der Version 3.1, da dieser alle benötigten Funktionen zur Verfügung stellt und noch zahlreiche Erweiterungsmöglichkeiten bietet. Die Konfiguration erfolgt hierbei in einer zentralen Datei, die Zugriff auf zahlreiche Module bietet, welche in Kamailio enthalten sind. [Kam12]

2.1 Network Impairment Generator

Um das Projekt abzuschließen, wurde in das vorhandene Testnetzwerk ein Network Impairment Generator integriert. Um einen geeigneten Network Impairment Generator zu finden, wurde eine Recherche auf dem Open Source Markt und eine anschließende Softwareevaluation durchgeführt. Hierzu wurden unterschiedlichste Kriterien ausgearbeitet, anhand derer die gefundenen Lösungen bewertet werden konnten. Eine optimale Lösung erzeugt Störungen wie Packet Loss, Delay und Jitter realitätsgetreu und in einer beliebigen Zusammenstellung, sowie Intensität. Sie verfügt über eine angemessene Konfiguration und eine Dokumentation in ausreichendem Umfang. Die Installation darf nicht zu komplex sein und keine großartigen Abhängigkeiten aufweisen. Im Idealfall gibt es noch weitere nicht in den Anforderungen genannte Zusatzfunktionen, die für eventuelle Erweiterungsmöglichkeiten genutzt werden können. Die einzelnen Kriterien wurden anhand ihrer Prioritäten gewichtet und anschließend wurde jede gefundene Lösung anhand dieser Kriterien bewertet, um so die optimale Lösung zu finden. Eine grobe Zusammenfassung der Softwareevaluation findet man in der folgenden Tabelle:

Rang	Lösung	Hauptkritik
1	WANulator	Keine
2	WANEM	Keine GUI
3	Netem	Keine GUI
4	NIST NET	Unübersichtliche Dokumentation
5	MGEN	Sehr Komplex

Um nun ein funktionsfähiges Netzwerk zu schaffen, müssen die einzelnen Bestandteile folgendermaßen angeordnet werden:

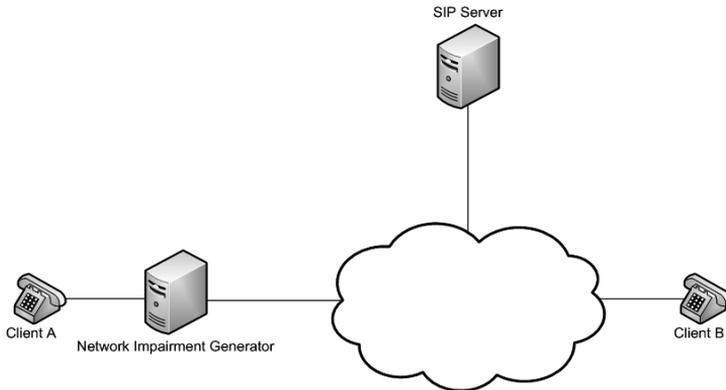


Abbildung 2: Netzaufbau

3 Ausblick

Da dieses Projekt unmittelbar vor einer Bachelorarbeit durchgeführt wurde, wird das hier aufgebaute Testnetzwerk mit integrierten Network Impairment Generator als Grundlage für diese Arbeit verwendet. Die anschließende Bachelorarbeit wird sich mit dem Einfluss von Packet Loss auf die Ende Sprachequalität in Voice over IP Netzen beschäftigen. Hierzu müssen zahlreiche Testszenarien entwickelt, durchgeführt und anschließend analysiert werden, um am Ende eine passende Aussage über die Beziehung von Packet Loss und Sprachequalität treffen zu können.

Literaturverzeichnis

- [Bit10] Bitkom; Markt und Statistik – Internet Telefonie, 2010.
http://www.bitkom.org/de/markt_statistik/64046_63174.aspx
- [Bit11] Bitkom; Markt und Statistik – Internet Telefonie, 2011.
http://www.bitkom.org/de/markt_statistik/64046_66760.aspx
- [Ke01] Kompendium Elektronik, VoIP - Voice over I, 2011
<http://www.elektronik-kompendium.de/sites/net/0503131.htm>
- [Jj07] Janak, Jan; SIP Introduction, 2007.
- [Bdt09] Le Boudec, Duda, Thiran; Real Time Protocol (RTP), 2009.
- [Tw07] Ulrich Trick, Frank Weber; SIP, TCP/IP und Telekommunikationsnetze, Oldebourg-Verlag, 2007
- [Ngr03] Network Working Group; RFC 3550, RTP: A Transport Protocol for Real-Time Applications, 2003
- [Ngs02] Network Working Group; RFC 3261, SIP: Session Initiation Protocol, 2002
- [Kam12] Kamailio, <http://www.kamailio.org>
- [Jit12] Jitsi, <http://jitsi.org/>

Smart Energy Grids: Ein interdisziplinäres Forschungsvorhaben an der Hochschule Darmstadt

Olaf Gaydoss, Stephan Schaible und Michael Massoth

Hochschule Darmstadt
Fachbereich Informatik
o.gajdoss@googlemail.com, StephanSchaible@gmx.de
michael.massoth@h-da.de

Art der Arbeit: Projekt Systementwicklung
Betreuer der Arbeit: Prof. Dr. Michael Massoth

Abstract: Dies ist eine Machbarkeitsstudie zu einem interdisziplinären Forschungsvorhaben zum Thema *Smart Energy Grids*. Der Beitrag dokumentiert eigene Vorarbeiten und den aktuellen Stand des Vorhabens. Insbesondere wird ein Konzept für die geplante Smart Energy Grid-Infrastruktur an der Hochschule Darmstadt vorgestellt.

1 Einleitung und Motivation

Smart Energy Grids (deutsch: intelligente Stromnetze) gehören zu den größten wissenschaftlich-technischen Herausforderungen und Investmentchancen unseres Jahrzehnts (2012+X). Die Hochschule Darmstadt möchte verstärkt Studierende in diesem Zukunftsbereich ausbilden.

2 Bedeutung der Informations- und Kommunikationstechnologie

Die intelligenten Stromnetze der Zukunft sind in hohem Maße von computerbasierten Steuersystemen abhängig. Diese Systeme sind zunehmend mit offenen IKT-Netzwerken, wie dem Internet verbunden, und dadurch den Gefahren und Risiken des Cyberspace ausgesetzt, wie zum Beispiel der Angriff des Computerwurms Stuxnet auf die Siemens-Steuerungstechnik für Kraftwerke belegt. Smart Grids zählen deshalb zu den sicherheitskritischen Infrastrukturen.

Der IKT-Bereich spielt bei der Realisierung und dem zuverlässigen Betrieb von Smart Grids eine entscheidende Rolle. Er dient als Steuerungs- und Kreislaufsystem zur Verbindung der verschiedenen Systeme und Geräte von Energiebereich bis Anwendungsbereich und übernimmt die Kommunikation und Telemetrie in beide Richtungen der Energiezulieferkette.

Im IKT-Bereich werden dieselben Internettechnologien genutzt, durch die bereits andere Hightech-Branchen (wie zum Beispiel Automotive und Maschinenbau) transformiert wurden. Diese Internettechnologien werden nun für intelligente Grids eingesetzt. Auf oberster Ebene kann der IKT-Bereich von Smart Grids in drei Bereiche unterteilt werden:

Energiebereich	Überwachung und Steuerung von Energieerzeugung, -übertragung, -transformation, -zwischenlagerung, -verteilungsnetz und -verbrauch.
Kommunikationsbereich	Netzwerke (WAN, LAN, Field und Home Area Network) und IKT-Infrastruktur der Energieversorger zur internen Kommunikation, für die Datenerfassung, Messdaten, Voice- und Dateninformationen, sowie IT-Sicherheit und Schutz der sicherheitskritischen Infrastrukturen.
Anwendungsbereich	Steuerung der Nachfragereaktion, Abrechnung, Ausfallkontrolle, Lastenüberwachung, Home Automation, Facility Management, in Echtzeit arbeitende Energiemärkte, sowie neue Anwendungen und Dienstleistungen für Endkunden, einschließlich Schutz der persönlichen Daten und der Privatsphäre der Anwender.

3 Die aktuelle Infrastruktur / das aktuelle Stromnetz

Das deutsche Stromnetz umfasst eine Länge von 1,73 Millionen Kilometern und ist in vier Spannungsebenen eingeteilt. Das Niederspannungsnetz (230 V oder 400 V), mit dem die allermeisten Endverbraucher versorgt werden, macht über eine Million Kilometer aus, während die Übertragungsnetze (Höchstspannungsnetz: 220 kV oder 380 kV) sich zu einer Länge von rund 35.000 km addieren. Das Hochspannungsnetz (Verteilnetz: 60 kV oder 100 kV) ist knapp 77 000 km lang, das Mittelspannungsnetz (3 kV bis 30 kV) fast 500.000 km. Diese Netze befinden sich auf der Niederspannungsebene hauptsächlich im Besitz von Stadtwerken und anderen lokalen Versorgern, die Übertragungsnetze gehören vor allem den großen Energiekonzernen. [Stand 2008]

Doch nicht nur Stadtwerke haben in der Vergangenheit die Stromnetze als Einnahmequellen genutzt. Nach Hochrechnungen haben die Netzbetreiber seit 2006 etwa 21 Milliarden Euro an Netzgebühren eingenommen, wovon lediglich ein Zehntel auch tatsächlich in die Instandhaltung der Netze geflossen sein soll. Der Zustand der teilweise veralteten Stromnetze erfordert aber nach Einschätzung der Bundesnetzagentur sehr viel höhere Investitionen, um die Netze ohne Stromausfälle betreiben zu können

4 Eigene Vorarbeiten und aktueller Stand

Am Fachbereich EIT (Prof. Dr. Metz, Energietechnik) gibt es bereits einen kompletten Aufbau zur Simulation von Smart Grid-Stromnetzen, die in den nächsten Jahren und Jahrzehnten dominieren werden. Dort wurde weiterhin in Zusammenarbeit mit der HSE AG ein erster Aufbau der Infrastruktur realisiert. Diese bereits getätigten Investitionen können im Rahmen geplanten Vorhabens umfangreicher genutzt werden.

Der Fachbereich Informatik (Prof. Dr. Massoth, Telekommunikation, BMBF-Forschungsprojekt NextFactor) verfügt bereits über Expertenwissen im Bereich Home Automation bei der intelligenten Steuerung und Überwachung von im Haushalt vorhandenen Geräten.

Am Fachbereich Informatik soll deshalb ein Smart Meter Gateway, sowie etwa 30 Smart Meter mit den entsprechenden Multi Utility Communication Controllern, angeschafft werden. Vgl. dazu auch Abbildung 1.

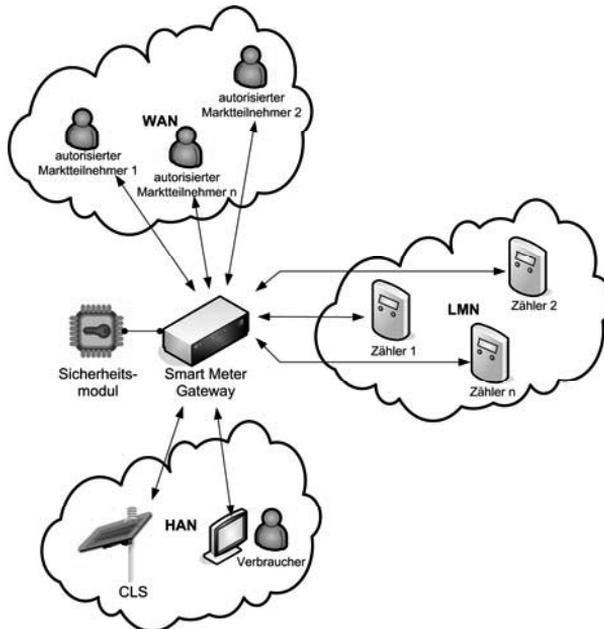


Abbildung 1: Das Smart Meter-Gateway

Durch Anschaffung des Smart Meter Gateways bekäme der Fachbereich Informatik erstmals ein computerbasiertes Steuersystem für die Lehre sowie für Forschung und Entwicklung. Damit wäre es möglich, sich mit dem bereits vorhandenen System zur Simulation von Smart Grid-Stromnetzen am Fachbereich EIT (Prof. Dr. Metz, Energietechnik) zu vernetzen und das System um den Bereich Home Automation zu erweitern.

5 Gesamtkonzept, Ziele und beabsichtigte Wirkung

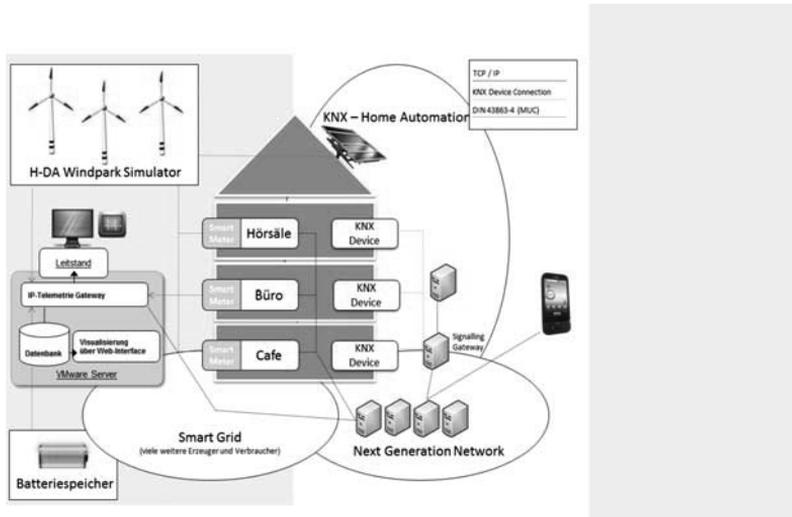


Abbildung 2: Gesamtkonzept der geplanten Smart Energy Grid-Infrastruktur an der Hochschule Darmstadt

In der Abbildung 2 ist das Gesamtkonzept für die geplante Smart Energy Grid-Infrastruktur an der Hochschule Darmstadt aufgezeigt. Damit wäre es erstmalig ein interdisziplinäres Projekt im Bereich Energietechnik und Informations- und Kommunikationstechnologie zum Thema *Smart Energy Grids* an der Hochschule Darmstadt möglich.

Die Hochschule Darmstadt möchte damit auch ihre Attraktivität im MINT-Bereich verbessern und verstärkt Studierende im Zukunftsbereich der Smart Energy Grids (deutsch: intelligente Stromnetze) ausbilden.

Literaturverzeichnis

- [1] SmartGrids – Europäische Forschungsplattform, www.smartgrids.eu
- [2] Bernhard Fenn, Dieter Metz; SmartGrids: Wege zu intelligenten Stromnetzen mit breitem Einsatz von Regenerativen Energien am Beispiel Darmstadt, Ringvorlesung der Hochschule Darmstadt, Januar 2009
- [3] Dieter Metz: Netzführung und Netztraining für SmartGrids, Marcus Evans, Netzleittechnik-Seminar, Hilton Köln, Juni 2009
- [4] Bernhard Fenn, Dieter Metz; SmartGrids mit dezentralen Speichern in Verteilnetzen, VWEW Jahrbuch 2010, VWEW-Verlag (erscheint Nov. 2009)

Hardware Concepts for Elliptic Curve based Zero-Knowledge Protocols

André Himmighofen
andre@ahimmighofen.de
Hochschule RheinMain
Master's Thesis

Betreuer:

Prof. Dr. Steffen Reith Prof. Dr. Jörn Steuding
(Hochschule RheinMain) (Universität Würzburg)
GI-Fachbereich Technische Informatik

Abstract: This thesis shows how to implement a zero knowledge protocol based on elliptic curve cryptography (ECC) for FPGAs in an efficient way. Its implementation is founded on twisted Edwards Curves which are a fast alternative to other elliptic curves. To gain an acceptable size the implementation uses block RAM and has only one unified point operation instead of addition and doubling of points. The performance is increased by the use of projective coordinates which avoids many inversions, thus resulting in an attractive performance/size ratio. As far as the author knows, this is the first FPGA implementation of the Schnorr protocol over elliptic curves.

1 Background

1.1 Elliptic Curve Cryptography

According to [HMV03] an elliptic curve E over a prime field \mathbb{F}_p , $p > 3$ is defined as $y^2 = x^3 + ax + b$ where $a, b \in \mathbb{F}_p$ satisfies $4a^3 + 27b^2 \not\equiv 0 \pmod{p}$. A point (x, y) that satisfies the equation is on the specified curve. Additionally the point of infinity (\mathcal{O}) is defined as on the curve. $E(\mathbb{F}_p)$ is defined as the set containing all points on the curve E over \mathbb{F}_p . Two points on the curve $P(x_P, y_P), Q(x_Q, y_Q) \in \mathbb{F}_p$ are added by drawing a straight line through them. With one exception the straight line has an intersection with the curve. The result R of the addition is this intersection mirrored by the x-axis.

If Q is $(x_P, -y_P)$, i.e. Q is equal to P mirrored at the x-axis, there is no intersection with the curve. The result of that exception is \mathcal{O} . In the case that P and Q are equal we call the operation point doubling. The only difference between point addition and point doubling is that instead of a straight line through P and Q a tangent at point P is used [Sil09]. The multiplication of a point $P \in E(\mathbb{F}_p)$ with a scalar value $n \in \mathbb{F}_p$ is defined as $n * P = \underbrace{P + P + \dots + P}_{n\text{-times}}$. Akin to the repeated multiplying and squaring for calculating g^a

efficiently we can implement the multiplication by adding and doubling the point [Sil09].

1.2 Zero Knowledge Protocols

Zero knowledge protocols are based on the idea that a communication attendee proves that he knows a secret without sharing the secret to the other communication attendees [MvOV00]. Many zero knowledge protocols are constructed around mathematical problems already used for cryptographic purpose. We chose a protocol based on the discrete logarithm problem over finite fields. The Schnorr identification protocol is the result of the attempt to prove the knowledge of a secret number by utilizing the discrete logarithm problem over a finite field. With a sufficient bit length the Schnorr identification protocol requires only three protocol steps to prove the knowledge of the secret with a sufficient probability [Sch90] [MvOV00] [CPSS11].

An advantage of elliptic curve cryptography results from the possibility to simply replace the discrete logarithm problem with the elliptic curve discrete logarithm problem. While Qi describes in [Qi09] the general suitability of elliptic curve cryptography for zero knowledge proofs, [CPSS11] describes how to modify the Schnorr protocol.

Let p a prime with $p > 3$ and \mathbb{F}_p the corresponding finite field. A elliptic curve E over \mathbb{F}_p is chosen. G is a generator in $E(\mathbb{F}_p)$. The private key of the prover A is a random $x \in \mathbb{F}_p$. The corresponding public key is Q with $Q = x * G$. The prover A generates a random r with $r \in \mathbb{F}_p$ and computes $P = r * G$. He sends P to the verifier B . On the other side B generates a random challenge c with $c \in \mathbb{F}_p$ and sends it to A . The prover A now computes $m = r + c * x \bmod p$ and sends m back to B . After the three messages B has gathered all information to verify if A knows x with sufficient probability. B computes $Z = m * G - c * Q$. If $Z = P$, B knows that A knows x [CPSS11].

2 Design

Zero knowledge identification protocols are used to authenticate attendees of a communication. That means that the authentication is only the first step of the communication. After the participants of the protocol have proven their identities to each other, there are further message exchanges. The resulting implementation of the Schnorr identification protocol is designed as a co-processor which can be used within the main communication. The design was split in three parts. Compute $P = r * G$ and $m = r + cx \bmod p$ on prover side and verify that $Z = P$ with $Z = mG - cQ$ on verifier side. For the scalar multiplications the ECC multiplication is utilized, the point subtraction in the calculation of Z and the multiplication of c and x for m are passed trough the ECC module.

Twisted Edwards curves, a type of elliptic curves introduced in [BBJ⁺08], are one of the most efficient curves known. The point addition and doubling can be joined to one operation, the unified point operation. While separate point doubling and adding needs 20 multiplications and 15 additions the unified point operation consumes only 14 multiplications and five additions [BMB⁺09] [Edw07]. To achieve a small implementation we need to reduce the number of implemented operations. In that case the unified point operation will save many slices. To avoid unnecessary inversions, points are transformed to projective coordinates. This saves the divisor in a third component and requires only one

division at the end of the multiplication. Calculation with affine points (points with two components) would require an inversion after each addition [BMB⁺09].

3 Implementation

The VHDL implementation is constrained to $\mathbb{F}_{p_{192}}$, a finite field over the NIST prime p_{192} . According to the special form of the NIST primes, introduced in [nis00], a very simple modulo reduction can be used which only requires four additions [Sol99] [HMV03]. The implementation of the point addition uses a multiplication constructed around the modulo reduction. To gain an efficient implementation the point addition uses block RAM to store its internal values. Compared to register-based storage the RAM utilizes nearly no slices and requires only few additional cycles for data gathering. The scalar multiplication, implemented as double-and-add-always, and the Schnorr protocol implementation store their values in block RAM, too. Additions and comparisons are implemented as full size 192bit operations which takes more slices but results in a much better performance.

4 Results

The resulting VHDL Code was synthesized for Xilinx Spartan 3 XC3S5000 FPGA, all results are based on that device.

The multiplication module for two scalar values, including the modulo p_{192} reduction, utilizes 2531 slices and requires 16 clock cycles. The point addition, implemented as unified point operation, requires 3162 slices and 352 clock cycles including the multiplication. Based on the point addition the scalar multiplication requires 6858 slices. To multiply a 192bit scalar with a point the multiplication processes 384 point additions per multiplication which results in a runtime of 138861 clock cycles. The module can be clocked with 20.8 MHz which means 20,800,000 cycles per second. The module can perform 149 multiplications per second [Him11].

Altogether the three steps of the Schnorr protocol require three scalar multiplications which results in an overall effort of 17219 clock cycles. According to the clock speed of 20.8 MHz the module can perform 49 identifications per second. It utilizes 8335 slices, including all other operations mentioned above.

The approach described in [BMB⁺09] has a size of 4654 slices and requires 1160 cycles to perform a unified point operation, which means 445440 cycles for a scalar multiplication. The own implementation described above utilizes 6858 slices but with 138861 clock cycles only requires one third of the time to process a scalar multiplication [Him11].

References

- [BBJ⁺08] Daniel J. Bernstein, Peter Birkner, Marc Joye, Tanja Lange, and Christiane Peters. Twisted Edwards Curves, 2008.
- [BMB⁺09] Brian Baldwin, Richard Moloney, Andrew Byrne, Gary Mcguire, and William P. Marnane. A Hardware Analysis of Twisted Edwards Curves for an Elliptic Curve Cryptosystem. In *Proceedings of the 5th International Workshop on Reconfigurable Computing: Architectures, Tools and Applications*, ARC '09, pages 355–361, Berlin, Heidelberg, 2009. Springer-Verlag.
- [CPSS11] Ioannis Chatzigiannakis, Apostolos Pyrgelis, Paul G. Spirakis, and Yannis C. Stamatiou. Elliptic Curve Based Zero Knowledge Proofs and their Applicability on Resource Constrained Devices. In *MASS'11*, pages 715–720, 2011.
- [Edw07] Harold M. Edwards. A normal form for elliptic curves. In *Bulletin of the American Mathematical Society*, pages 393–422, 2007.
- [Him11] Andre Himmighofen. Hardware Concepts for Elliptic Curve based Zero-Knowledge Protocols. Master's thesis, Hochschule RheinMain, 2011.
- [HMV03] Darrel Hankerson, Alfred Menezes, and Scott Vanstone. *Guide to Elliptic Curve Cryptography*. Springer, 2003.
- [MvOV00] Alfred J. Menezes, Paul C. van Oorschot, and Scott A. Vanstone. *Handbook of applied cryptography*. CRC Press, 2000.
- [nis00] Digital Signature Standard (DSS), Federal Information Standards Processing Publication 186-2, 2000.
- [Qi09] Chengming Qi. A Zero-Knowledge Proof of Digital Signature Scheme Based on the Elliptic Curve Cryptosystem. In *Intelligent Information Technology Application, 2009. IITA 2009. Third International Symposium on*, volume 3, pages 612–615, nov. 2009.
- [Sch90] C. Schnorr. Efficient Identification and Signatures for Smart Cards. In Jean-Jacques Quisquater and Joos Vandewalle, editors, *Advances in Cryptology - EUROCRYPT '89*, volume 434 of *Lecture Notes in Computer Science*, pages 688–689. Springer Berlin / Heidelberg, 1990.
- [Sil09] Joseph H. Silverman. *The Arithmetic of Elliptic Curves*. Springer, 2009.
- [Sol99] Jerome A. Solinas. *Generalized Mersenne numbers*. Research report (University of Waterloo. Faculty of Mathematics). Faculty of Mathematics, University of Waterloo, 1999.

Constraint Modeling and Matching in a Dynamic Context Service Discovery and Binding Process

Alexander Kohout

University of Kassel
Distributed Systems Group, Prof. Dr. Kurt Geihs
alexanderkohout@gmail.com

Type of Work: Bachelor's Thesis
Advisor: Dipl. Inf. Michael Wagner

Abstract: Nowadays a lot of different information sources exist that provide information about our context. These information sources can be abstracted as context service, allowing for dynamic discovery and binding based on context requests of information consumers. Thereby, constraints are a powerful means to specify consumer needs and service characteristics. This paper presents an introduction of how constraints can be applied for dynamic context service discovery. Furthermore, an approach is provided to model and match constraints which are formulated as complex logical expressions. The approach is integrated in an architecture to automatically perform discovery and binding of context services.

1 Introduction

Modern mobile devices are equipped with various internal sensors and other information sources providing information about the current context, e.g. microphone; GPS sensor; camera(s); calendar; etc. Additional external context provider will be available in *ubiquitous computing environments* [Wei91], while it is possible that multiple internal and external context sources offer the same information. However, these context sources provide mobile applications with context information allowing for *context-awareness* and an according *self-adaption* [Abo99]. To use the best service for a certain context information, Wagner [Wag11] presents an approach to automatically discover and bind services at run-time with the help of a middleware. Internal and external context sources are abstracted as context service and are described by context offer. Applications query for needed information and the middleware matches these queries with available offers, choosing the best fitting service for receiving the queried information.

A better specification of application needs and service capabilities is reached with constraints, while it is possible to combine (atomic) constraints as complex logical expressions. Figure 1 demonstrates the application of constraints in an example scenario: *Context Query 1* asks for the *position of user Paul in cartesian coordinates*. This query sets constraints for *accuracy* and *battery costs*. Both depicted services, a GPS sensor on

the left and a Cell-ID based location sensor on the right side, offer the wanted information (different representations are automatically transformed [Rei10]). While both services satisfy the *accuracy* constraint, the GPS sensor violates the *battery cost* constraint. As a result, *Context Offer 2* will be chosen by the middleware.

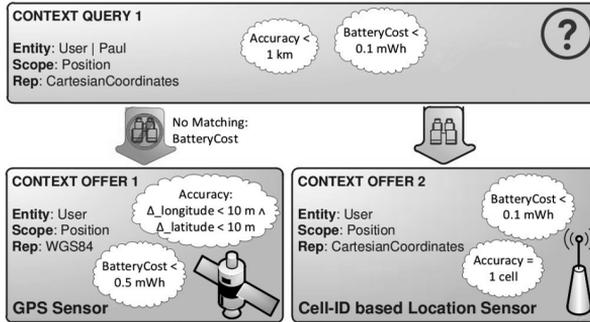


Figure 1: Example Application of Constraints

The application of constraints means that the matching mechanism has to guarantee that all application (query) constraints are satisfied by a service (offer). To realize this constraint matching, constraints of query and offer are first modeled as first-order formulas. In a next step the matching analyzes the conjunction of query and offer constraints for contradictions, i.e. violations of constraints. The remainder of this paper will give an introduction to these two steps. Section 2 starts with the transformation of constraints to first-order formulas. Section 3 deals with the according constraint matching and section 4 will give a short summary of this work.

2 Constraint Modeling using First-order Logic

Context information is generally expressed as a three-tuple consisting of *entity*, *scope* and *representation* [Wag11]. Entities describe what the information is about, scopes define the type of information, and representations refer to an ontological description of format, value ranges and unit (see figure 1). However, context queries and offers are specified using the XML-based *Context Offering and Query Language* (COQL), and *COQL documents* for query and offer serve as input for the constraint matching problem. An example how constraints are expressed within COQL is given in figure 2.

```
<MetadataConstraints xsi:type="coql:CompositeMetadataConstraint"
  MetadataConstraintID="query1_metadata_composite1" operator="OR">
  <Constraints xsi:type="coql:AtomicNumericalMetadataConstraint"
    MetadataConstraintID="query1_metadata_atomic1" metadata="Accuracy"
    representation="PolarCoordinates" operator="LT" delta="0" value="1000",
  <Constraints xsi:type="coql:AtomicNumericalMetadataConstraint"
    MetadataConstraintID="query1_metadata_atomic2" metadata="Reliability"
    representation="Percent" operator="GE" delta="0" value="97"/>
</MetadataConstraints>
```

Figure 2: Constraints expressed in COQL

Within figure 2, two kinds of constraints can be identified: atomic constraints for accuracy and reliability, and a composite constraint combining both atomic constraints as disjunction. Atomic constraints will be modeled as first-order predicates, using the operator as predicate symbol, value and delta as value border. This yields the following two predicates, where LT stands for less than and GE for greater or equal:

$$GE(\text{reliability}, 97) \quad LT(\text{accuracy}, 1000)$$

Both predicates describe through their predicate symbol and their value border a value range for the metadata given as first parameter (accuracy and reliability). The composite constraint can now easily be modeled using the defined logical operator:

$$LT(\text{accuracy}, 1000) \vee GE(\text{reliability}, 97)$$

This way of creating first-order formulas for constraints is done for query and offer. Both formulas serve as input for the following constraint matching.

3 Constraint Matching using Analytic Tableaux

The idea here is to analyze the conjunction of incoming offer and query formulas for contradictions. This is done using an analytic tableaux [Smu71] algorithm which breaks complex formulas down into subformulas by creating a tree structure. A contradiction is found if one branch in this tree is not satisfiable, and a found contradiction implies a violation of query constraints. Creating the tableau tree structure is done with defined tableau expansion rules (see figure 3).

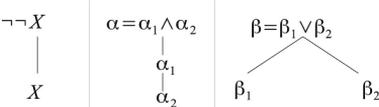


Figure 3: Tableaux Expansion Rules

With these three rules it is possible to break down complex constraint formulas into its subformulas and analyze these subformulas for contradictions. Figure 4 demonstrates the application of these rules on two example constraint formulas. The left side of the conjunction (1) contains the context query, the right side contains a potential context offer. The tableau splits this first conjunction into two subformulas (2) and (3). Applying the expansion rules to (2) causes a splitting of the tree into two new branches. Already expanded rules will be inherited by these branches. As a result, the two new emerged branches contain only predicates, so no more expansion rules can be applied. Now it is time to compare these atomic constraints with each other, analyzing whether value ranges of constraints that describe the same metadata are consistent with each other. In this example both branches show up contradictions. While the left side contains the contradiction accuracy of „less than 1000“ and „equal 1600“ (EQ stands for equal), the right side contains reliability of „greater or equal 97“ and „equal 95“.

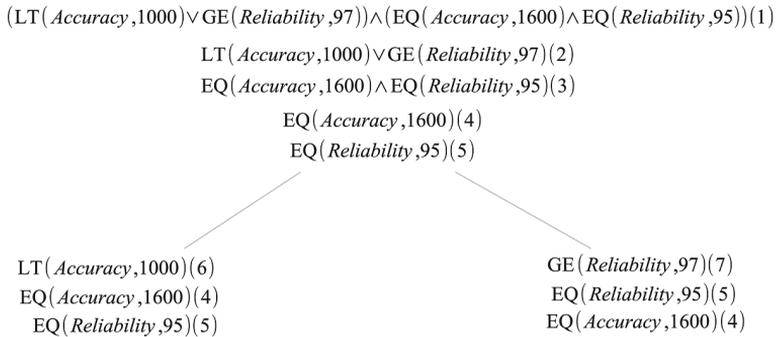


Figure 4: Example Tableau-Expansion

4 Conclusion

This paper gives a short overview about how constraints are applied within Wagner's work [Wag11], and presents in brief the according modeling and matching of constraints. Constraints are modeled using first-order logic. The according matching is done by searching for contradictions in the conjunction of two constraint formulas using an analytic tableaux [Smu71] algorithm. Found contradictions means that constraints are not satisfied and therewith that constraints do not match with each other. The evaluation of this approach [Koh11] shows a good performance on mobile platforms: matching one hundred constraints can be done in under one second, while the expected amount of used constraints per query and offer lies under ten constraints.

Literaturverzeichnis

- [Abo99] Abowd, Gregory et al.: Towards a Better Understanding of Context and Context-Awareness, in: Hans-W. Gellersen (Herausgeber) *Handheld and Ubiquitous Computing*, Bd. 1707 von *Lecture Notes in Computer Science*, Springer Berlin / Heidelberg (1999), S. 304–307
- [Koh11] Kohout, Alexander: *Context Constraints in a Dynamic Context Service Discovery and Binding Process*; Bachelor's Thesis; 2011
- [Rei10] Reichle, Roland: *Information Exchange and Fusion in Heterogenous Distributed Environments*, Dissertation, Universität Kassel, (2010)
- [Smu71] Smullyan, Raymond M.: *First-Order Logic*, Bd. 43 von *Ergebnisse der Mathematik und ihrer Grenzgebiete*, Springer-Verlag, 2. Aufl. (1971)
- [Wag11] Wagner, Michael; Reichle, Roland and Geihs, Kurt: *Context as a Service - Requirements, Design and Middleware Support*, in: *Proceedings of the Workshop on Middleware and System Support for Pervasive Computing (PerWare 2011)*
- [Wei91] Weiser, Mark: *The computer for the 21st century*. *Scientific American* (1991)

Implementierung eines interrupt-gesteuerten OSEK-Betriebssystemkerns auf einem ARM-Cortex-M3-Mikrocontroller

Rainer Müller
sirmuel@stud.informatik.uni-erlangen.de

Studienarbeit im Diplomstudiengang Informatik
Friedrich-Alexander-Universität Erlangen-Nürnberg

Abstract:

Im Rahmen dieser Arbeit wurde das SLOTH-Konzept für einen interrupt-gesteuerten Betriebssystemkern auf dem Mikrocontroller ARM Cortex-M3 implementiert. Bei diesem Ansatz werden für das Einplanen und Einlasten von Threads anstatt eines Software-Schedulers die Hardware-Komponenten zur Behandlung von Interrupts eingesetzt. Vor der Umsetzung wurde zunächst untersucht, wie die zur Verfügung stehende Hardware angesteuert und genutzt werden kann. Die fertige Implementierung wurde in der Evaluation mit einem herkömmlichen System mit einem software-basiertem Scheduler verglichen, um die positiven Effekte des SLOTH-Konzepts auf die Verwaltung von Threads zu bestätigen.

1 Das Sloth-Konzept

Ein Betriebssystem unterscheidet üblicherweise zwischen synchron aktivierten Threads, die von einem Scheduler in Software verwaltet werden, und Interrupt Service Routines (ISRs), welche durch Hardware ausgelöst, die CPU bei der Ausführung des aktuellen Kontrollflusses asynchron an beliebigen Stellen unterbrechen können. Durch diese Unterscheidung entsteht ein zweigeteilter Prioritätsraum, in dem eine ISR mit semantisch niedriger Priorität einen Thread mit hoher Priorität verdrängen [dFMAdN06], was insbesondere in Echtzeitsystemen zu Problemen führen kann.

Das SLOTH-Konzept stellt einen Weg vor, dieses Problem zu lösen, indem die Unterscheidung zwischen Threads und ISRs aufgehoben wird [HLSSP09, HLSP11]. Indem alle Kontrollflüsse als Unterbrechungsbehandlungen implementiert werden, entsteht ein einheitlicher Prioritätsraum, der es möglich macht, Prioritäten für Threads und ISRs aufgrund ihrer semantischen Funktion zu vergeben, ohne dabei Rücksicht auf die Art des Kontrollflusses nehmen zu müssen. Der größte Vorteil ist jedoch, dass die Hardware für Unterbrechungsbehandlung die Arbeit eines Schedulers übernehmen kann und damit eine Implementierung des Schedulers in Software wie in herkömmlichen Betriebssystemen unnötig macht.

Indem zum Einplanen und Einlasten der Kontrollflüsse Hardware-Komponenten verwendet werden, sollen in SLOTH gegenüber Systemen mit einem software-basierten Sche-

duler schnellere Kontextwechsel erreicht werden. Durch den Rückgriff auf den Interrupt-Controller ist dafür keine speziell angefertigte Hardware nötig, sondern für die Umsetzung entstehen lediglich Anforderungen an die jeweilige Zielplattform: diese muss verschiedene Prioritätsstufen für jeden Interrupt anbieten und es muss möglich sein, Interruptanforderungen in Software zu generieren, um synchron spezifische Interrupts auszulösen.

Viele moderne Plattformen erfüllen diese Anforderungen, wie zum Beispiel der ARM Cortex-M3, für den im Rahmen meiner Arbeit eine Implementierung des SLOTH-Konzepts umgesetzt wurde. Da das Konzept vor allem auf die Anwendung in ereignisgesteuerten, eingebetteten Echtzeitsystemen mit statisch konfigurierten Prioritäten abzielt, wird beispielhaft die OSEK¹-OS-Spezifikation implementiert, die vor allem in der Automobilindustrie Verbreitung findet [OSE05].

2 Implementierung auf dem ARM Cortex-M3

Der ARM Cortex-M3 ist ein Mikrocontroller mit einer 32-Bit-Architektur, der mit seinem einfachen Programmiermodell für den Einsatz in eingebetteten Systemen vorgesehen ist. Als festen Bestandteil des Designs enthält dieser den *Nested Vectored Interrupt Controller* (NVIC), der bis zu 240 externe Interrupt-Quellen mit bis zu 256 frei konfigurierbaren Prioritätsstufen verwalten kann. Damit werden die oben genannten Bedingungen für eine SLOTH-Implementierung erfüllt.

Die Aktivierung eines OSEK-Tasks wird in SLOTH durch einfaches Setzen des entsprechenden Bits im *Interrupt Set-Pending Register* des NVIC durchgeführt. Daraufhin prüft die Hardware selbsttätig die anliegenden Interruptanforderungen und bestimmt aus diesen die höchste Priorität. Ist die aktuelle Priorität des sich in Ausführung befindenden Kontrollflusses niedriger, wird der Prozessor unterbrochen und eine Unterbrechungsbehandlung durchgeführt, wobei direkt die Taskfunktion aus der Interruptvektortabelle angesprungen wird. Dadurch wird in SLOTH die Einlastung des aktivierten Tasks durch die Hardware vorgenommen. Die Terminierung eines Tasks entspricht dem Verlassen der Unterbrechungsbehandlung. Dabei muss darauf geachtet werden, dass der unterbrochene Kontext korrekt wiederhergestellt wird, um eine transparente Verdrängung zu gewährleisten.

In OSEK-OS gibt es neben den bisher beschriebenen einfachen (*basic*) auch erweiterte (*extended*) Tasks, welche zusätzlich blockieren können, um auf ein Ereignis zu warten. Ein blockierter Task gibt die CPU auf und erlaubt so anderen Tasks – auch mit niedrigerer Priorität – verarbeitet zu werden. Dieses Blockieren widerspricht an sich dem Ablaufmodell von Unterbrechungsbehandlungen, da diese bis zum Ende durchlaufen müssen. Die Priorität der laufenden Behandlung müsste also gesenkt werden, um Unterbrechungen mit niedriger Priorität zu erlauben. Dies ist beim NVIC des Cortex-M3 aber nicht vorgesehen, weshalb in SLOTH der komplette Kontext dieses Tasks gesichert und die Interruptquelle deaktiviert wird, um danach die Unterbrechungsbehandlung beenden zu können. Dann können andere Tasks mit niedriger Priorität weiterlaufen, während die Deaktivierung der Interruptquelle sicherstellt, dass der blockierte Task nicht erneut eingelastet wird. Erst

¹Abkürzung für "Offene Systeme und deren Schnittstellen für die Elektronik im Kraftfahrzeug"

Ausgewählte Testfälle		Task-Art	SLOTH	OSEK	Faktor
ActivateTask ()	ohne Einlastung	<i>einfach</i> und <i>erweitert</i>	9	233	25,9
ActivateTask ()	mit Einlastung	<i>einfach</i>	40	507	12,7
TerminateTask ()	mit Einlastung	<i>einfach</i>	24	310	12,9
ChainTask ()	mit Einlastung	<i>einfach</i>	61	471	7,7
ActivateTask ()	mit Einlastung	<i>erweitert</i>	60	505	8,4
TerminateTask ()	mit Einlastung	<i>erweitert</i>	30	312	10,4
WaitEvent ()	mit Einlastung	<i>erweitert</i>	96	325	3,3
SetEvent ()	mit Einlastung	<i>erweitert</i>	87	407	4,7

Tabelle 1: Ausführungszeit ausgewählter Systemaufrufe gemessen in Prozessorzyklen im Vergleich zwischen SLOTH und einer herkömmlichen OSEK-Implementierung auf dem ARM Cortex-M3.

wenn das entsprechende Ereignis zum Deblockieren signalisiert wird, führt dies zu einer Freigabe der Interrupt-Quelle und der Task wird erneut aktiviert. Der gesicherte Kontext wird von einer speziell generierten Funktion wiederhergestellt, die als Prolog vor die eigentliche Taskfunktion eines erweiterten Tasks gestellt wird.

Die SLOTH-Implementierung für den ARM Cortex-M3 unterstützt die Klassen BCC1 und ECC1 der OSEK-OS-Spezifikation, die neben Tasks und Ereignissen auch noch die Koordination von Tasks mittels Ressourcen und Alarme für die zeitgesteuerte Aktivierung von Tasks umfassen.

3 Evaluation

Die Implementierung wurde in verschiedenen Konfigurationen evaluiert und die Ergebnisse mit einem quell-offenen, kommerziell vermarkteten System mit einem Scheduler in Software verglichen, welches ebenso die OSEK-Schnittstelle implementiert. Verschiedene Anwendungskonfigurationen wurden aufgesetzt und die Anzahl der benötigten Prozessorzyklen für die Taskverwaltungsaufgaben gemessen. Dabei sticht vor allem die einfache Aktivierung eines Tasks niedrigerer Priorität hervor, bei der eine Beschleunigung um den Faktor 25,9 erreicht wird, da dies in SLOTH lediglich eine Schreiboperation in das Register des Interrupt-Controllers darstellt. Aber auch bei Aktivierung eines höher priorisierten Tasks mit anschließendem Einlasten erzielt SLOTH eine um den Faktor 12,7 verbesserte Laufzeit. Weiterhin wurde das Verhalten bei Blockieren und Wiederaufnahme in einer Konfiguration mit erweiterten Tasks gemessen. Obwohl das Blockieren von Tasks nicht dem Ablaufmodell von Unterbrechungsbehandlungen entspricht, erreicht SLOTH dabei im Vergleich zum software-basiertem Scheduler eine Beschleunigung von 3,3. Tabelle 1 zeigt eine Übersicht der beschriebenen, sowie weiterer ausgewählter Testfälle, wobei sämtliche gemessenen Beschleunigungsfaktoren im Bereich von 3,3 bis 25,9 liegen.

Der Vorteil des SLOTH-Betriebssystems für eine echte Anwendung hängt vom Anteil der Taskverwaltung im Programm ab. Dennoch ist es wichtig, dass vor allem eingebettete Betriebssysteme möglichst wenig Ressourcen verbrauchen und einen schnellen Durchsatz bei der Ausführung erreichen. Zusätzlich erlaubt SLOTH durch die Einführung eines gemeinsamen Prioritätsraums für Threads und ISRs, Prioritäten der Kontrollflüsse im System unabhängig von der synchronen oder asynchronen Aktivierung zu vergeben.

Literatur

- [dFMAAdN06] Luis E. Leyva del Foyo, Pedro Mejia-Alvarez und Dionisio de Niz. Predictable Interrupt Management for Real Time Kernels over conventional PC Hardware. In *Proceedings of the 12th IEEE International Symposium on Real-Time and Embedded Technology and Applications (RTAS 2006)*, Seiten 14–23, Los Alamitos, CA, USA, 2006. IEEE Computer Society.
- [HLSP11] Wanja Hofer, Daniel Lohmann und Wolfgang Schröder-Preikschat. Sleepy Sloth: Threads as Interrupts as Threads. In *Proceedings of the 32nd IEEE Real-Time Systems Symposium (RTSS 2011)*, Seiten 67–77, Los Alamitos, CA, USA, 2011. IEEE Computer Society.
- [HLSSP09] Wanja Hofer, Daniel Lohmann, Fabian Scheler und Wolfgang Schröder-Preikschat. Sloth: Threads as Interrupts. In *Proceedings of the 30th IEEE Real-Time Systems Symposium (RTSS 2009)*, Seiten 204–213, Los Alamitos, CA, USA, 2009. IEEE Computer Society.
- [OSE05] OSEK/VDX Group. Operating System Specification 2.2.3. Bericht, OSEK/VDX Group, Februar 2005. <http://portal.osek-vdx.org/files/pdf/specs/os223.pdf>.

Mental Stress Detection Using Multimodal Sensing in a Wireless Body Area Network

Marina Rodríguez Aliberas

Technische Universität Berlin
Telecommunication Networks Group (TKN)

Universitat Ramon Llull
Dept. d'Electrònica i Comunicacions de Enginyeria i Arquitectura La Salle

rodriguez@tkn.tu-berlin.de, marinara@gmail.com

Master Thesis submitted Sept. 2011 (GI-Fachbereich Technische Informatik)
Supervisors: Dipl.Inf. Jan-Hinrich Hauer, M.Sc. Rosa Maria Alsina Pagès,
Prof. Dr.-Ing. Adam Wolisz

Abstract: This thesis presents the design and implementation of a software framework that allows to collect and store physiological data on small sensor devices attached to a human being (forming a Wireless Body Area Network). It describes how sensor data from two biosensors (GSR and ECG) is collected in a number of field experiments, where subjects are exposed to different emotional situations. The experiments are done together with a group of researchers from the department of psychology at Freie Universität Berlin, FUB. The collected data is then used to (post factum) extract a set of features, which are used as input for existing machine learning algorithms to study how mental stress of human beings can be detected automatically. The evaluation of the algorithms shows that multimodal sensing can improve the stress detection accuracy.

1 Background

Mental stress is an emotion which has become frequent in world's population, to the point of affecting health and causing diseases. However, often people do not recognize or underestimate the level of stress they suffer. Consequently, the interest for monitoring body responses to better understand the level of stress people are experiencing in every day life has increased considerably [LWB07]. Stress and emotions have complex body responses and one single sensor is likely not enough to detect stress [HNS10]; it is more promising to analyse signals from several sources of the body to understand better the reaction to a specific emotion [P00].

The goal of this Master thesis was to develop a low-cost distributed platform for measurement of different biological signals and demonstrate that results of these measurements processed by properly selected machine learning algorithms (MLA) deliver correct results. For this purpose a software framework was developed, allowing to collect sensor data on sensor devices attached to human body. A wireless Body Area Network (BAN, [LBM11]) was used to synchronize the data from

the sensor nodes. The platform was used to perform experiments together with a group of researchers from the field of psychology. In the experiments the subjects were exposed to different emotional situations.

Major contributions of the thesis are the design and implementation of a software framework in the TinyOS 2 [LG09] operating system to collect human bio-signals; the design and implementation of a wireless time synchronization protocol to synchronize data from all components of a BAN; characterization of bio-signals using feature extraction; and the study of machine learning algorithms for stress recognition.

2 System Development and Experiments

Some BAN components needed for this project were already available: an operating system was given (TinyOS 2), wireless body sensors (Shimmer2R, [BGMOKASC10]), and some device drivers. ECG (electrocardiogram) and GSR (galvanic skin response) sensors were chosen for their good results showed in other research works focused on physiological responses related to emotions ([HCA11] , [HRB04] , [HNS10]). Once the hardware was set, there was the need to design and implement a software framework to allow an application to first collect the data from the body sensors and then store it in persistent memory.

The first step of the development was to analyse the hardware and given software (operating system). Afterwards a software framework for collecting the bio-signal data was designed and implemented according to the possibilities and limitations of the hardware. The software framework was structured in three main functional parts: a component to sample and get the data from the sensors, a component to store that data on an SD card and a component to synchronize the devices of the BAN; all those parts were coordinated by an application written in nesC [LG09].

There were some synchronization issues to be solved: how to have the same time reference in all devices, how to make this times accurate enough and how to remove the effect of 'clock drift' (clock drifts are provoked by a slightly differences on the oscillation frequency of a clock, compared to its theoretical work frequency). To solve those issues a time synchronization algorithm was designed, and also implemented in TinyOS 2. The solution proposed is based on a global time reference provided by a device with an accurate local oscillator which is broadcast by a base station to the rest of the BAN nodes in order to have a global time equivalence with their local time. In the process of sending a timestamp periodically, time delay errors where minimized to the order of milliseconds which resulted to be acceptable accuracy for our application.

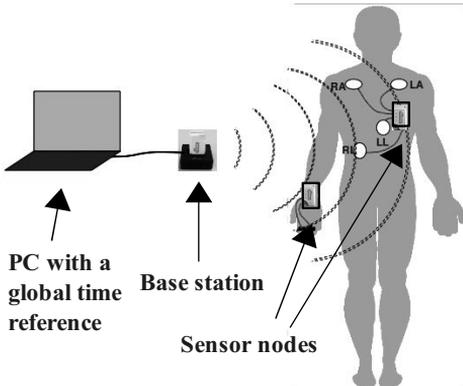


Figure 1. Platform schematic

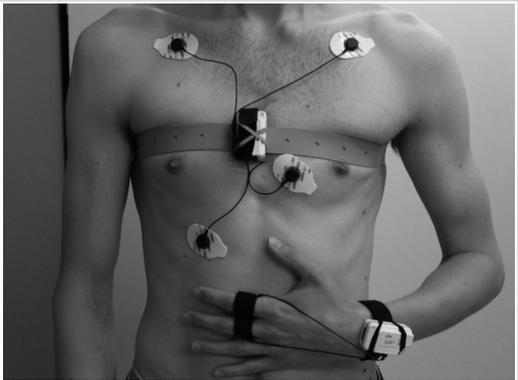


Figure 2. Participant with sensors/electrodes

The platform (see Figure 1) is composed by a synchronization device (a PC), a base station that broadcasts the synchronization packets (Shimmer2R node connected to the PC), and two sensor nodes attached to the body of the participant (one placed in the chest, ECG, and the other one in the wrist, GSR, see Figure 2).

After validating and testing the platform's performance, a series of experiments with 20 participants were done along with experts on emotion recognition (a group of researchers from the Cluster of Excellence '*Languages of Emotion*' from the Freie Universität Berlin). Each participant did two consecutive sessions (of approximately 20 minutes each) with a computer software used by the psychologists. For the period of a session the participant (wearing the body sensors) sat alone in front of a computer which showed a sequence of standardized pictures with emotional content (IAPS, [CBL96]). In parallel the platform was extracting bio-signals of the participant, so later on the all data fragments (GSR and ECG signals) were assigned to a user and an standardized rating.

3 Stress Detection with Learning Algorithms and Conclusions

Once the data from all the experiments was collected, it was processed by different MLA in order to know their ability to automatically stress. Usually MLAs do not use the raw signals to classify entries, but they work on some properties of the signal, so-called features. Examples of MLA used for emotion detection are [G09], [SKC10], [HCA11], [HCC06]. Before evaluating the algorithms, an important part was to select the features that best characterize both signal types (e.g. RMSSD or SCL).

This evaluation was carried out in Matlab along with existing feature extraction and classification libraries. Four steps summarize this process: Parsing and pre-processing the raw data from all participants, feature extraction of the parsed data, training of the algorithms and selection of the most suitable MLA for stress detection (through the analysis of the results). Four MLA were analysed: Bayesian Network, Random Decision Tree, K Nearest Neighbour using No-nested Generalized Examples (NNge) and Lazy IB1 instance based learning algorithm, [DHS01]. In order to train and compare the algorithms a software called Weka [DCS11] was used.

Instance-based algorithms (Lazy IB1) obtained the best results of classification, using a multimodal combinations of features (4 features: two GSR and two ECG features) to classify the data, with an accuracy of 69%. Better accuracies were obtained but with a worse performance in detecting stressed situations (instead they had a better detection of non-stress situations). Percentages of detection (slightly lower than in other experiments related to emotions) are considered to be improved by having a more even amount of stress and non-stress situations (50%-50%), normalizing the features extracted and having a bigger amount of experiment's data (just 75% of ECG and 40% of GSR experiment's data could be used for classification).

Although the system has some limitations and further research needs to be done to perform a more precise emotional features classification, this project and its experimental results show that physiological signals (ECG and GSR) provide a promising basis for detecting emotions. It also confirms that multimodal systems improve the accuracy of the detection, in response to the high complexity of emotions themselves. Future lines of work are suggested to improve the flexibility and accuracy of the platform (for example to design of a GUI for the platform, replacing the PC with a smartphone or supporting more interactive psychological test)

Literaturverzeichnis

- [HNS10] Healey J.; Nachman L.; Subramanian S.; Shahabdeen J.; Morris M. : Out of the Lab and into the Fray: Towards Modelling Emotion in Everyday Life. *Pervasive Comp.*, 2010. Vol. 6030/2010, p. 156-173. DOI: 10.1007/978-3-642-12654-3_10
- [P00] Picard R.W.: *Affective Computing*. 1st ed. Massachusetts: Massachusetts Institute of Technology Press, 2000. ISBN 0-262-16170-2
- [DCS11] Dept. of Computer Science: Weka Documentation. [online]. Waikato University, June 2011. <http://www.cs.waikato.ac.nz/>.
- [LG09] Levis P.; Gay D.: *TinyOS Programming*. Cambridge CB2 8RU, UK, 2009. ISBN: 978-0-521-89606-1
- [HCC06] Hsieh C.W.; Chen C.T.; Chao Y.P.; Chen J.H.; Study of Human Affective Response on Multimedia Contents. *World Congress on Medical Physics and Biomedical Engineering*, 2006. vol. 14, num. 6.
- [G09] Gao Y.: *A Digital Signal Processing Approach for Affective Sensing of a Computer User through Pupil Diameter Monitoring*. Florida International University, 2009.
- [HCA11] Hussain S; Calvo R.A.; AlZoubi O.; D’Mello S.: *Affect Detection from Multichannel Physiology during Learning Sessions*. National ICT Australia (NICTA), Australia, 2011.
- [CBL96] Cuthbert B.; Bradley M.; Lang P.: *Probing picture perception: Activation and emotion*. University of Florida, 1996. vol.33, num. 2 : p. 103-111.
- [DHS01] Duda R.O; Hart P.E; Stork D.G.: *Pattern Classification*. John Wiley & Sons, New York, 2001. ISBN: 0-471-05669-3
- [HRB04] Hjortskov N.; Rissén D.; Blangsted A.K; Fallentin N.; Lundberg U.; Sogaard K.: The effect of mental stress on HRV and blood pressure during computer work. *European Journal Applied to Physiology*, February 2004. vol. 92, num.: p. 84-89.
- [SKC10] Sun F.T.; Kuo C.; Cheng, H.T.; Buthpitiya S.; Collins P.; Griss M.: *Activity-aware Mental Stress Detection Using Physiological Sensors*. Silicon Valley Campus, 2010. Paper 23.
- [LBM11] Latré B; Braem B; Moerman I; Blondia C; Demeester P.: A survey on wireless body area networks. *Wireless Networks* , vol. 17, Issue 1, January 2011. DOI: 10.1007/S11276-010-0252-4
- [LWB07] Loew N.; Winzer k.; Becher G.; Schönfuß D.; Falck Th.; Uhrich G.; Ketterle M.; Scheller F.: *Medical Sensors of the BASUMA Body Sensor Network*. IFMBE Proceedings, 2007. vol. 3, 3rd session.: p. 171-176 . DOI: 10.1007/978-3-540-70994-7_30
- [BGM10] Burns A; Greene B; McGrath M; O’Shea T; Kuris B; Ayer S; Stroiescu; Cionca V.: *SHIMMER- A Wireless Sensor Platform for Noninvasive Bionedical Research*. *IEEE Sensor Journal* , 2010. vol.10 : p. 1527-1534.

Implementierung und Evaluierung eines verteilten Algorithmus für die Kanaluweisung in drahtlosen Maschennetzen

Simon Seif

Institut für Informatik
Freie Universität Berlin
Takustraße 9
14195 Berlin
seif.simon@googlemail.com

Abstract: Die Verbreitung drahtloser Maschennetze hat im letzten Jahrzehnt stark zugenommen. Die Leistung drahtloser Maschennetze ist allerdings durch entstehende Interferenzen beschränkt. Kanaluweisungsalgorithmen reduzieren die netzwerkweite Interferenz unter Verwendung orthogonaler Kanäle und erhöhen somit die Leistungsfähigkeit des Netzwerks. In dieser Arbeit wurden Leistungsmetriken und Benchmarkszenarien für die experimentelle Evaluation von Kanaluweisungsalgorithmen entwickelt. Ein verteilter Kanaluweisungsalgorithmus wurde implementiert und auf dem DES-Testbed der Freien Universität Berlin mit 106 Knoten evaluiert. Die Auswertung mit Hilfe des Benchmarks hat gezeigt, dass der Algorithmus in allen Szenarien eine signifikant höhere Leistung als eine Ein-Kanal-Konfiguration erzielt.

1 Einleitung

Einer der zentralen beschränkenden Faktoren für die Leistung drahtloser Maschennetze ist die Interferenz simultaner Übertragungen [GK00, JPPQ03]. Interferenz bezeichnet hierbei die gegenseitige Wechselwirkung, die simultane Übertragungen auf einem drahtlosen Medium aufeinander ausüben. Interferenz wird durch die Verwendung von Medienzugriffssteuerungsmechanismen, wie z.B. Carrier Sense Multiple Access/Collision Avoidance (CSMA/CA), vermieden. Allerdings sorgt der exklusive Zugriff auf das Übertragungsmedium dafür, dass der aggregierte Durchsatz mehrerer simultaner Übertragungen reduziert wird. Verschiedene Funktechnologien, wie z.B. IEEE 802.11 (WLAN), stellen orthogonale Kanäle, deren Frequenzspektren nicht überlappen, bereit. Kanaluweisungsalgorithmen versuchen die netzwerkweite Interferenz zu minimieren und somit die Leistung des Netzwerks zu erhöhen, indem orthogonale Kanäle für potentiell interferierende Verbindungen verwendet werden. Verteilte Kanaluweisungsalgorithmen treffen ihre Entscheidungen dabei basierend auf lokal vorhandenen Informationen und kommen somit ohne Kontrollinstanz aus. Ein wesentlicher Trade-off zwischen Konnektivität und Interferenz entsteht aus dem Sachverhalt, dass zwei Netzwerkteilnehmer nur dann in der Lage sind, miteinander zu kommunizieren, wenn sie auf mindestens einer Netzwerkschnittstelle

einen gemeinsamen Kanal verwenden. Ziel der Arbeit ist es, das Potential von Kanalzuweisungsalgorithmen für die Leistungssteigerung in großen Maschennetzwerken zu evaluieren.

1.1 Distributed Greedy Algorithm

Der in dieser Arbeit implementierte *Distributed Greedy Algorithm* [SGDC08] erfasst zunächst die in seiner m -hop Nachbarschaft verwendeten Kanäle. Mittels einer Kostenfunktion, die die spektrale Distanz zweier Kanäle auf ein reelles Kostenmaß abbildet, kann der Algorithmus die Gesamtinterferenz seiner m -hop Nachbarschaft berechnen. Jede Instanz des Algorithmus weist ihren Schnittstellen nun Kanäle so zu, dass die Gesamtinterferenz in der Nachbarschaft minimiert wird. Jegliche Änderung der Kanalzuweisung wird den m -hop Nachbarn mitgeteilt, damit diese wiederum ihre Kanalzuweisung anpassen können. Dieses Vorgehen wird so lange wiederholt, bis keine Instanz mehr in der Lage ist, die Gesamtinterferenz zu minimieren. Die Konvergenz des Verfahrens wird mit einem verteilten Sperrprotokoll sichergestellt. Eine grundlegende Konnektivität wird durch die Verwendung eines dedizierten Kanals, der von allen Instanzen auf einer Netzwerkschnittstelle verwendet wird, aufrecht erhalten.

2 Methodik

Eine universell anerkannte Methodik für die experimentelle Evaluation von Kanalzuweisungsalgorithmen existiert bislang nicht. Bekannte Metriken und Szenarien für die Evaluation von Netzwerken (wie z.B. Sättigungsdurchsatz und Latenz) liefern zwar Aussagen über die Leistungsfähigkeit des Netzwerks, adressieren aber nicht kanalzuweisungsspezifische Parameter wie Interferenz. Das Fehlen einer solchen Methodik macht den Vergleich verschiedener Evaluationen schwer bis unmöglich. Um diese Lücke zu füllen, wurde eine Methodik für die Leistungsevaluation von Kanalzuweisungsalgorithmen entwickelt und in Form eines Benchmarks implementiert. Der Benchmark umfasst drei neu entwickelte Metriken für Interflussinterferenz, Intraflussinterferenz und Sättigungsdurchsatz. Um die Auswirkung des Algorithmus auf *Interflussinterferenz* zu ermitteln, wird eine Gruppe von m örtlich benachbarten, knotendisjunkten Verbindungen gewählt. Der aggregierte Durchsatz dieser m Verbindungen wird zunächst in m sequentiellen Durchsatzmessungen ermittelt. Dieser Wert stellt den aggregierten Durchsatz in einer interferenzfreien Umgebung dar. Anschließend wird der aggregierte Durchsatz bei höchstmöglicher Interflussinterferenz ermittelt - hierfür wird der Durchsatz aller m Verbindungen simultan gemessen. Das Verhältnis beider Werte gibt Aufschluss darüber, welchen Anteil des möglichen Durchsatzes die erzeugte Kanalzuweisung unter Einfluss von Interflussinterferenz aufrecht erhalten konnte. Die Auswirkung von *Intraflussinterferenz* wird analog ermittelt, allerdings unter Verwendung einer k -langen, zyklenfreien Route. Die Metrik für den *Sättigungsdurchsatz* normalisiert den erreichten Sättigungsdurchsatz gegen den einer Ein-Kanal-Konfiguration. Auf diese Art wird der Vergleich von Ergebnissen testbedübergreifend ermöglicht.

2.1 DES-Testbed

Die Implementierung und Evaluation des Algorithmus wurde auf dem drahtlosen Labor Distributed Embedded Systems Testbed (DES-Testbed) [BGJH11, GBS08a, GBS08b, GJB⁺09] durchgeführt. Das DES-Testbed ist ein Multi-Radio Hybrid Wireless Mesh Network an der Freien Universität Berlin. Der Forschungsschwerpunkt liegt in der ganzheitlichen Erforschung von drahtlosen multi-hop Maschennetzen und Sensornetzen. Das DES-Testbed ist unter anderem Teil der WiSEBED, OPNEX und G-Mesh-Lab Projekte. Das Testbed besteht aus 128 drahtlosen Multi-Radio-Maschenknoten, die in einer irregulären Topologie über mehrere Gebäude auf dem Campus verteilt und mit jeweils drei IEEE 802.11 Netzwerkschnittstellen ausgestattet sind.

3 Ergebnisse

Der implementierte Algorithmus wurde mit einer Ein-Kanal-Konfiguration verglichen, in der netzwerkweit genau ein Kanal für alle Netzwerkschnittstellen verwendet wurde. Es wurden insgesamt 40 Replikationen des Benchmarks mit jeweils 106 Knoten durchgeführt. Der evaluierte Algorithmus zeigte eine Leistungssteigerung um etwa 50% bei dem Intraflussinterferenzszenario. Im Interflussinterferenzszenario wurde eine Leistungssteigerung von etwa 60% erreicht. Die Ergebnisse des Sättigungsszenarios zeigten, dass der Durchsatz der einzelnen Verbindungen bis hin zur Sättigung des Netzwerks stabil gehalten werden konnte. Die Ein-Kanal-Konfiguration zeigte hier hingegen signifikante Durchsatzeinbußen. Der Gesamtdurchsatz des Netzwerks war im Sättigungszustand um etwa 20% höher. Ein parallel betriebenes IEEE 802.11 Netzwerk des zentralen Rechnerbetriebs der Freien Universität Berlin sorgte für externe Interferenz im 2 GHz Band. Im 2 GHz Band erreichten die Verbindungen etwa die Hälfte des Durchsatzes der Verbindungen im 5 GHz Band.

4 Schlussfolgerung

Es konnte gezeigt werden, dass der evaluierte Kanalzuweisungsalgorithmus, in allen von dem Benchmark erfassten Szenarien, erhebliche Leistungssteigerungen gegenüber einer Ein-Kanal-Konfiguration erzielen konnte. Die Auswirkung externer Interferenz machte sich beim Sättigungsdurchsatz bemerkbar. Hierdurch wird deutlich, dass eine Notwendigkeit für Algorithmen besteht, die externe Interferenz beachten. Dies wird insbesondere von steigender Bedeutung sein, da in naher Zukunft mit der Entwicklung des Internets der Dinge eine steigende Dichte an drahtlosen Maschennetzen in städtischen Gebieten erwartet werden kann. Der Benchmark ist leicht zu erweitern und lässt sich auf weitere Testbeds portieren. Die Arbeit liefert eine Grundlage, weitere Kanalzuweisungsalgorithmen zu implementieren und diese mit dem entwickelten Benchmark zu evaluieren.

Literatur

- [BGJH11] Bastian Blywis, Mesut Günes, Felix Juraschek und Oliver Hahm. Properties and Topology of the DES-Testbed, March 2011.
- [GBJS08a] Mesut Güneş, Bastian Blywis, Felix Juraschek und Philipp Schmidt. Concept and Design of the Hybrid Distributed Embedded Systems Testbed. Bericht TR-B-08-10, Freie Universität Berlin, 2008.
- [GBJS08b] Mesut Güneş, Bastian Blywis, Felix Juraschek und Philipp Schmidt. Practical Issues of Implementing a Hybrid Multi-NIC Wireless Mesh-Network. (TR-B-08-11), 2008.
- [GJB⁺09] Mesut Güneş, Felix Juraschek, Bastian Blywis, Qasim Mushtaq und Jochen Schiller. A Testbed for Next Generation Wireless Networks Research. *Special Issue PIK on Mobile Ad-hoc Networks*, 34(4), 2009.
- [GK00] Piyush Gupta und P. R. Kumar. The Capacity of Wireless Networks. *Information Theory, IEEE Transactions on*, 46 Issue:2:388 – 404, 03 2000.
- [JPPQ03] Kamal Jain, Jitendra Padhye, Venkata N. Padmanabhan und Lili Qiu. Impact of interference on multi-hop wireless network performance. In *Proceedings of the 9th annual international conference on Mobile computing and networking*, MobiCom 03, Seiten 66–80, New York, NY, USA, 2003. ACM.
- [Phi10] Matthias Philipp. A Framework for Distributed Channel Assignment in Wireless Mesh Networks. Diplomarbeit, Freie Universität Berlin, 2010.
- [SGDC08] Anand Prabhu Subramanian, Himanshu Gupta, Samir R. Das und Jing Cao. Minimum Interference Channel Assignment in Multiradio Wireless Mesh Networks. *IEEE Transactions on Mobile Computing*, 7:1459–1473, 2008.

Continuous Diameter Measurement of Vascular Structures in Ultrasound Video Sequences

Matthias Bremser

matthias@bremser.eu

Abstract: Diameter measurements of vessel structures are of interest for a number of cardiovascular examinations. To support manual analysis of duplex ultrasound (US) images of human vessels and to improve the measurement accuracy and reproducibility, a semi-automatic approach has been developed which applies image processing techniques to compute reliably the vessel diameter.

The proposed approach presents an interactive tool for measuring vessel diameters from US image sequences. A first derivative of a Gaussian (DoG) is applied for vessel wall detection followed up by a skeletonization step, and computation of horizontal line segments. Within a final classification step those line segments which show the highest likelihood are selected. The classification score is computed based on length, edge strength, and linearity information. The overall approach has been evaluated at several sample sequences showing results outperforming „manual” measurements. It is currently applied in physiological studies.

1 Introduction

The function and structure of the human vessel system is a fundamental factor regarding cardiovascular health. Both elements are highly adaptable to its environmental conditions, for example micro-gravity. Vascular ultrasound scanning is a proven method to assess human vessel structure and function [BDRe05, GSEe10].

US assessments are usually performed on computer aided basis. Ordinary US devices are capable of analyzing the diameter of vascular structures under certain requirements during acquisition. For some in particular scientific applications however, repeated analysis of very precise measurements are necessary, to determine vascular changes e.g. during exercise conditions. Therefore, the proposed approach has been developed to extend the basic capabilities of standard video output out of different standard US devices by semi-automatic stand-alone post-acquisition analysis procedures.

2 Materials and Methods

During the development process US data of several areas of healthy adult human bodies (male + female; arteria femoralis superficialis, arteria brachialis) at a resolution of

330×290, 8-bit grayscale, have been used. US data are preprocessed by automatic cropping of the central area of interest from the raw image sequence. Figure 1 shows a sample B-mode US image and a schematic illustration of the notation of near wall, far wall, lumen, and diameter. US images in general suffer from noise and artifacts and may differ strongly

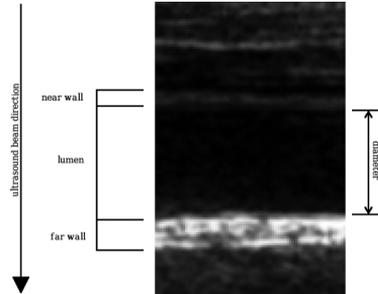


Figure 1: Sample US image area (enlarged) of a healthy adult human including a schematic illustration of the used notation and setup.

in quality from subject to subject and device to device. Depending on the individual habitus, higher amounts of tissue surrounding the blood vessel usually leads to poor quality of the US images. Most noted kind of noise is the so called speckle noise [GAGHe97, DMGe07, LPCe05]. Blood backscattering the US waves, strong reflecting walls, or tissue absorbing the US signal are other noted sources of noise [DMGe07, GAGHe97].

Principle steps of the approach are to first apply edge detection methods, then identify both edges on the vessel walls and finally measure the diameter between the two. A region of interest (ROI) has to be defined where diameter measurement should be executed. This ROI selection has to be done in a way, that it is appropriate for the entire set of frames under investigation. In the next step the image sequences are subsequently analyzed frame-by-frame. To detect the vessel walls, a longitudinal first derivative of a Gaussian filter is applied. It has been shown, that a DoG parameterization of $\sigma_{xx} \leftarrow 10$ and $\sigma_{yy} \leftarrow 5$ for $(7\sigma_{xx} \times 7\sigma_{yy})$ kernel sizes result in meaningful values for the provided data sets. In the next process step the edge images are skeletonized. Adjacent edge elements are computed to horizontal „Line” structures (horizontal lines are defined quite generously (see figure 2)). Each line is considered as a candidate for a vessel wall. During a classification step the most likely line segment pair is determined as near and far vessel wall representatives. Figure 2 shows typical pattern of computed line candidates (all horizontal lines) and the final selection of the vessel wall representatives (dashed lines).

The classification of the line segments as edges of the vessel walls is computed on the basis of length, edge strength, linear regression coefficient, and distance to the vertical center of the ROI, for each line. Length, edge strength and linear regression coefficient are mandatory for the classification process while the distance of the line to the vertical center of the ROI is weighted by two tolerance factors α and β have to be given. Having the two vessel walls identified as straight lines in the ROI, the diameter can be computed by the vertical distances between both lines. This approach assumes, that the vessel is horizontal

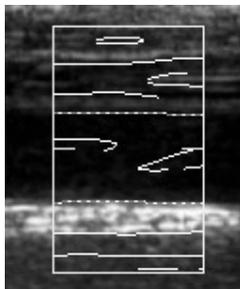


Figure 2: Computed line segments and selected near and far vessel wall candidates (dashed) of arteria femoralis superficialis.

Deviation	mm	Std.
Intra-tester manual	0.209	0.232
Intra-tester automatic	0.130	0.251
Inter-tester manual	0.173	0.202
Inter-tester automatic	0.014	0.016

Table 1: Intra-tester and inter-tester deviation (manually and automatic): For intra-tester absolute difference between 2 measurements; for inter-tester mean of value of 2 measurements per frame and tester respectively; Average deviation over all testers and frames.

and not tilted more than 5° in the frame. This assumption is evaluated by the system for every single frame before the diameter is calculated. The limitation is considered as a compromise between simplicity of the algorithm and given user requirements.

3 Results

An evaluation of the approach has been conducted at DLR aerospace center. Manual measurements of image sequences of 6 subjects performed by 2 examiners, summing up to 15,608 frames, have been compared to diameter measurements obtained by the approach as well operated by the both raters. All measurements were obtained twice. Manual measurements were obtained on the basis of a reduced set of frames.

The chart in figure 3 shows the measurement results of one sample examined sequence. The measurement results computed by the approach, on average, lie between both results of manual measurement, concluding a lower inter-individual deviation. For the evaluations conducted, the ground truth of the vessel diameter should be considered as rather constant. However due to the current of the blood flow within a living subject the measurements may vary naturally in a range of $\pm 10\%$. Variations of this magnitude, however, may not occur within the range of preselected frames as we have seen it during purely manual measurements.



Figure 3: Measurement results of a sample sequence obtained by both examiners, manually and by the proposed approach. The abscissa shows frame numbers. The ordinate axis shows measured lumen diameter in mm. The values shown, are the mean of both iterations with the support of the approach and manual reference. Offset of start and stop frames were defined by the operators due to the quality of the video sequences. It is clearly visible that the diameter variations of the manual references vary more than the approach supported ones.

The evaluation shows, that the results for semi-automatic diameter detection vary less or similar in comparison to the manual results. Comparison to manual measurements have been the basis for evaluation of the results. Not only the coherence of the manual and semi-automatic results, was considered as convincing enough for a practical application.

References

- [BDR05] MW Bleeker, PC De Groot, GA Rongen, and et al. Vascular adaptation to deconditioning and the effect of an exercise countermeasure: results of the berlin bed rest study. *J Appl Physiol*, 99:1293–1300, 2005.
- [DMGe07] S Delsanto, F Molinari, P Giustetto, and et al. Characterization of a Completely User-Independent Algorithm for Carotid Artery Segmentation in 2-D Ultrasound Images. *IEEE Trans Instrum Meas*, 56:1265–1274, 2007.
- [GAGHe97] T Gustavsson, R Abu-Gharbieh, G Hamarneh, and et al. Implementation and comparison of four different boundary detection algorithms for quantitative ultrasonic measurements of the human carotid artery. In *Comput Cardiol*, pages 69–72, Lund, Sweden, 1997. IEEE.
- [GSEe10] DJ Green, A Swart, A Exterkate, and et al. Impact of age, sex and exercise on brachial and popliteal artery remodelling in humans. *Atherosclerosis*, 210:525–530, 2010.
- [LPCe05] CP Loizou, CS Pattichis, Christodoulou, and et al. Comparative evaluation of despeckle filtering in ultrasound imaging of the carotid artery. *IEEE Trans Ultrason Ferroelectr Freq Control*, 52:1653–1669, 2005.

Development of a phylogenetic transcriptome atlas of *Arabidopsis thaliana* embryo development

Hajk-Georg Drost

Thesis Supervisors : Prof. Dr. Ivo Große and Dr. Marcel Quint
Martin-Luther-University Halle-Wittenberg
hajk-georg.drost@student.uni-halle.de

Abstract: Linking phylogeny and ontogeny on the basis of available transcriptome data for *Arabidopsis thaliana* embryo development I argue that during the embryogenesis of plant and animal organisms common evolutionary patterns determine a potent development of an organism. To process large data sets necessary for broad analyses the programming language *R* has been used.

1 Introduction

Carl-Ernst von Baer (1791-1876) a comparative embryologist discovered that at a specific stage of embryogenesis embryos of different vertebrates seem to pass through a stage at which their morphological appearance is very similar [A.97]. This conspicuous stage was first proposed by K. Sander as the *phylotypic stage* [K.83]. Today the *phylotypic stage* is defined as “*the stage of development at which all major parts are represented in their final positions as undifferentiated cell condensations, or the stage at which all members of the phylum show the maximum degree of similarity.*”[W⁺93].

Reflecting the fact that the *phylotypic stage* is in the middle of embryogenesis and not at the beginning, which could be thought of to be the stage at which the evolutionarily oldest genes are transcribed, shows the complexity of linking embryogenesis as part of ontogeny and phylogeny as part of evolutionary processes. The chronology of vertebrate embryogenesis and the corresponding morphologies between different phyla can be pictured as a developmental hourglass. This hourglass model has been introduced to illustrate the morphological and molecular behavior of embryos during early, mid, and late embryogenesis. It demonstrates that the similarities of morphology and molecular processes between phyla during early and late embryogenesis are more resistant against evolutionary changes than the highly conserved middle stage [A.96].

So far, the discussion on to what extent the hourglass model reflects the embryogenetic *phylotypic stage* is confined to vertebrates and echinoderms. Following the evidence that there exists a correlation between ontogeny and phylogeny based on molecular studies one could ask if those phyla are only exceptions in the vast diversity of organisms or if a logical pattern of evolution is found in all organisms which pass through embryogenesis. Comparative studies between plants and animals have shown that aside from the fact that

plants and animals developed independently they share logical patterns of development and metabolism [M.02]. Assuming that plants and animals share similar logical processes which evolved separately, it is justified to ask whether those common logical patterns exist in plant and animal embryogenesis. Answering this question could lead to a basic understanding of how development within all multicellular species might have evolved.

In my thesis, I mainly focused on studying the existence of a *phylotypic stage* in *Arabidopsis thaliana* (thale cress). This approach is based on my assumption that plants and animals share common logical patterns during embryogenesis. I performed broad analyses based on the work done by Domazet-Loso and Tautz and its applicability on plant organisms to support my hypothesis [TD10]. I used the *phylostratum* [DLTD07] of *A. thaliana* provided by A. Gabel [A.11] and self-implemented *R* [R D10] functions to process large data sets provided by Xiang et. al. [D⁺11] to look for supporting facts for my hypothesis. *R* is a specialized programming language for statistical computing, data analysis and graphic handling. To process large transcriptome data sets (up to 16188 x 61 dimensional) and data base queries of large data bases, *R* was providing powerful functions to quickly implement the necessary pipelines for data analyses. An example for a self-implemented *R* function is presented in *Figure 1*. It shows the *R* implementation of the *Transcriptome age index* calculating function [TD10].

```
# Transcriptome age index

TAI <- function(expression_data, phylostratum){
  i <- 1;
  TAI_s <- 1;
  for(i in 1:dim(expression_data)[2])
    TAI_s[i] <- sum(phylostratum*expression_data[1:dim(expression_data)[1],i]) /
    sum(expression_data[1:dim(expression_data)[1],i]);
  return (TAI_s);
}
```

Figure 1: *R* implementation of the *Transcriptome age index*.

2 Transcriptome age index

The origin of a gene has been defined as the *BLAST* hit which has been found in the oldest phylostratum. The method searches for the first appearance of a protein and ranks it based on its first appearance in a given phylostratum. Multi-domain proteins had been left out [DLTD07]. As a result, each gene within a given genome has a phylogenetic rank based on its assignment by the *phylostratigraphic* approach [DLTD07]. Using the genes selected by *phylostratigraphy* and linking them to their gene expression values determined by microarray analyses the age of a transcriptome could be measured during defined stages. To combine the phylogenetic and transcriptome approach Domazet-Loso and Tautz [TD10] modeled a *Transcriptome Age Index* (TAI). Domazet-Loso and Tautz introduced this age index to combine the ranked ages of genes with their expression intensity during a spe-

cific ontogenetic stage. The expression levels represented the weights in the TAI formula which then has been introduced as weighted mean to determine the age (position) of a transcriptome based on the given phylogenetic tree. *Figure 2* shows the TAI results for the *A. thaliana* embryo development. The TAI function of *A. thaliana* embryo development is also following the embryonic hourglass pattern which had been shown for animal model organisms [A^+ 10, TD10].

$$TAI_s = \frac{\sum_{i=1}^n e_{is} \cdot ps_i}{\sum_{i=1}^n e_{is}} \quad (1)$$

where e_{is} denotes the gene expression value of gene i , in stage s , ps_i denotes an integer value representing the phylostratum of gene i , and n denotes the total number of genes available in the data set. A small ps_i value describes an old phylostratum and a high ps_i value a younger phylostratum [TD10].

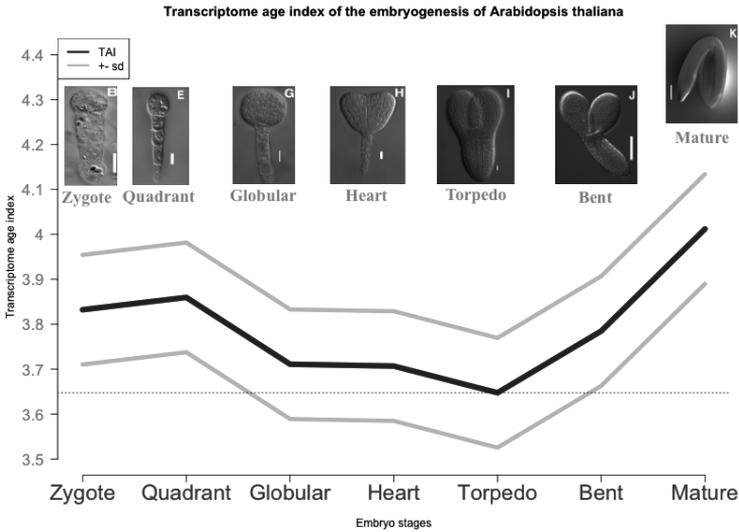


Figure 2: Transcriptome age index of *A. thaliana*. The x-axis shows the embryo stages of *A. thaliana*. The y-axis shows the TAI value. The grey colored function represents the \pm standard deviation of the data. The global minimum of the TAI curve is located at the *Torpedo* stage returning a TAI value of 3.64. This result indicates that the *phylotypic stage* in *A. thaliana* embryo development lies at the *Torpedo* stage. Pictures of *A. thaliana* embryo development have been taken from [D⁺11].

3 Conclusion

The existence of a *phylotypic stage* had never been discussed in the plant world. In my thesis I tried to approach this issue based on the TAI approach for several stages covering

A. thaliana embryo development. Another aspect was the comparability of evolutionary processes during animal and plant embryogenesis. As a result, I could show for *A. thaliana* that TAI calculations during defined embryo stages returned a hourglass shaped pattern with a global minimum at the *Torpedo* stage. This result indicated that this stage expresses the largest amount of evolutionary conserved genes. This assumption had been supported by analyses on highly expressed top level genes which had been shown to be causing the shape of the hourglass pattern and the degree of conservation during the *Torpedo* stage. I was able to show that those top level genes yielded an overrepresentation of old genes responsible for highly conserved housekeeping activities during the embryogenesis of *A. thaliana*. Similar results had been achieved for *Danio rerio* (zebrafish). Extracting top level genes from the data set, the hourglass pattern formed by TAI calculations disappeared and had been leading to an overall increase of TAI values.

Summarizing these results, I was able to show that in animals and plants alike the existence of highly conserved embryo stages might be responsible for a potent development of an organism.

References

- [A.96] Raff R. A. *The Shape of Life*. The University of Chicago Press, Chicago, 1996.
- [A.97] Müller W. A. *Developmental Biology*. Springer-Verlag, New York, 1997.
- [A⁺10] Kalinka A. et al. Gene expression divergence recapitulates the developmental hourglass model. *Nature*, 468:811–814, December 2010.
- [A.11] Gabel A. Phylostratigraphy analysis of the Arabidopsis thaliana genome. *Bachelor's thesis*, 2011.
- [D⁺11] Xiang D. et al. Genome-wide analysis reveals gene expression and metabolic network dynamics during embryo development in Arabidopsis. *Plant Physiology Preview*, 156(1):346–356, March 2011.
- [DLTD07] Brajkovic J. Domazet-Loso T. and Tautz D. A phylostratigraphy approach to uncover the genomic history of major adaptations in metazoan lineages. *TRENDS in Genetics*, 23(11):533–539, Oktober 2007.
- [K.83] Sander K. *The evolution of patterning mechanisms: Gleanings from insect embryogenesis and spermatogenesis*. Cambridge University Press, New York, 1983.
- [M.02] Meyerowitz E. M. Plants Compared to Animals: The Broadest Comparative Study of Development. *Science*, 295:1482–1485, February 2002.
- [R D10] R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2010. ISBN 3-900051-07-0.
- [TD10] Domazet-Loso T. and Tautz D. A phylogenetically based transcriptome age index mirrors ontogenetic divergence patterns. *Nature*, 468:815–818, December 2010.
- [W⁺93] Slack J. M. W. et al. The zootype and the phylotypic stage. *Nature*, 361:490–492, February 1993.

A phylostratigraphic analysis of the *Arabidopsis thaliana* genome

Alexander Gabel

Alexander.Gabel@student.uni-halle.de

Supervisors: Prof. Dr. Ivo Große and Dr. Marcel Quint

Abstract: The aim of this work was the development and implementation of a pipeline implemented in Perl to determine the evolutionary age of each *A.thaliana* gene on the basis of BLAST comparisons against its reconstructed tree of life. Depending on these results of similarity searches each gene should assign to a specific category, which represent its evolutionary age.

1 Introduction

In October 2007 Domazet-Lošo and Tautz [TJD07] published the introduction of a new method to study macroevolutionary trends. They called this method phylostratigraphy. This method should illustrate an additional level of analysis compared to the traditional approaches to study macroevolutionary trends for example with fossil analysis or comparative morphological analyses. The phylostratigraphy allows to trace the evolutionary origin of each gene in the genome of the organism of interest. It assigns each gene in the genome to a specific category in the phylogeny that defines its evolutionary age. This procedure was made possible with the availability of genome sequences and associated data from an increasing diversity of taxa [TJD07]. The whole procedure is based on similarity searches. Those were done by BLAST [FWW⁺90, FLA⁺97] to find homologous sequences in the organisms which represent the internodes of the phylogeny.

In their first phylostratigraphic approach Domazet-Lošo and Tautz used this method to determine the evolutionary age of each gene of *Drosophila melanogaster*. They could show *that grouping genes by their phylogenetic origin can uncover footprints of important adaptive events in evolution* [TJD07].

Additional phylostratigraphic studies [TD08, TD10a, TD10b, TD11] mainly concentrated only on animal organisms. This bachelor's thesis uses the method of phylostratigraphy to trace the evolutionary origin of each gene of the *Arabidopsis thaliana* genome. The aim of this work is the creation of a pipeline to determine the evolutionary age of each *A.thaliana* gene on the basis of phylostratigraphy. The resulting phylostratigraphic map shall give indications of important events in the evolution of plants and especially in the evolution of *A.thaliana*.

2 Methods

The phylostratigraphy is a general approach to trace evolutionary innovations using data from genome projects. The aim of the genomic phylostratigraphy in this work was to assign each gene in the *A.thaliana* genome to a specific category that defines its evolutionary age. Such a category called phylostratum and is a set of genes from the organism of interest that coalesce to founder genes having common phylogenetic origin [TJD07]. Thereby, a founder gene is the first emergence of a gene forming the basis of a new gene lineage or gene family. The emergence of founder genes in this analysis is based on two main models: first gene duplication followed by divergence and, second de-novo evolution of non-coding sequences. To identify the new evolved genes as founder genes they may not have any sequence similarity to their ancestors.

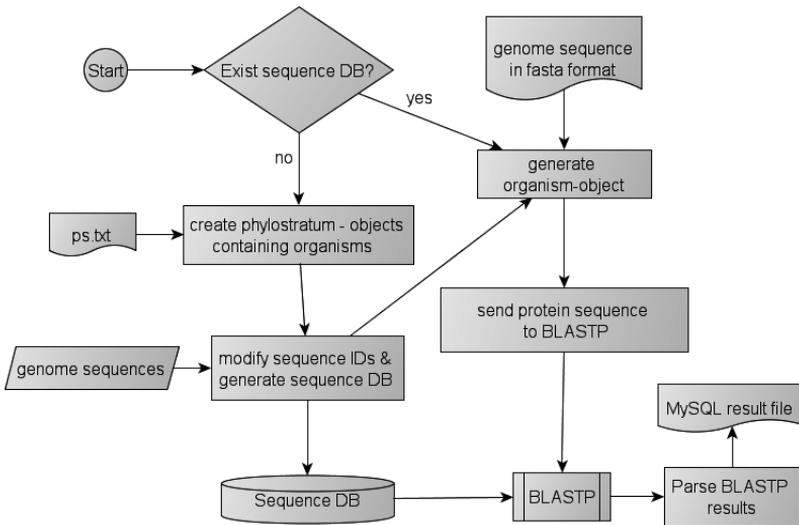


Figure 1: Flowchart of the pipeline for phylostratigraphic analysis. This flowchart shows the work of the main script, which took each genome sequence and generated with the information of their phylostratum (ps.txt) and organism name the BLAST sequence database. The next step was to load the *A.thaliana* genome and send each protein sequence to BLAST which compared the sequence against the previously created database. The last step was to parse the BLAST results into a MySQL-file, which contains all hit sequences to a given *A.thaliana* query sequence.

For the assignment of each gene into a specific phylostratum it is necessary to trace its evolutionary origin. This was done by using BLAST comparisons of each *A.thaliana* gene against a sequence database in which the internodes of the tree of life of *A.thaliana* were represented by one or more fully annotated genomes. For similarity searches I used an $1e-5$ E-value cutoff. All genes were then distributed into 14 phylostrata according to the emergence of their founders in the phylogeny.

The pipeline for the phylostratigraphic analysis shown in Figure 1 was implemented in

Perl. Using different BioPerl-packages [EDBK02] the script created a sequence database corresponding to the tree of life of *A.thaliana* and performed the BLAST similarity searches of each *A.thaliana* gene against the sequence database. Afterwards, all BLAST results were stored in a MySQL-database. Accessing this database each gene was sorted into a phylostratum according to the phylogenetically most distant genome in which homologs were identified.

Also, for each phylostratum the number of founder genes was calculated. For that I used the following equation [TJD07],

$$G_f = \sum_{i=1}^G \frac{1}{H_i} , \tag{1}$$

where G denotes the number of genes in one phylostratum, H_i denotes the obtained number of hits of every gene and G_f is the number of founder genes, while $1 \leq H \leq G$. When all genes in the phylostratum are related, G_f has the lowest value one. Otherwise, if all genes in the phylostratum are founder genes, the G_f is equal to G .

With the help of the number of founder genes the amount of evolutionary novelties in each phylostratum could be represented. Additionally, for each internode in the phylogeny I took the divergence times from *timetree.org* [BJS06] and estimated the fixation rate of founder genes by dividing the number of founder genes by the divergence time between two phylostrata. This fixation rate shows how fast new genes evolved and also if they were locked up in a certain pathway or disappeared in the evolution.

3 Results

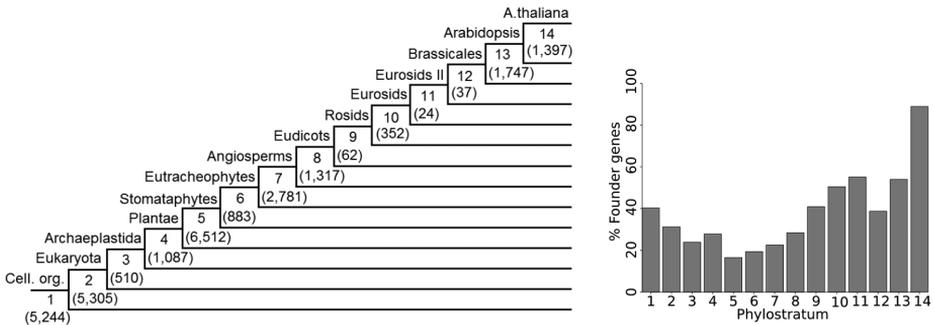


Figure 2: Complete phylostratigraphy with *A.thaliana*. Numbers in the parenthesis denote the number of genes that were ordered in a specific phylostratum. Obviously the largest number of genes has arisen in the lower phylostrata 1 and 2. The internodes in this tree represent the different taxa which lead to *A.thaliana*.

After performing the BLAST comparisons, the *A.thaliana* genes were distributed into the 14 phylostrata. While looking at Figure 2 one can see that a majority of genes were sorted

into the older phylostrata. One can assume that more than half of all genes are older than 565 million years in respect to the given divergence times [BJS06]. Also there are a lot of genes in phylostrata 1 and 2. This indicates that on one side many genes of *A.thaliana* evolved long time before plants developed and on the other side that these genes are also present in other non-plant eukaryotes. This could underpin their evolutionary importance.

The phylostratigraphic analysis of *A.thaliana* shows that many new genes evolved in the oldest phylostrata and that these genes were fixed over a long period of time. In contrast, the youngest phylostrata showed also many new evolved genes and also many new evolved founder genes. This indicates that the development of new genes is very fast but it shows also that many of these genes would disappear in the ongoing evolution if they were not fixed in a particular pathway.

References

- [BJS06] Hedges S. B., Dudley J., and Kumar S. TimeTree: A public knowledge-base of divergence times among organisms. *Bioinformatics*, 22:2971–2972, 2006.
- [EDBK02] Stajich J. E., Block D., and et al. Boulez K. The Bioperl Toolkit: Perl modules for the life sciences. *Genome Research*, 12:1611–1618, October 2002.
- [FLA⁺97] Altschul St. F., Madden T. L., Schäffer A. A., Zang J., Zang Z., Miller, and Lipman D. J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, 25:3389–3402, 1997.
- [FWW⁺90] Altschul St. F., Gish W., Miller W., Myers E. W., and Lipman D. J. Basic Local Alignment Search Tool. *Journal of Molecular Biology*, 215:403–410, 1990.
- [TD08] Domazet-Lošo T. and Tautz D. An Ancient Evolutionary Origin of Genes Associated with Human Genetic Diseases. *Molecular Biology and Evolution*, September 2008.
- [TD10a] Domazet-Lošo T. and Tautz D. A phylogenetically based transcriptome age index mirrors ontogenetic divergence patterns. *Nature*, 468, December 2010.
- [TD10b] Domazet-Lošo T. and Tautz D. Phylostratigraphic tracking of cancer genes suggests a link to the emergence of multicellularity in metazoan. *BioMed Central Biology*, 8, 2010.
- [TD11] Domazet-Lošo T. and Tautz D. The evolutionary origin of orphan genes. *Nature Reviews Genetics*, August 2011.
- [TJD07] Domazet-Lošo T., Brajković J., and Tautz D. A phylostratigraphic approach to uncover the genomic history of major adaptations in metazoan lineages. *Trends in Genetics*, 23, October 2007.

Vereinfachte 3D-Rekonstruktion aus Sequenzen unkalibrierter Bilder

Sergey Cheremukhin

FernUniversität in Hagen
Mensch-Computer-Interaktion
Sergey.Cheremukhin@FernUni-Hagen.de

Art der Arbeit: Studienarbeit

Betreuer der Arbeit: Dr. Klaus Häming, Prof. Dr. Gabriele Peters

Abstract: Diese Arbeit stellt einen neuen Ansatz vor, eine Rekonstruktion von 3D-Geometrie aus einfachen Fotos durchzuführen. Diese Fotos wurden von Kameras aufgenommen, deren interne Abbildungsparameter unbekannt sind (d.h. nicht kalibriert sind). Das Grundelement dieses Ansatzes sind Bildpaare, im Gegensatz zu den im unkalibrierten Fall sonst üblichen Bildtripel. Daraus ergibt sich eine Vereinfachung des Rekonstruktionsprozesses. Es wird gezeigt, dass bereits Korrespondenzen zwischen Bildpaaren ausreichen, um von diesen zunächst hinreichend genau auf die internen Kameraparameter zu schließen. Anschließend wird das Zusammenfügen der aus einer Vielzahl an Bildpaaren gewonnenen Teilrekonstruktionen beschrieben.

1 Einführung

Die Rekonstruktion von 3D-Geometrie aus einfachen Fotos verspricht einen stark vereinfachten Zugang zur Generierung von 3D-Modellen für jedefrau und jedermann. Die Aufnahmen können dabei von einfachen, unkalibrierten Kameras stammen.

Die grundlegende Vorgehensweise bei der Rekonstruktion macht sich die geometrischen Beziehungen zunutze, die zwischen Bildpunkten, 3D-Szene und den verwendeten Kameras gelten. Wie in [HZ03] beschrieben, lässt sich daraus ein Modell ableiten, dessen Parameter sich schrittweise bestimmen lassen. Eine konkrete Umsetzung dieser Vorgehensweise wurde bspw. in [PH10] vorgestellt. Diese Umsetzung diente als Ausgangspunkt für das hier beschriebene System.

Das verwendete Kameramodell ist eine projektive Kamera der Form $\mathbf{P} = \mathbf{KR}[\mathbf{I}_{3 \times 3} | -\mathbf{C}]$.

Die Matrix $\mathbf{K} := \begin{bmatrix} f & s & x_0 \\ & af & y_0 \\ & & 1 \end{bmatrix}$ ist dabei die sogenannte Kalibrierungsmatrix und stellt

die internen Abbildungsparameter dar. Dabei ist f die Brennweite (focal length), a das Seitenverhältnis (aspect ratio), s ein Scherungsparameter, sowie $(x_0 \ y_0)^T$ der Bildmittelpunkt. \mathbf{R} ist eine 3×3 Drehmatrix und \mathbf{C} ist das Kamerazentrum.

Die Eingabedaten sind lediglich Bildmerkmale, die wir mittels eines Harris-Merkmal-detektors [HS88] in einem Scale-Space-Rahmenwerk [Lin94] ermitteln. Um Beziehungen zwischen unterschiedlichen Bildern herzustellen, werden Merkmale, die das gleiche Szenenelement abbilden zu Korrespondenzen zusammengefasst. Dazu beschreiben wir die Merkmale mittels SIFT-Deskriptoren [Low04] und gleichen diese in einem kd-Baum mit Best-Bin-First-Heuristik [BL97] ab. Die dabei unvermeidbaren Falschzuordnungen werden in einer robusten Kombination aus LO-RANSAC und PROSAC [HP10] herausgefiltert.

2 Eigene Entwicklungen

Wir verfolgen den Ansatz, aus jedem Bildpaar eine metrische Rekonstruktion zu gewinnen, also die zugehörigen Kameras automatisch zu kalibrieren. Der übliche Ansatz wäre, nach der Berechnung von Bildpaaren zunächst zu Bildtripel [SSG⁺10] überzugehen. Wir sparen also einen Schritt im Rekonstruktionsprozess ein. Jedoch hat die Kalibrierung von Bildpaaren allein auf der Grundlage von Korrespondenzen im allgemeinen keine eindeutige Lösung. Insbesondere ergeben sich aus der rein mathematisch vorhandenen Möglichkeit, dass die Szenenstruktur, die zu einem abgebildeten Merkmal gehört, sowohl vor als auch hinter einer Kamera liegen kann, bereits vier mögliche Konstellationen.

Im ersten Schritt werden also Bildmerkmale in beiden Bildern eines Bildpaares gefunden und zu Korrespondenzen zusammengefasst. Anschließend werden fehlerhafte Korrespondenzen herausgefiltert, indem nur solche akzeptiert werden, die tatsächlich zu ein und derselben Kamerakonstellation gehören können. Eine solche Konstellation wird zu diesem Zeitpunkt durch eine Fundamentalmatrix beschrieben, die aus den Korrespondenzen berechnet werden kann.

Im zweiten Schritt wird der lineare Autokalibrierungsalgorithmus aus [PVG02] verwendet. Es werden zwei konkrete Kameramatrizen gewählt, die zu der Fundamentalmatrix aus dem ersten Schritt passen. Diese werden genutzt, um aus den Korrespondenzen vorläufige 3D-Punkte zu triangulieren. Danach wird eine symmetrische 4×4 Matrix Ω^* mit Rang 3 gesucht, so dass die Gleichung $\mathbf{K}\mathbf{K}^T \sim \mathbf{P}\Omega^*\mathbf{P}^T$ erfüllt ist. Im Fall lediglich zweier Bilder ist die Lösung unterbestimmt, und der Lösungsraum wird durch $\Omega_{\mathbf{a}}^* + \gamma\Omega_{\mathbf{b}}^*$ aufgespannt. Um die Rang-3-Bedingung zu erfüllen, werden nun solche Werte für γ gesucht, die die Determinante von $\Omega_{\mathbf{a}}^* + \gamma\Omega_{\mathbf{b}}^*$ verschwinden lassen [PKv98]. Wir wählen dann aus max. 16 möglichen Lösungen (bis zu 4 Lösungen für γ für jede von 4 möglichen Kamerakonstellationen) die beste aus. Dazu wird aus den möglichen Ω^* jeweils eine 3D-Homographie \mathbf{H} errechnet. Die beste Lösung führt zu einem \mathbf{H} , bei dem sich die meisten transformierten 3D-Punkte $\mathbf{H}^{-1}\mathbf{X}_n$ vor den transformierten Kameras $\mathbf{P}_i\mathbf{H}$ befinden. Dieser Kalibrierungsalgorithmus schlägt fehl, wenn die Bildpunkt-Korrespondenzen nicht allgemein genug, also bspw. lediglich Abbildungen von Punkten ein und derselben Ebene im Raum sind, die Bildpunkt-Korrespondenzen zu viele Falsch-Zuordnungen enthalten oder die internen Abbildungsparameter der Kameras sich zu stark unterscheiden. In diesen Fällen wird versucht, ein Bild im betrachteten Bildpaar durch ein anderes passendes Bild zu ersetzen (sofern genug Bildpunkt-Korrespondenzen gefunden werden können) und er-



Abbildung 1: Rekonstruktionsbeispiel. Gezeigt ist eine Rekonstruktion aus einer Bildfolge, von der einige Bilder links abgebildet sind.

neut eine Rekonstruktion zu finden. Diese Versuche werden solange wiederholt, bis eine Kalibrierung gelingt oder bis keine weiteren passenden Bilder vorhanden sind.

Im dritten Schritt wird das Zusammenfügen der Teilrekonstruktionen zu einer größeren Szene vorbereitet, indem der Einfluss der Kalibrierungsmatrizen entfernt wird. Dazu werden die rekonstruierten 3D-Punkte X_n mit einer geeigneten 3D-Homographie

$$T_i := \begin{bmatrix} R_i & -R_i C_i \\ e_4^T & \end{bmatrix}^{-1} \begin{bmatrix} P_i \\ e_4^T \end{bmatrix}, \text{ mit } e_4^T = (0\ 0\ 0\ 1)$$

transformiert, so dass diese die neue Position $T_i X_n$ annehmen.

Basierend auf diesen transformierten Punkten werden immer zwei Teilrekonstruktionen betrachtet, die sich mindestens eine Ansicht teilen. Zudem müssen sich über diese gemeinsame Ansicht mindestens zwei Korrespondenzenpaare einander zuordnen lassen. Da jeder Korrespondenz durch die oben genannte Triangulierung ein 3D-Punkt zugeordnet ist, bedeutet dies, dass zwei 3D-Punkte gefunden worden sind, die jeweils den selben Punkt der betrachteten Szene repräsentieren. Gibt es mehr als zwei solcher Punktepaare, werden in einem PROSAC-Sampling die besten ausgewählt. Zu jedem 3D-Punktepaar wird ein weiterer, virtueller, Punkt hinzugefügt, so dass der Schwerpunkt dieser Punkttripel im Kamerazentrum liegt. Dadurch wird die später berechnete relative Drehung eine Drehung um das Kamerazentrum. Anschließend wird die relative Orientierung (Drehung, Translation und Skalierung) zwischen diesen zwei Punkttripel berechnet. Für die Berechnung wird der SVD-Algorithmus eingesetzt [AHB87]. Dieser Algorithmus erweist sich in empirischen Versuchen als stabil und wird auch in der Literatur [LEF95] als stabil angesehen.

Im letzten Schritt führen wir eine metrische Variante des Bundle-Adjustment-Algorithmus [TMHF00] aus, um die Rekonstruktion weiter zu optimieren. Wir verwenden hier eine kompakte Parametrisierung durch 7 Parameter für die Kameras (3 für die Drehung, 3 für die Verschiebung und 1 für die Brennweite) und 3 Parameter für die 3D-Punkte. Abbildung 1 zeigt das Ergebnis für eine konkrete Bildsequenz.

Das vorgestellte Verfahren zur 3D-Rekonstruktion eignet sich gut zur schnellen Berechnung von 3D-Geometrie aus Bilddaten. Das inkrementelle Zusammenfügen von kalibrierten Teilrekonstruktionen ermöglicht es zudem, bereits Zwischenergebnisse zu visualisieren sowie den Rekonstruktionsprozess zu verfolgen und interaktiv zu kontrollieren.

Literatur

- [AHB87] K. S. Arun, T. S. Huang und S. D. Blostein. Least-Squares Fitting of Two 3-D Point Sets. *IEEE Trans. Pattern Anal. Mach. Intell.*, 9:698–700, Mai 1987.
- [BL97] Jeffrey S. Beis und David G. Lowe. Shape indexing using approximate nearest-neighbour search in high-dimensional spaces. In *In Proc. IEEE Conf. Comp. Vision Patt. Recog.*, Seiten 1000–1006, 1997.
- [HP10] Klaus Häming und Gabriele Peters. Structure-from-Motion Reconstruction Pipeline - A Survey with Focus on Short Image Sequences. *Kybernetika*, 46(5):926–937, 2010.
- [HS88] Chris Harris und Mike Stephens. A Combined Corner and Edge Detector. In *4th ALVEY Vision Conference*, Seiten 147–151, 1988.
- [HZ03] Richard I. Hartley und Andrew Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521540518, second. Auflage, 2003.
- [LEF95] A. Lorusso, D. W. Eggert und R. B. Fisher. A comparison of four algorithms for estimating 3-D rigid transformations. In *Proceedings of the 1995 British conference on Machine vision (Vol. 1)*, BMVC '95, Seiten 237–246, Surrey, UK, UK, 1995. BMVA Press.
- [Lin94] Tony Lindeberg. Scale-space theory: A basic tool for analysing structures at different scales. *Journal of Applied Statistics*, 21(2):224–270, 1994.
- [Low04] David G. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. *Int. J. Comput. Vision*, 60(2):91–110, 2004.
- [PH10] Gabriele Peters und Klaus Häming. Fast Freehand Acquisition of 3D Objects and their Visualization. *Journal of Communication and Computer*, 7(2-3), 2010.
- [PKv98] Marc Pollefeys, Reinhard Koch und Luc J. van Gool. Self-Calibration and Metric Reconstruction in Spite of Varying and Unknown Internal Camera Parameters. In *ICCV*, Seiten 90–95, 1998.
- [PVG02] Marc Pollefeys, Frank Verbiest und Luc Van Gool. Surviving Dominant Planes in Uncalibrated Structure and Motion Recovery. In *Johansen (Eds.) Computer Vision - ECCV 2002, 7th European Conference on Computer Vision, Lecture Notes in Computer Science, Vol.2351*, Seiten 837–851. Springer-Verlag, 2002.
- [SSG⁺10] Noah Snavely, Ian Simon, Michael Goesele, Richard Szeliski und Steven M Seitz. Scene Reconstruction and Visualization From Community Photo Collections. *Proceedings of the IEEE*, 98(8):1370–1390, 2010.
- [TMHF00] Bill Triggs, Philip F. McLauchlan, Richard I. Hartley und Andrew W. Fitzgibbon. Bundle Adjustment - A Modern Synthesis. In *ICCV '99: Proceedings of the International Workshop on Vision Algorithms*, Seiten 298–372, London, UK, 2000. Springer.

Automatische Pflanzenhöhenermittlung mittels Kinect

Rakulan Subramaniam, André Kunz,

Hochschule Osnabrück
Fakultät Ingenieurwissenschaften und Informatik
rakulan.subramaniam@hs-osnabrueck.de
andre.kunz@hs-osnabrueck.de

GI-Vertrauensdozent: Prof. Dr. Frank M. Thiesing

Betreuer/in der Arbeit: Dr. Hans-Peter Römer, Prof. Dr.-Ing. Bernhard Lang

Abstract: In dem vom BMBF geförderten Forschungsprojekt „Zusammenführung automatisch erfasster Klima- und Pflanzenparameter in einer Bewertungsplattform zur optimierten Steuerung von Gewächshauskulturen“ (kurz KliPa) werden Pflanzenparameter, wie zum Beispiel Höhen, mittels Kinect automatisch erfasst. Die Kinect liefert ein Histogramm mit allen möglichen Höhenwerten, welches anschließend bewertet wird, um die Pflanzenhöhe daraus zu bestimmen.

1 Einleitung

Das moderne Gewächshaus verwendet Klimacomputer (KC) zum Regeln und Dokumentieren der Klimaführung. Derzeit werden, aufgrund subjektiver Eindrücke am Pflanzenbestand, Anpassungen der Klimaparameter manuell vom Gärtner am KC vorgenommen. Eine objektive Erfassung von Pflanzenparametern wäre jedoch sinnvoll, um somit bei einer anschließenden Bewertung bzw. Vergleich zu anderen Kulturzeiträumen bessere Ergebnisse zu erzielen. Aus Sicht der Bildverarbeitung bieten sich hier viele interessante Möglichkeiten. Der Höhenzuwachs ist z.B. eine wichtige Eigenschaft für die Bewertung der Pflanzenentwicklung. Diese manuell zu erfassen ist allerdings sehr arbeitsintensiv und fehleranfällig. Daher werden in der vorliegenden Arbeit die Höhendaten über Histogramme, die eine Microsoft Kinect liefert, mit entsprechenden mathematischen Funktionen ermittelt.

2 Die Kinect

Die Kinect (Abbildung 1) ist eine Hardware, die von Microsoft zusammen mit der israelischen Firma PrimeSense für die Steuerung der Videokonsole Xbox 360 entwickelt wurde. Diese besteht aus einer RGB-Kamera, einem Infrarot-Sensor, einem Infrarot-Projektor und vier 3D-Mikrofonen (multi-microphone-array) [Ma10]. Der Name der Kinect leitet sich von den Begriffen „connect“ und „kinetic“ ab.

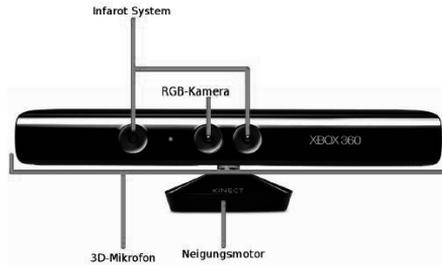


Abbildung 1: Aufbau der Kinect [3D11]

2.1 Technische Details

Die eingebaute RGB-Kamera liefert Farbbilder mit einer Auflösung von 1280x1024 Pixeln. Der Tiefensensor ist ein CMOS-Chip (complimentary metal-oxide semiconductor), hat eine Auflösung von 640x480 Pixeln und liefert Bilder mit einer Frequenz von 30 FPS (Frames per second). Die Genauigkeit liegt bei 2m Abstand vom Objekt zum Sensor, im X- u. Y-Koordinatenbereich bei ca. 3mm und in der Z-Koordinate bei ca. 10mm. Der Sensor arbeitet im Bereich von ca. 0,8 – 3,5m. [MR11]

2.2 Funktionsweise

Der Infrarot-Projektor strahlt ein Pseudo-zufälliges Punktmuster über eine große Fläche im Raum. Diese Punkte werden an den Oberflächen von Objekten reflektiert (Abbildung 2) [3D11] und vom CMOS-Sensor aufgenommen, welcher in der Lage ist, die Tiefe aller Pixel zu berechnen. Die Berechnung erfolgt über die bekannte Methode der Stereotriangulation [JB97]. Um die Berechnung nach dieser Methode durchführen zu können, werden zwei Kameras benötigt. Für die Tiefenmessung müssen korrespondierende Punkte in beiden Bildern gefunden werden. Sobald dies der Fall ist, kann die Disparität zwischen beiden Bildern bestimmt werden. Mit der Disparität kann über die Triangulation die Tiefe der Punkte berechnet werden [Ri10].

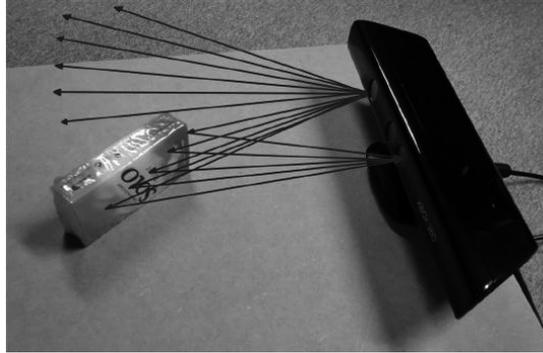


Abbildung 2: Funktionsweise des Infrarot-Systems [3D11]

Hier kommt die Einzigartigkeit der Kinect ins Spiel, welche in der Lage ist, die Berechnung für die Tiefe mit nur einer Kamera durchzuführen. Streng genommen ist der Infrarot-Projektor keine Kamera. Da dieser jedoch ein generiertes Punktmuster projiziert, welches der Kinect bekannt ist, ergeben sich somit zwei Bilder die miteinander verglichen werden können. Über diese Bilder wird letztlich die Höhe berechnet.

3 Aufbau

Im KliPa-Projekt wird die Kinect aktuell im Gewächshaus in der Versuchsanlage in Osnabrück/Haste eingesetzt (Abbildung 3). Hier wird eine komplette Wachstumsphase einer Kultur mit Stereokameras und der Kinect ausgewertet. Dabei befinden sich die Kameras in einer definierten Höhe über den Pflanzen und speichern stündlich Bilder und Höhenhistogramme des Bestandes, welche anschließend automatisch ausgewertet werden.



Abbildung 3: Versuchsaufbau mit Stereokameras und der Kinect im Gewächshaus

4 Höhenberechnung

Wie alle im Projekt eingesetzten Kameras, liefert auch die Kinect nicht „die Höhe“ einer Pflanze, sondern ein Histogramm mit allen möglichen Höhenwerten. Dieses Histogramm muss anschließend bewertet werden, um die Pflanzhöhe zu bestimmen. Im Verlaufe des Projektes stellte sich heraus, dass es nicht trivial ist, eine Höhe festzulegen, die „die Höhe“ der Pflanze ist. Wird über dem Höhenhistogramm einer Aufnahme der Mittelwert gebildet, so ähnelt der Mittelwert einer Glockenkurve (Abbildung 4). Die Höhe der Pflanze liegt dabei stets an der rechten Seite der Glocke. In diesem Fall bei ca. 333 mm einschließlich Topf.

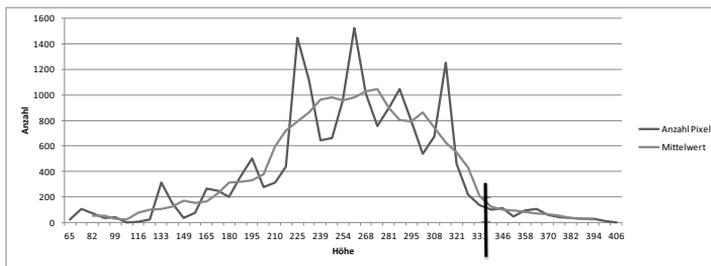


Abbildung 4: Höhenhistogramm[mm] und Mittelwert

Um die Höhe zu ermitteln, wird die zweite Ableitung der Mittelwertkurve gebildet. Diese gibt an, an welchen Punkten die Steigung stark abzufallen beginnt (oben auf der Glockenkurve) und an welchem Punkt die Steigung wieder abflacht (am Ende der Glocke). Anschließend werden die beiden gefundenen Punkte auf der 2. Ableitungskurve durch eine Gerade verbunden. An der Stelle wo diese Gerade die X-Achse schneidet, liegt die Höhe der Pflanze.

5 Zusammenfassung

Im Rahmen dieses Projektes wurde die Kinect zur Höhenermittlung von Pflanzen als günstige Alternative zu teuren 3D-Kameras mit aufwendigen Algorithmen verwendet. In den ersten Tests lieferte diese gute Ergebnisse, die im weiteren Verlauf des Projektes noch ausführlich analysiert werden.

Literaturverzeichnis

- [Ri10] Rivera, K.: Explain how the 3D camera in the Microsoft Kinect sensor operates. 2010.
- [3D11] Reinhardt, M.; Dach, S.; Backasch, M.; Schlegel, C.; Schröder, T.: 3D-Scan 2.0. Technische Universität Bergakademie Freiberg, Freiberg 2011.
- [MR11] Kinect for Windows. Microsoft Research, 2011.
- [Ma10] Marshall, D.: Inside The Kinect Hardware. Microsoft Research, 2010. URL: <http://www.trueachievements.com/n2124/inside-the-kinect-hardware.htm>
- [JB97] Jiang, X.; Bunke H.: Dreidimensionales Computersehen. Springer-Verlag 1997.

Klassifizierung von Anzeigeprozessen in AR-Anwendungen anhand von definierten Paradigmata

David A. Plecher

plecher@in.tum.de

Fachgebiet Augmented Reality

Technische Universität München, Fakultät Informatik

1 Motivation

In den letzten Jahren ist die Anzahl von Anwendungen im Bereich der Augmented Reality (AR) stark gestiegen. Die Einsatzmöglichkeiten sind äußerst vielfältig und umfassen unter anderem Navigationshilfen, Warnsysteme in Autos oder optische Hilfen bei medizinischen Operationen. Jedoch mehren sich dadurch auch die Anforderungen an das User Interface, da es sich im Gegensatz zur Virtual Reality (VR) mit zwei 3D Welten beschäftigt. Dies erfolgt durch die Einbettung virtueller Informationen in die reale Welt z. B. mittels hochentwickelter Displays. Die Frage ist nun, ob neben der Displaytechnik auch die Anzeigemöglichkeiten weiter entwickelt werden oder bereits ausgeschöpft sind. Um dies festzustellen, bedarf es zuerst einer genauen Klassifizierung von Anzeigeprozessen.

2 Definition der Klassen

Auf Basis einer früheren Publikation von Tönnis, Plavšić und Klinker [TPK09] wurden folgende sechs unterschiedliche Klassen der Anzeigeprozessen, die in jeder AR-Anwendung anzutreffen sind, definiert: Tempus, Raum, Registrierung, Perspektive, Referenzierung und Mounting. Grundsätzlich werden hierbei die Anzeigedauer und die Relation des virtuellen Objekts zur realen Umgebung und zum Benutzer untersucht. Jede Klasse lässt sich in zwei oder drei Entitäten gliedern; als Beispiele werden im Folgenden zwei ausgewählte Klassen näher beschrieben. Die temporäre Klasse (I) aus den Entitäten *diskret* und *kontinuierlich*. Während eine diskrete Darstellung ihre Information ereignisabhängig, z.B. das ortsabhängige Einblenden von Pfeilen durch eine Navigationssoftware in ein Head-Up-Display, freigibt, ist eine kontinuierliche Anzeige omnipräsent. Die Registrierung (III) untersucht den Bezug zwischen dem virtuellen Objekt und der realen Umgebung und unterteilt diesen in dreierlei Entitäten. Eine Anzeige ist *unregistriert*, wenn diese keinen räumlichen Bezug zur realen Welt hat. Wenn das dargestellte virtuelle Objekt ein reales Element exakt überlagert und dadurch mit der Wirklichkeit interagiert, wird

es als *registriert* kategorisiert. Sollte es zusätzlich noch in der genauen räumlichen Tiefe angezeigt werden, erfüllt es die Voraussetzungen der *kontakt-analogen* Registrierung.

Durch eine Kombination der jeweils zutreffenden Entitäten lässt sich eine komplette Anzeige beschreiben. Daraus ergeben sich 324 mögliche Kombinationen, welche durch die Konkatenation von Abkürzungen der Entitäten, sogenannte Codewörter, dargestellt werden können. Die Kurzdefinitionen und die jeweiligen Abkürzungen sind der Tabelle am Ende des Abstracts zu entnehmen. Für eine exakte und detaillierte Beschreibung sei auf die Arbeit von Tönnis und Plecher [TP11] verwiesen. Als kurzes Beispiel der Klassifizierung soll die virtuelle Teekanne aus Abb. 1 dienen. Diese wird *kontinuierlich* (I) als virtuelles *3D-Objekt* (II) der realen Welt hinzugefügt. Die Überlagerung des realen Markers, welcher die Position der Teekanne im Raum klar definiert, stellt die *kontakt-analoge* Registrierung (III) dar. Die Augpunkte der virtuellen Kamera - bzgl. des Objekts - und der realen Kamera bzw. auch der Augen des Benutzers (je nach Displaytyp) sind gleich, so dass die Perspektive (IV) *egozentrisch* genannt wird. Das Objekt ist zudem *direkt* zum Benutzer *referenziert* (V), da es sich unverdeckt in dessen Blickfeld befindet. Abschließend lässt sich feststellen, dass es sich um ein lokal begrenztes Trackingvolumen handelt und es daher dem *Environment-Mounting* (VI) zugeordnet wird. Somit kann die gesamte Anzeige durch das Codewort *c3caegodirenv* repräsentiert werden.

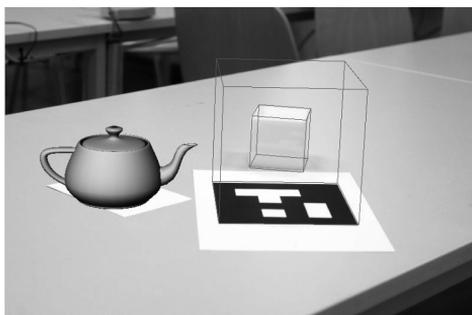


Abbildung 1: Klassische Kombination von Anzeigeprinzipien [T10]

3 Empirie und Evaluation

Nach Festlegung der Definitionen und Erstellung eines Klassifizierungsschemas wurde diese nun auf eine Vielzahl von Anwendungen, die sich mit AR beschäftigen, übertragen. Es wurden Paper und auch Poster der ISMAR-Konferenzen aus den Jahren 2002 - 2010 gesichtet, wovon sich 173 zur Klassifizierung eigneten. Die Übrigen enthielten keinerlei Anzeigeprinzipien. Aus den gesammelten Daten wurde auf Basis der Codewörter eine Statistik über die verwendeten Kombination erstellt, für die Unbenutzten hingegen wurde eine Einteilung bzgl. ihrer Umsetzbarkeit vorgenommen. Die Statistik besagt, dass überwiegend die klassische Kombination *c3caegodirenv* verwendet wurde; siehe dazu Abb. 2.

KLASSEN	KURZDEFINITION	CODE
I. Temporär		
Kontinuierlich	Virtuelles Objekt jederzeit sichtbar	<i>c</i>
Diskret	Anzeige ist ereignisabhängig	<i>d</i>
II. Raum		
2D	Anzeige in 2D	<i>2</i>
3D	Anzeige in 3D	<i>3</i>
III. Registrierung		
Unregistriert	Anzeige hat keinen räumlichen Bezug zur realen Welt	<i>un</i>
Registriert	Virtuelles Objekte wird richtig überlagert angezeigt	<i>reg</i>
Kontakt-analog	Virtuelles Objekt wird zusätzlich in der exakten Tiefe angezeigt und integriert sich geschmeidig in die reale Welt	<i>ca</i>
IV. Perspektive		
Egozentrisch	Augpunkte von Augen/realer Kamera und virtueller Kamera sind gleich; Kompensationsaufgaben möglich	<i>ego</i>
Egomotion	Augpunkte von Augen/realer Kamera und virtueller Kamera sind verschieden; Kompensationsaufgaben dennoch möglich	<i>m</i>
Exozentrisch	Augpunkte von Augen/realer Kamera und virtueller Kamera sind verschieden; keine Kompensations-, nur Folgeaufgaben möglich	<i>exo</i>
V. Referenzierung		
Direkt	Objekt ist sichtbar, befindet sich im Blickfeld	<i>dir</i>
Indirekt	Virtuelles Objekt weist auf ein reales Objekt hin, welches sich im Blickfeld befindet, aber verdeckt ist	<i>ind</i>
Referenziert	Virtuelles Objekt weist auf ein reales Objekt hin, welches sich nicht im Blickfeld befindet	<i>ref</i>
VI. Mounting		
Human	Koordinaten des virtuellen Objekts werden in Abhängigkeit der Position des Benutzer transformiert	<i>hum</i>
Environment	Koordinaten des virtuellen Objekts werden in das Koordinatensystem des Environments transformiert; Trackingvolumen lokal begrenzt	<i>env</i>
World	Koordinaten des virtuellen Objekts werden in das Welt-Koordinatensystem transformiert; Trackingvolumen „unbegrenzt“	<i>w</i>

Tabelle 1: Kurzdefinitionen der Klassen

ESARC - Referenzmodell und Ontologie für Enterprise Architecture Management

Sebastian Bourscheidt, Thomas Breuer, Tobias Brunner, Ben Fetler, Georg Fogel

Fakultät Informatik
Hochschule Reutlingen
Alteburgstraße 150
72762 Reutlingen

Sebastian.Bourscheidt@Student.Reutlingen-University.de

Thomas.Breuer@Student.Reutlingen-University.de

Tobias1.Brunner@Student.Reutlingen-University.de

Ben.Fetler@Student.Reutlingen-University.de

Georg.Fogel@Student.Reutlingen-University.de

Art der Arbeit: Forschungsarbeit

Betreuer: Prof. Dr. rer. nat. Alfred Zimmermann

Abstract: Unstrukturierte Unternehmensarchitekturen verursachen hohe Kosten und geben Anlass zur Evaluierung der vorhandenen Architekturframeworks. In Vorbereitung der Klassifikation, Reifegrad-Analysen, zyklischen Evaluierungen und Bewertungen von Service-orientierten Unternehmens- und Software-Architekturen werden Architektur-Referenzmodelle, basierend auf dem ESARC - Enterprise Services Architecture Reference Cube, mit der Hilfe eines Metamodell-Ansatzes zusammengeführt und integriert. Des Weiteren wird die Entwicklung einer konkreten wissensbasierten Ontologie vorgestellt, die in Zukunft eine intelligente Navigation in Architekturmodellen ermöglichen soll.

1 Einleitung – ESARC

Der Enterprise Services Architecture Reference Cube (ESARC), der abgeleitet von TOGAF eine Detaillierung auf acht Referenzarchitekturen bietet, ist als ein neues Architektur-Framework zu betrachten und adressiert aktuelle Konzepte im Enterprise Architecture Management (EAM). Hierbei werden Referenzarchitekturen zu Themen wie Cloud Computing, Sicherheit und vor allem für die serviceorientierte Modellierung angeboten. Die ESARC Referenzarchitekturen sind an bewährte Standards wie TOGAF und ITIL angelehnt und helfen bei der Erstellung und Definition von Architekturartefakten für die Evaluation, das Monitoring, die Durchführung und Erstellung sowie die Weiterentwicklung und Änderung von Unternehmensarchitekturen [ZI11].

Auf Basis des ESARC und dessen Referenzarchitekturen wurden in einem einjährigen Masterprojekt, detaillierte Metamodelle und wissensbasierte Ontologien modelliert. Die Metamodelle verknüpfen die Entitäten der Referenzarchitekturen auf konsistente Art und stellen eine formale Struktur dar (siehe Kapitel 2, Kapitel 3). Die daraus abgeleiteten Ontologien legen ein einheitliches Vokabular fest und klären die Relationen zwischen den Entitäten. Somit ist es möglich die Metamodelle über die modellierte Ontologie maschinell zu interpretieren (siehe Kapitel 4). In der Vision ergibt sich die intelligente Navigation bei einer heterogenen und dynamischen Vielfalt von Architekturkonzepten.

In diesem Paper wird beispielhaft auf eine Referenzarchitektur, die Business & Information Architecture, des ESARC eingegangen (siehe Kapitel 3). Weitere Referenzarchitekturen und deren Integration sind in einem separaten Dokument zu finden [FA12].

2 Meta-Metamodell der Open Group

Zur Formalisierung und Vereinheitlichung der Modellbildung werden Metamodelle eingesetzt. Unsere Überlegungen erweitern den Ausgangspunkt der Open Group und liefern auf diese Weise eine Basis für die erstellten Ontologien. Die Zielsetzung eines Metamodells liegt darin, Modellkonzepte und deren wesentliche Relationen darzustellen, diese besser verständlich zu machen, einen Konsens über ihre Verwendung und Bedeutung herbeizuführen und die Modellelemente präzise zu definieren.

Das entwickelte ESARC Metamodell spezialisiert das Metamodell der Open Group für service-orientierte Architekturen [OG10] und ergänzt dieses Basismodell um die wesentlichen Architekturen des Enterprise Architecture Management. Über das vorhandene Meta-Metamodell erfolgte die Integration aller entwickelten ESARC-Metamodelle. Darüber hinaus wird die vollständige Interoperabilität zwischen den einzelnen Referenzarchitekturen auf konsistente Art gewährleistet.

3 Business & Information Reference Architecture

Das entwickelte Metamodell und die zugehörige Ontologie wird in exemplarischer Detailliertheit für die Business & Information Reference Architecture des ESARC skizziert und in [FA12] umfassend dargestellt. Abbildung 1 zeigt die Entitäten der Business & Information Architecture anhand des entwickelten Metamodells.

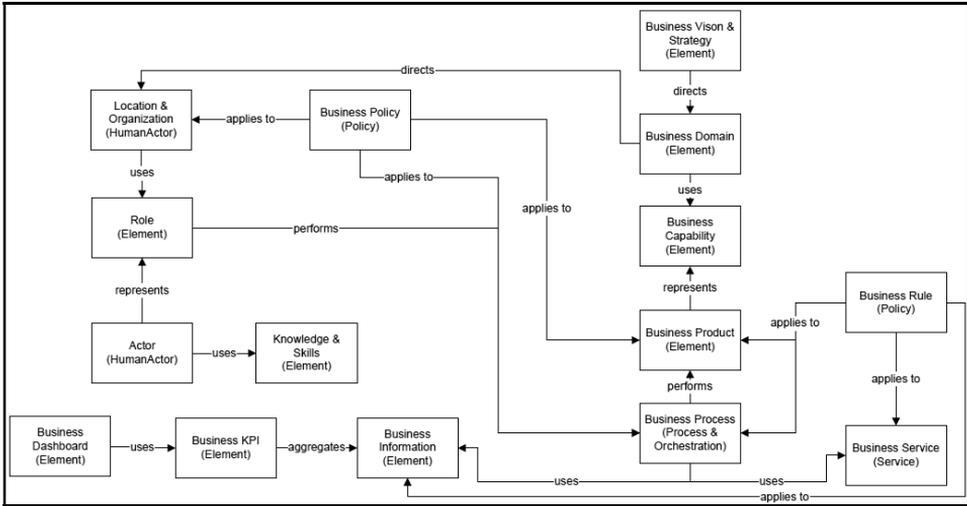


Abbildung 1: Business & Information Reference Architecture

Auf der Ebene der Business & Information Architecture werden die Unternehmensziele und -visionen festgelegt und über Geschäftsprozesse sowie Workflows abgebildet. Die Geschäftsprozesse ergeben sich aus der Fähigkeit der Unternehmung eine Leistung zu erbringen. Die Fähigkeit der Leistungserstellung ergibt sich aus der Integration verschiedener Produktionsfaktoren (Informationen & Wissen, Akteure, Lokationen etc.). Darüber hinaus sind Geschäftsregeln und Grundsätze zu beachten. Hinzu kommt, dass Unternehmensprozesse beständig kontrolliert, aufbereitet und anhand bestimmter Indikatoren (zum Beispiel: Key Performance Indicators) gemessen und vergleichbar gemacht werden müssen.

Neu an unseren Überlegungen ist die Spezialisierung und Detaillierung der allgemeinen Metamodellkonzepte aus [OG10] unter Zuhilfenahme des ESARC [ZI11].

4 Ontologie der Business & Information Reference Architecture

Die Ontologie der Business & Information Reference Architecture vereint die spezifischen Modellkonzepte und deren Relationen, in den von Protégé [PR12] zur Verfügung gestellten Basismechanismen (grafische Repräsentation). Die daraus abgeleitete symbolische Repräsentation (OWL) basiert auf einer XML-Grundstruktur und standardisierten OWL-Konzepten des W3C [OW12].

Die Ontologie umfasst sowohl das Metamodell der Open Group, als auch unsere spezifischen Erweiterungen zu den verschiedenen Modellen und stellt diese in einer Klassenhierarchie dar [FA12]. Um eine eindeutige Abgrenzung zu gewährleisten, wurden die Klassen der Business & Information Architecture durch das Präfix „Business“ gekennzeichnet. Durch die Definition von geeigneten Properties und Kardinalitäten wurden zudem die Beziehungen zwischen den Entitäten festgelegt.

5 Ausblick

Auf Basis des ESARC und dessen Architektur-Referenzmodellen wurden detaillierte Metamodelle und Ontologien entwickelt, die über ein Modell der Open Group [OG10] integriert werden. Die entstandenen Lösungselemente stellen aufgrund des serviceorientierten Ansatzes des ESARC, dessen Detaillierung der Referenzarchitekturen und der somit vielfältig vorhandenen Entitäten, eine neue und bisher nicht vorhandene Innovation für das Enterprise Architecture Management (EAM) dar.

Auf Grundlage der neu entwickelten Metamodelle ist eine konsistente Navigation, zwischen den acht Referenzarchitekturmodellen unter Berücksichtigung der spezifischen Relationen möglich. Dies bildet die Basis für standardisierte Analysen, Evaluierungen, Optimierungen und Neuausrichtungen der IT-Unternehmensarchitekturen. Durch die Formalisierung des Metamodells in einer Ontologie wird eine Grundordnung im EAM geschaffen. Über die in der Ontologie definierten Regelwerke und der Operation auf einem konsistenten Metamodell werden der Aufbau und die Navigation einer Unternehmensarchitektur für einen Anwender deutlich vereinfacht.

Im strategischen Ausblick ergibt sich die intelligente Navigation bei einer heterogenen und dynamischen Vielfalt von Architekturkonzepten. So sind Systeme zur Verwaltung der Unternehmensarchitektur denkbar, die anhand der definierten Ontologie, Konsistenzprobleme in der Unternehmensarchitektur automatisch erkennen und Vorschläge zur Erstellung und Veränderung bereitstellen. Die vorgestellten Lösungselemente bieten einen ersten Ansatz zur Integration der vorhandenen Standards in einem neuen Detaillierungsgrad. Des Weiteren bilden sie die Basis für Simulationen, automatisch unterstützte Klassifikationen, Diagnoseprozesse, Navigation und bessere Visualisierung integraler und erweiterter Konzepte für EAM.

Bis zu einer produktiven Inbetriebnahme müssen die vorhandenen Ansätze vervollständigt und weiterentwickelt werden. Zudem sollte eine umfangreiche multiperspektivische Evaluierung stattfinden, um so die Einsatzmöglichkeiten der Metamodelle und Ontologien zu prüfen.

Literaturverzeichnis

- [FA12] Forschungsarbeit; Bourscheidt, Breuer, Brunner, Fetler, Fogel: Ausarbeitung, ESARC-Referenzmodell und wissensbasierte Ontologie für Enterprise Architecture Management, Reutlingen, 2012.
- [OG10] The Open Group: Technical Standard, Service-Oriented Architecture Ontology, Berkshire, October 2010.
- [ZI11] Alfred und Gertrud Zimmermann, ESARC - Enterprise Services Architecture Reference Cube for Capability Assessments of Service-oriented Systems, Reutlingen, 2011.
- [PR12] Protégé, Stanford University, Official Website, <http://protege.stanford.edu/>, letzter Zugriff: 10.01.2012
- [OW12] World Wide Web Consortium, OWL Web Ontology Language Guide, 12.11.2009, <http://www.w3.org/TR/owl-guide/>, letzter Zugriff: 10.01.2012

Marktanalyse einer Authentication-as-a-Service-Lösung bei Telefonaten mittels des neuen Personalausweises

Steffen Brandt, Torsten Wiens und Michael Massoth

Hochschule Darmstadt
Fachbereich Informatik
sbrandt@fbihome.de, torsten.wiens@h-da.de
michael.massoth@h-da.de

Art der Arbeit: Masterarbeit
Betreuer/in der Arbeit: Prof. Dr. Michael Massoth

Abstract: Bei Telefonaten lassen sich Gesprächspartner an ihrer Stimme erkennen - vorausgesetzt, man kennt die Person, mit der man telefoniert. Ist dies jedoch nicht der Fall, muss man darauf vertrauen, dass sie die Wahrheit sagt oder durch technische Verfahren ein minimales Vertrauen schaffen. Diese Situation stellt tagtäglich in vielen Unternehmen ein Problem dar. Das Projekt *Universal Call Authentication* stellt eine Lösung für dieses Problem auf Basis des neuen Personalausweises und eines Smartphones bereit. Im Rahmen der Abschlussarbeit wurde eine Marktabschätzung für dieses Produkt durchgeführt.

Motivation

Führen zwei einander bekannte Personen ein Telefonat, besteht zwischen ihnen eine Vertrauensbasis. Diese besteht bei einem Telefongespräch zwischen zwei Personen, die einander unbekannt sind, nicht. Ungeachtet dessen existiert auch im zweiten Fall das Bedürfnis nach Sicherheit. Dies gilt für Privatpersonen und für Geschäftsleute. Mit Hilfe des Produkts *Universal Call Authentication* wird die Basis für gegenseitiges Vertrauen durch eine sichere Authentifizierung der Gesprächspartner geschaffen. Dieses Produkt ermöglicht mit Hilfe des neuen Personalausweises (nPA) eine eindeutige Identifizierung von Gesprächsteilnehmern in Telefonaten. Zusätzlich zum nPA [PA10] wird nur ein Smartphone mit einem NFC-Lesegerät (Near Field Communication) benötigt [NFC12]. Im Rahmen der vorgestellten Abschlussarbeit wurde eine Marktanalyse für *Universal Call Authentication* erstellt.

Zwar existieren Konkurrenzprodukte, die eine Authentifikation ermöglichen. Jedoch richten sich diese Lösungen an Dienstleister, die bestehenden Kunden ein Mehr an Sicherheit ermöglichen wollen. Personen, die bislang keine Kunden des Dienstleisters sind, bleibt diese Lösung verschlossen.

Universal Call Authentication ermöglicht es, eine Authentifikation durchzuführen, ohne zuvor ein spezielles Gerät o.ä. an den Kunden auszuliefern. Der Nutzerkreis erstreckt sich somit nicht nur auf ein Dienstleister-Kunden Verhältnis, sondern ebenfalls auf verschiedene Personen- und Berufsgruppen, die in der Kommunikation zur Geheimniswahrung verpflichtet sind und denen die Möglichkeit der sicheren Authentifikation durch Konkurrenzprodukte nicht gegeben ist. Durch den Einsatz des untersuchten Produkts wird das Schutzbedürfnis gestillt und der Wunsch nach Sicherheit erfüllt.

Aktuelle Marktprognosen zeigen sehr positive Entwicklungsperspektiven für das Projektvorhaben. Unter anderem wird der Absatz der Komponenten, die zur Nutzung des Produkts benötigt werden, zukünftig weiter zunehmen. Hierbei sind besonders die Absatzzahlen für Smartphones mit NFC-Technologie, die Nutzung von (mobilen) Internetzugängen und die Nutzung von Internettelefonie zu beachten. Für die erstgenannten Produkte wird ein Wachstum von 93,2 Mio. Geräten im Jahr 2011 auf 544,7 Mio. Geräte im Jahr 2015 prognostiziert [IS11]. Dies entspricht einem Wachstum von fast 585%. Ebenfalls nimmt die Nutzung mobiler Internetzugänge und von Internetzugängen im Allgemeinen zu [AZ11]. Während die regulären Internetanschlüsse seit 2003 einem stetigen jährlichen Wachstum von 2% unterliegen, wird für 2011 ein Zuwachs von 4% erwartet. Bei mobilen Internetzugängen verursachten Nutzer im Jahr 2010 im Schnitt einen Datenverkehr von 91 MB pro Monat. Dieser Wert wird voraussichtlich auf 166 MB pro Anschluss im Jahr 2011 ansteigen [DC11]. Dies entspricht einem Wachstum von ca. 82%, mit steigender Tendenz. Die Zahl der Nutzer von Internettelefonie in Deutschland betrug 8,19 Mio. im Jahr 2007. Bis 2010 stieg die Zahl auf 13,06 Mio. Nutzer (s. Abbildung 1) [DC11]. Diese Zahlen zeigen durchweg positive Tendenzen auf den für *Universal Call Authentication* bedeutsamen Märkten auf.

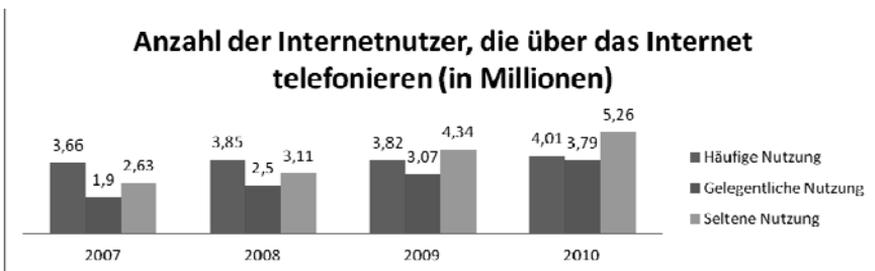


Abbildung 1: Marktentwicklung der Internettelefonie

Um eine Präsentation des beschriebenen Produkts zu ermöglichen, wurde bereits mit einer prototypischen Implementierung begonnen. Die Entwicklung wird bis zur CeBIT 2012 abgeschlossen sein, um mit Hilfe des Demonstrators Geschäftskontakte zu knüpfen. Anschließend wird die Applikation um die fehlenden Funktionen erweitert und getestet. Die daraus resultierenden Ergebnisse müssen analysiert werden und in eine erneute Implementierung einfließen. Spätestens Ende 2012 soll ein marktfähiges Produkt fertiggestellt werden.

Da das Produkt neben der Authentizität auch die Integrität des Gesprächs wahrt, ist es für jeden interessant, der sein Bedürfnis an Sicherheit in Telefonaten erfüllt sehen möchte. Somit werden nicht nur Kunden von Dienstleistern angesprochen, sondern zum Beispiel auch Gewerbetreibende oder Ärzte.

Die aufgeführten Marktprognosen und die Resonanz bezüglich der Präsentation des Produkts auf der CeBIT 2011 zeigen eine positive Entwicklung. Einem erfolgreichen Marktstart stehen nur zwei Risiken im Weg: Einerseits besteht die Gefahr, dass ein Wettbewerber ein ähnliches Produkt am Markt platziert. Dies stellt ein Problem dar, weil sich Kunden beim Thema Sicherheit eher an am Markt etablierte Anbieter wenden [FC09][CW11]. Andererseits besteht die Gefahr, dass die Kundenakquise generell schwierig ist, da der Anbieter am Markt für Sicherheitsprodukte unbekannt ist. Beide Risiken können jedoch reduziert werden, indem entweder in einem frühen Stadium große Referenzkunden gewonnen werden oder eine Kooperation mit einem bekannteren Unternehmen eingegangen wird.

Eine im Rahmen der Ausarbeitung durchgeführte Marktabschätzung ergab, dass bei einer optimistischen Marktentwicklung im Jahr 2012 ein Umsatz von 19.000 € erreicht werden kann, der bis zum Jahr 2015 auf 1,361 Mio. € wächst. Bei einer pessimistischen Abschätzung kann für das Jahr 2012 ein Umsatz von 4.725 € erwartet werden. Im Jahr 2015 wächst der Umsatz des adressierbaren Marktes auf 453.704 €.

Eine Kostenbetrachtung für das erste Geschäftsjahr, in der die einmaligen und die laufenden Kosten analysiert wurden, zeigte, dass mit einmaligen Aufwendungen in Höhe von 32.050 € und laufenden Kosten von 85.240 € zu rechnen ist. Dies ergibt eine Summe von 117.290 € für das erste Geschäftsjahr. Selbst wenn die Kosten mit der Zeit steigen, wäre der Umsatz sogar bei der pessimistischen Marktabschätzung ab dem dritten Geschäftsjahr höher als die Kosten; für das vierte Geschäftsjahr würde wahrscheinlich kein Fremdkapital mehr benötigt.

Ausblick

Aufgrund der durchgeführten Markt Betrachtung kann dem Produkt *Universal Call Authentication* ein guter Marktstart prognostiziert werden. Mit Hilfe der durchgeführten Beispielrechnung konnte der Rahmen des Finanzbedarfs grob abgesteckt und eine Abschätzung der möglichen Marktpositionierung durchgeführt werden.

Literaturverzeichnis

- [AZ11] ARD/ZDF: Entwicklung Onlinenutzung. <http://www.ard-zdf-onlinestudie.de/index.php?id=onlinenutzung0>, abgerufen 12.01.2012
- [CW11] Computerwoche: Security-as-a-Service ist für Mittelständler ideal. <http://www.computerwoche.de/security/1903672/>, erstellt 10.11.2011, abgerufen 30.11.2011
- [DC11] Dialog Consult/VATM: 13 gemeinsame TK-Marktanalyse 2011. http://www.vatm.de/fileadmin/publikationen/studien/2011-10-27_TK-Marktstudie-2011.pdf, abgerufen 11.01.2012
- [FC09] Forrester Consulting: Authentication-As-A-Service. <http://www.verisign.co.uk/static/auth-as-a-service.pdf>, erstellt 20.03.2009, abgerufen 21.09.2011
- [IS11] Rebello, Jagdish: US Wireless Carriers Partner with Big Credit Card Companies, Boosting Cell Phone NFC Market. <http://www.isuppli.com/Mobile-and-Wireless-Communications/News/Pages/US-Wireless-Carriers-Partner-with-Big-Credit-Card-Companies-Boosting-Cell-Phone-NFC-Market.aspx>, erstellt 12.05.2011, abgerufen 09.01.2012
- [NFC12] International Standardization Organisation: <http://www.iso.org/iso/search.htm?qt=NFC&searchSubmit=&Search&type=simple&published=on>, abgerufen am 13.01.2012
- [PA10] Personalausweisportal. <http://www.personalausweisportal.de/>, abgerufen 13.01.2012
- [UC12] Universal Call Authentication Projektseite. <http://u-can.h-da.de>

Ein patternbasiertes Vorgehensmodell für den Enterprise Services Architecture Reference Cube (ESARC)

Tobias Brunner

Enterprise Architecture Management
Hochschule Reutlingen
Alteburgstraße 150
72762 Reutlingen

Tobias1.Brunner@Student.Reutlingen-University.de

Art der Arbeit: Master-Thesis

Betreuer der Arbeit: Prof. Dr. rer. nat. Alfred Zimmermann

Abstract: Vorhandene Enterprise Architecture (EA) Frameworks bieten bisher keinen ausreichenden Umfang bei der Bereitstellung von Referenzarchitekturen und Vorgehensmodellen zur umfassenden und gesteuerten Entwicklung von Unternehmensarchitekturen. Daraus ergibt sich, dass in den meisten Modellierungsansätzen nur allgemeingültige Aussagen über den Aufbau und die Gestaltung von Referenzarchitekturen getroffen werden. Der Enterprise Services Architecture Reference Cube (ESARC), der abgeleitet von TOGAF eine Detaillierung auf acht Referenzarchitekturen bietet, ist als ein neues Architektur-Framework zu betrachten. Auf dieser Basis wurde ein patternbasiertes Vorgehensmodell entwickelt, das bewährte Standards integriert und erweitert. Im Resultat ergibt sich ein Patternkatalog, der einen Einstieg zur Modellierung sowie die geführte Entwicklung der acht Referenzarchitekturen des ESARC vorgibt.

1 Einleitung

Das entwickelte Vorgehensmodell integriert Patterns, die eine Iteration über den Enterprise Services Architecture Reference Cube (ESARC) [ZIM11] vorgeben und die Gestaltung grundlegender Architekturartefakte darstellen. Darüber hinaus werden klare Ergebnisdokumente gefordert. Alle Patterns wurden unter Berücksichtigung aktuell vorhandener Enterprise Architecture Frameworks, Vorgehensweisen und Best Practices aus dem Enterprise Architecture Management (EAM) -Umfeld erstellt. Als ein wesentlicher Ausgangspunkt diente die von TOGAF entwickelte Architecture Development Method (ADM) [TOG09] und der Service-Lifecycle von ITIL [OLB08].

Ziel dieser Arbeit ist die Bereitstellung eines Core-Patternkatalogs, der eine strukturiert geführte Entwicklung einer Unternehmensarchitektur vorgibt. Jedes Unternehmen kann den bereitgestellten Patternkatalog erweitern und auf die eigenen Bedürfnisse anpassen. In diesem Paper wird das Vorgehensmodell beschrieben und ein Auszug aus dem vorliegenden Patternkatalog dargestellt. Aktuell beinhaltet der Katalog 28 Patterns zur umfassenden Entwicklung einer Unternehmensarchitektur.

2 Das patternbasierte Vorgehensmodell

Das patternbasierte Vorgehensmodell bietet einen Einstiegspunkt sowie eine navigierbare Vorgehensweise zur Iteration über die acht Referenzarchitekturen des Enterprise Services Architecture Reference Cube (ESARC). Diese Vorgehensweise ist in einem Pattern beschrieben¹.

Des Weiteren wird auf jeder zu durchlaufenden Referenzarchitektur eine einheitliche Vorgehensweise dargestellt, die die Entwicklung der aktuell vorhandenen Referenzarchitektur vorgibt. Zu jeder Iterationsphase innerhalb der Referenzarchitekturen existiert ein Pattern.

2.1 Process Model Patterns

Die Patterns, die sich mit der Iteration über den gesamten ESARC sowie innerhalb der Referenzarchitekturen befassen, werden als „Process Model(PM) Patterns“² bezeichnet. Die folgende Abbildung verdeutlicht das beschriebene Vorgehensmodell.

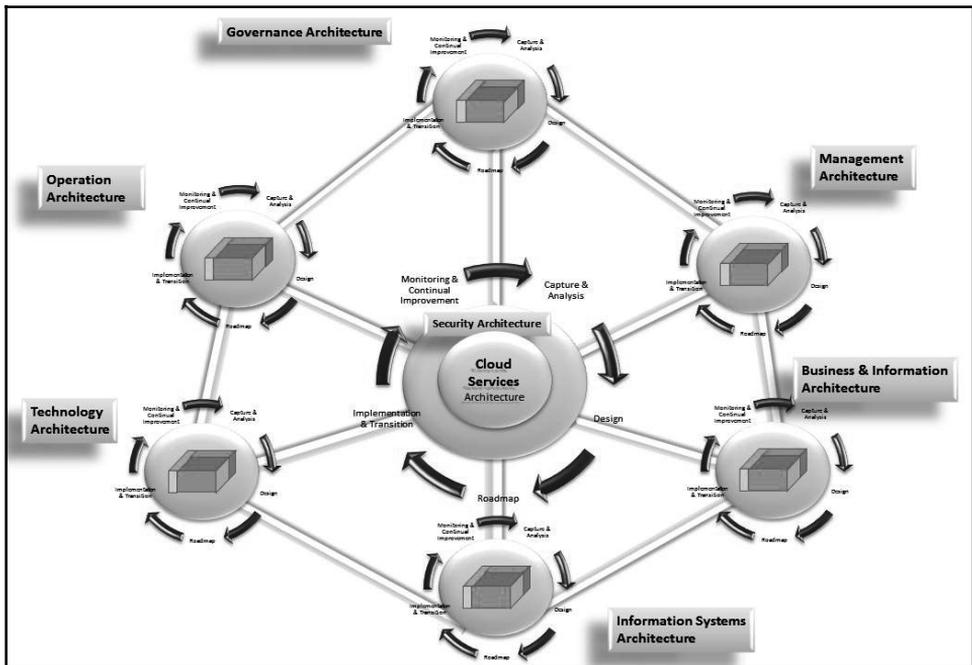


Abbildung 1: ESARC – Vorgehensmodell

¹ PM – Approach Lifecycle Pattern

² PM – Architecture Capture & Analysis Pattern, PM – Architecture Design Pattern, PM – Architecture Roadmap Pattern, PM – Architecture Implementation & Transition Pattern, PM – Architecture Monitoring & Continual Improvement Pattern

2.2 Reference Architecture Patterns

Zusätzlich zu den „Process Model Patterns“, die eine Vorgehensweise zur Iteration über den ESARC abbilden, existieren sogenannte „Reference Architecture(RA) Patterns“, die zur Erstellung und Definition von Architekturartefakten für die Evaluation, das Monitoring, die Durchführung und Erstellung sowie die Weiterentwicklung und Änderung der acht Referenzarchitekturen des ESARC von Bedeutung sind.

2.3 Bereitstellung des Patternkatalogs

Die Patterns wurden mit dem Ziel entwickelt, einem Unternehmensarchitekten eine klare, leichtverständliche und standardisierte Handlungsanweisung mit vorgegebenen Ergebnisdokumenten zur Verfügung zu stellen. Alle entwickelten Patterns wurden über eine Webanwendung bereitgestellt. So können zu dem dargestellten Vorgehensmodell Erklärungen der einzelnen Architekturebenen des ESARC sowie der Patternkatalog und die zugehörigen Ergebnisdokumente abgefragt werden.

3 Auszug aus dem Patternkatalog

Im folgenden Abschnitt werden zur Entwicklung, der durch den ESARC vorgegebenen „Management Reference Architecture“ und der „Cloud Services Reference Architecture“ jeweils ein „Reference Architecture (RA) Pattern“ dargestellt. Diese Patterns adressieren wichtige Architekturartefakte zur Erstellung einer Unternehmensarchitektur. Das im Folgenden dargestellte „Service Lifecycle Pattern“ bezieht sich auf die „Management Reference Architecture“. Das „Cloud Service Model Pattern“ bezieht sich auf die „Cloud Services Reference Architecture“.

3.1 Service Lifecycle Pattern

Wie kann die Entwicklung von unternehmensweiten Services standardisiert, zielorientiert und erfolgreich ein- und umgesetzt werden?

Problem: Werden externe sowie interne Services ohne konkrete Ausrichtung an ein strukturiertes und einheitlich definiertes Vorgehen entwickelt, kann nicht gewährleistet werden, dass diese zielorientiert, redundanzfrei vorhanden sind und dem vorgegebenen Strategieplan folgen.

Lösung: ITIL beschreibt einen kontinuierlichen Verbesserungsprozess der als Service Lifecycle bezeichnet wird. Ausgehend von der strategischen Ausrichtung des Unternehmens (Service Strategy) werden die Services entwickelt (Service Design) und in den operativen Betrieb überführt (Service Transition). Anschließend werden die Services betrieben und den Geschäftsprozessen zur Verfügung gestellt (Service Operation). Diese Phasen werden von einem kontinuierlichen Verbesserungsprozess (Continual Service Improvement) umschlossen. [OLB08]

Erklärung: Die ausführliche Beschreibung des Service Lifecycles ist in den fünf Publikationen über ITIL nachzulesen. Diese fünf Core-Publikationen bilden die Phasen eines iterativen und mehrdimensionalen Lifecycle für die im Unternehmen vorhandenen oder zu erstellenden Services.

3.2 Cloud Service Model Pattern

*Wie können Unternehmensressourcen als Services (*aaS) in einer Cloud repräsentiert werden?*

Problem: Aufgrund der aktuell vorhandenen und nicht hinreichenden Servicemodellen (IaaS, PaaS, SaaS) können umfangreiche Unternehmensarchitekturen nicht abgebildet werden.

Lösung: Die bisher vorhandenen Servicemodelle müssen erweitert werden. Die Servicemodelle müssen sich an den Unternehmensstrukturen orientieren. Hierbei müssen Services aller Ebenen einer Unternehmensarchitektur in der Cloud abgebildet werden können.



Erklärung:

Die dargestellte Unterteilung von Services (*aaS) veranschaulicht eine Möglichkeit, Unternehmensressourcen detailliert abzubilden.

Abbildung 2: Servicemodelle für eine Unternehmensarchitektur [LIN09]

4 Fazit und Ausblick

Die Vision, die mit dem Aufbau des Core-Patternkatalog verfolgt wird, ist ein fortwährend wachsender Katalog, auf diesen zentral über eine Webanwendung zugegriffen werden kann. Bei einer produktiven Anwendung des Patternkatalogs muss dieser um weitere Patterns vervollständigt werden, da momentan noch nicht alle acht Referenzarchitekturen des ESARC berücksichtigt werden. Zudem ist eine multiperspektivische Evaluierung des Vorgehensmodells und den zugehörigen Patterns vorzunehmen, um deren Einsatzmöglichkeiten zu prüfen.

Literaturverzeichnis

- [LIN09] David S. Linthicum: Cloud Computing and SOA Convergence in Your Enterprise, 2009.
- [OLB08] Alfred Olbrich: ITIL kompakt und verständlich, Wiesbaden, 2008.
- [TOG09] The Open Group: TOGAF 9 Foundation Study Guide, September 2009.
- [ZIM11] Zimmermann, A., Zimmermann, G.: ESARC-Enterprise Services Architecture Reference Cube for Capability Assessments of Service-oriented Systems. In SERVICE COMPUTATION 2011 - The Third International Conferences on Advanced Service Computing, September 25-30, Rome, Italy, 2011.

Ontologie zur Repräsentation und Auswertung von Problemdiagnose- und Problemlösungswissen im Kontext eines semantikerweiterten Kennzahlendashboards

Michael Falkenthal, Dierk Jugel

Hochschule Reutlingen
Fakultät Informatik

Michael.Falkenthal@Student.Reutlingen-University.de
Dierk.Jugel@Student.Reutlingen-University.de

Art der Arbeit: Masterthesis
Betreuer: Prof. Dr. rer. nat. Alfred Zimmermann

Abstract: Kennzahlensysteme sind im Zusammenspiel mit Kennzahlen- bzw. Performancedashboards zu wichtigen Steuerungsunterstützungssystemen in Unternehmen geworden. Diese Arbeit zeigt die Limitierung herkömmlicher Kennzahlendashboards auf die mathematisch-logische Verdichtung von Informationen im Kontext eines Kennzahlensystems auf und stellt eine Erweiterung von Kennzahlendashboards um semantische Informationsverdichtungsmechanismen zur automatisierten Diagnose von Problemsituationen im Unternehmen vor.

1 Einleitung und Stand der Technik

Kennzahlensysteme haben sich in den letzten Dekaden in beinahe allen Unternehmen als zentrale Steuerungsunterstützungsinstrumente¹ für das Management etabliert [Gle08]. Mit steigenden Rechenkapazitäten und immer engerer Vernetzung der Daten in Unternehmen durch Client-Server-Architekturen, fanden Kennzahlensysteme in den letzten 20 Jahren immer öfter eine IT-technische Umsetzung in sogenannten Kennzahlen- bzw. Performancedashboards [Eck11]. Dies beruht insbesondere auf der Tatsache, dass heutzutage alle wesentlichen Geschäftsprozesse eines Unternehmens eine Implementierung in IT-Systemen erfahren. So werden beispielsweise Geschäftsprozesse in der Regel in Enterprise Resource Planning (ERP) Systemen wie SAP ERP abgebildet und damit einer IT-Unterstützung zugeführt. Durch die IT werden so Datenpools geschaffen, die als Grundlage für Performancedashboards dienen können. Kennzahlen werden auf Performancedashboards als Informationseinheiten dargestellt, welche den Mitarbeitern intuitiv und schnell einen Überblick über den aktuellen Zustand des Unternehmens ermöglichen

¹Als Fachterminus wird in der Literatur häufig das Akronym MSS für **M**anagement **S**upport **S**ystems verwendet [Gle08]

sollen. Basisinformationen werden in Kennzahlensystemen verdichtet, indem Kennzahlensysteme formal Kennzahlen und deren mathematisch-logische Verknüpfungen vorgeben, oft auch gepaart mit einer Vorgehensweise, die beschreibt, wie Kennzahlen oder sogar ein ganzes Kennzahlensystem von der Unternehmensstrategie abgeleitet werden können [Eck11, Erl04, HK98, KN96, KN09, Lau10, Sch07]. Bekannte Beispiele hierfür sind das DUPONT SCHEMA [Sch07, Fri96] und die BALANCED SCORECARD [KN09]. In Unternehmen besteht der Bedarf nach verdichteten Informationen, da einfache Informationen meist nicht dem gewünschten Detaillierungsgrad entsprechen, der für Analysen eines Gesamtkontextes im Unternehmen notwendig ist. So werden beispielsweise angefallene Verkaufsdaten über die Zeit zu Absatz und Umsatz im letzten Monat verdichtet und als Gesamtbild für die Verkaufperformance des Unternehmens in jeweils einer aggregierten Kennzahl abgebildet.

Die Auswertung des Zustands des Unternehmens wird durch Mitarbeiter vorgenommen, indem die aggregierten Kennzahlen auf dem Kennzahlendashboard dargestellt und gegen von der Strategie abgeleitete Sollwerte abgeglichen werden. Auf dem Dashboard wird so neben der Analyse einzelner Kennzahlen durch den Mitarbeiter eine semantische Verknüpfung von formal voneinander getrennten Informationen vorgenommen, um die Einzelinformationen in einem Gesamtkontext auszuwerten. Durch sein Fachwissen kann ein Mitarbeiter auf im Unternehmen vorliegende Probleme schließen, wobei die Einzelinformationen in Form von Kennzahlen als Symptome von Problemen interpretiert werden können, wenn sie zu stark vom geplanten Sollwert abweichen. Der Prozess der Problemdiagnose muss von Mitarbeitern selbstständig und intuitiv bewältigt werden, da Wissen zur Identifikation von Problemen meist nicht dokumentiert im Unternehmen oder gar im Kennzahlensystem zur Verfügung steht. Das Kennzahlendashboard als MSS ist nicht selbstständig in der Lage Informationen in Form von Kennzahlen semantisch zu verknüpfen und zur automatisierten Problemdiagnose zu nutzen.

2 Informationsverdichtung zur Problemdiagnose und Inferenz von Problemlösungen

Wie bereits aufgezeigt dienen Kennzahlensysteme in Kombination mit Kennzahlendashboards der Aggregation und Verdichtung von Informationen zur Entscheidungsunterstützung in Unternehmen. Der Prozess der Problemdiagnose und Lösungsinferenz geht jedoch weit über die einfache mathematisch-logische Verknüpfung von Informationen hinaus.



Abbildung 1: Informationsverdichtungsprozess mit Problemdiagnose und Lösungsinferenz

Abbildung 1 zeigt die einzelnen Stufen des Informationsverdichtungsprozesses mit Pro-

blemdiagnose und Lösungsinferenz, die wir im Zuge dieser Arbeit identifiziert haben, ausgehend von (I) der Erhebung atomarer grundlegender Informationen, über (II) die Aggregation der Messwerte und Fakten zu Kennzahlen bis hin (III / IV / V) zur Inferenz von Problemdiagnosen und Problemlösungen. Die Stufen I und II werden heutzutage bereits von allen Performancedashboards implementiert, die weiteren Stufen der Informationsverdichtung zur Identifikation von Problemsituationen im Unternehmen hingegen sind Gegenstand dieser Arbeit. Die grundlegenden Prinzipien und Methoden der Anwendung von Wissen zur Diagnose von Problemen sind aus der Disziplin der Medizin entliehen [SS09]. Zur Diagnose von Problemen muss zunächst eine Wissensbasis geschaffen werden, in der alle diagnostizierbaren Probleme formalisiert werden. Ein Problem π_i wird in einer OWL Ontologie als Klasse konzeptualisiert, indem es durch eine Menge notwendiger Symptome Σ_i hinreichend beschrieben wird $\pi_i \Leftrightarrow \Sigma_i$. Ein Symptom σ_i definiert den Wertebereich einer Kennzahl κ_i , der als notwendiges Indiz für das Auftreten eines Problems modelliert wird. Ein Problem kann somit als komplexe semantische Verdichtung von Kennzahlen definiert werden, wobei eine Kennzahl aufgrund ihrer Ausprägung durch eine logische Definition als Symptom klassifiziert werden kann. Die Menge der Kennzahlausprägungen K definiert den Zustand ζ des Unternehmens zum Zeitpunkt t ($\zeta_t = (\kappa_{1_t}, \dots, \kappa_{i_t} | \kappa_n \in K_t; 1 \leq n \leq i; n, i \in \mathbb{N})$). Basierend auf diesem Modellierungsprinzip einer Ontologie, kann der Zustand des Unternehmens zu einem Zeitpunkt, beschrieben durch die Ausprägungen aller Kennzahlen, durch einen Reasoner klassifiziert werden. Durch die Klassifikation inferiert das System ausgehend vom Zustand auf Probleme und führt damit eine automatisierte Problemdiagnose durch. Wird abhängig von den Problemen auch Lösungswissen in die Ontologie aufgenommen, ist das System sogar in der Lage, Lösungen für inferierte Probleme aufzuzeigen.

3 Das semantikerweiterte Kennzahlendashboard

Wird die oben beschriebene Funktionalität auf ein herkömmliches Kennzahlendashboard angewendet, kann dieses um eine semantische Informationsverdichtung und automatisierte Problemdiagnosen erweitert werden. Zur Zusammenführung von Monitoringinformationen aus verschiedenen Systemen, nutzt der von uns implementierte Prototyp eines semantikerweiterten Kennzahlendashboards ein Metamodell, anhand dessen Instanziierung eine globale und systemunabhängige Beschreibung von Informationen im Kontext eines Kennzahlensystems realisiert wird. Das Metamodell definiert INFORMATIONEN, die aus ATTRIBUTEN bestehen. Ein Attribut ist Träger einer Information in Form eines Datums, wobei das Datum alphanumerisch, logisch, oder eine andere Information sein kann. Die Instanziierung dieses Metamodells definiert KENNZAHLEN als Informationsaggregate, deren mathematisch-logische Zusammensetzungen durch weitere META-INFORMATIONEN beschrieben werden. Auf diese Weise kann ein Kennzahlensystem definiert und modellgetrieben berechnet werden. Kennzahlausprägungen werden der Ontologie zur Problemdiagnose als INDIVIDUEN hinzugefügt. Die Menge der Individuen in der Ontologie beschreibt so den Zustand des Unternehmens, durch dessen Klassifikation das semantikerweiterte Kennzahlendashboard auf vorliegende Probleme schließen kann.

4 Fazit und Ausblick

Durch die Formalisierung von Problemwissen in einer Ontologie, kann ein Kennzahlen-dashboards zu einem System erweitert werden, das vom Zustand des Unternehmens, widergespiegelt durch Kennzahlen, durch automatisierte Diagnosen auf Probleme schließen kann. Die Auswertung von Kennzahlen zur Identifikation von Problemen muss damit nicht mehr intuitiv von Mitarbeitern übernommen werden. Der Arbeitsaufwand eines Mitarbeiters lässt sich so von der reinen Problemdiagnose hin zur Problemlösung verschieben. Ein Ausbau des vorgestellten Ansatzes könnte die Modellierung von Problemlösungswissen auf eine formale und maschinenverständliche Art und Weise adressieren, so dass ein semantikerweitertes Kennzahlendashboards nicht nur automatisiert auf Probleme schließen kann, sondern auch in die Lage versetzt wird, selbstständig Problemlösungswissen anzuwenden.

Literatur

- [Eck11] Wayne W. Eckerson. *Performance dashboards : measuring, monitoring, and managing your business*. Wiley, Hoboken, NJ, 2. ed.. Auflage, 2011.
- [Erl04] Leon Erlanger. Monitoring Your Business IN REAL TIME. *InfoWorld*, 26(30):43 – 47, 2004.
- [Fri96] George Thomas Friedlob. *Understanding return on investment*. Wiley, New York [u.a.], 1996.
- [Gle08] Ronald Gleich, Hrsg. *Management Reporting : Grundlagen, Praxis und Perspektiven*. Haufe Fachpraxis. Haufe, Freiburg, 1. aufl.. Auflage, 2008.
- [HK98] Péter Horváth und Lutz Kaufmann. Balanced Scorecard - Ein Werkzeug zur Umsetzung von Strategien. *HARVARD BUSINESSmanager*, 5/1998:39 – 48, 1998.
- [KN96] Robert S. Kaplan und David P. Norton. Using the Balanced Scorecard as a Strategic Management System. *Harvard Business Review*, January-February:75 – 85, 1996.
- [KN09] Robert S. Kaplan und David P. Norton. *The balanced scorecard : translating strategy into action*. Harvard Business School Press, Boston, Mass., 2009.
- [Lau10] Gerhard Laußer. *Nagios - Das Praxisbuch : Open Source-Monitoring im Unternehmen*. Open source library. Addison-Wesley, München, 2010.
- [Sch07] Maximilian Schreyer. *Entwicklung und Implementierung von Performance Measurement Systemen*. Dissertation, Universität Bayreuth, Wiesbaden, 2007.
- [SS09] Cord Spreckelsen und Klaus Spitzer, Hrsg. *Wissensbasen und Expertensysteme in der Medizin : KI-Ansätze zwischen klinischer Entscheidungsunterstützung und medizinischem Wissensmanagement*. Vieweg + Teubner Verlag / GWV Fachverlage GmbH, Wiesbaden, Wiesbaden, 2009.

Effiziente Nutzung von Cross-Selling im Bankumfeld

Markus Fortak
Schloßstraße 6
83278 Traunstein
markus.fortak@gmx.de

Gutachter: Prof. Dr. Herbert Fischer
HDU - Hochschule Deggendorf

Abstract: Der vorliegende Beitrag setzt es sich zum Ziel, einen grundlegenden Ansatz für Cross-Selling im Privatkundengeschäft aufzuzeigen. Dabei soll vor allem verdeutlicht werden, das Cross-Selling nicht ausschließlich auf Berechnungen aktueller Algorithmen und Software-Tools basiert, sondern das Zusammenspiel aus Geschäftsprozessen, Technologien und organisatorischen Aspekten maßgeblich ist. Den fachliche Rahmen spannt das Customer Relationship Management (CRM) hinsichtlich einer kundenorientierten Strategie auf.

1 Einführung

Cross-Selling wird in der Praxis häufig diskutiert, aber nur selten konsequent als Strategie in Banken verfolgt. Trotz fehlender Ansätze wird Cross-Selling als zentraler Lösungsansatz zur Überwindung der Ertragsproblematik im Privatkundengeschäft angesehen (vgl. [WKS03, S.66]). Die Konzeption basiert primär auf den Erkenntnissen der gängigen Literatur ([Sch02, Wit10]) und führt diese Gedanken auf der operativen Ebene fort.

2 Cross-Selling-Prozesse im Customer Relationship Management

Aus Sicht der Bankinstitute sollen Kunden langfristig gebunden und die Cross-Selling-Quote erhöht werden, um die Profitabilität des Retailsegments zu verbessern, ohne hohe variable Investitionen pro Kunde zu tätigen. Der Kunde erhält so einen starken Allfinanzdienstleister, der ihn mit bedarfsgerechten Produkten in verschiedenen Lebensphasen unterstützt und so die kundeneigenen Beschaffungskosten senkt. Im Kontext dieses Beitrages lassen sich drei Oberziele für die Gestaltung eines Cross-Selling-Prozesses benennen:

- Cross-Selling strukturiert durchführen
- Integration in eine CRM-Strategie
- Einfache und transparente Abläufe

Abbildung 1 zeigt die grundsätzlichen Phasen des zyklischen Cross-Selling-Prozesses. Dieser Hauptprozess wird im Folgenden nach dem „Top-Down-Vorgehen“ in seine Kernprozesse zerlegt und beschrieben.



Abbildung 1: Phasen des Cross-Selling-Prozesses in Anlehnung an [BNP08, S. 238]

Informationen sammeln

Als Informationsquelle dient ein CRM-System, das als unabdingbare Grundlage für alle Cross-Selling-Aktivitäten gilt. Besonderer Bedeutung kommt hier der Datenqualität zu, da eine unzureichende Datenbasis den Erfolg aller Maßnahmen gefährdet. Für ein effizientes Vorgehen ist vor allem ein ausgeprägtes Wissen über das „Cross-Selling-Potenzial“ der Kunden notwendig. Dabei ist nach der „Cross-Selling-Bereitschaft“ und des „Cross-Selling-Bedarfs“ zu differenzieren. Die Bereitschaft steht in unmittelbarem Zusammenhang mit der Qualität der angebotenen Leistung, während der Bedarf die reine Notwendigkeit eines Produkts widerspiegelt (vgl. [TRS06, S. 134]). Der Bedarf orientiert sich am Kundenlebenszyklus und ist für Banken nur in geringem Maße zu beeinflussen. Die Cross-Selling-Bereitschaft kann durch die Kundenzufriedenheit und ein emotionales Verbundgefühl weitere Vertragsabschlüsse stimulieren. Folglich ist der Bedarf an weiteren Finanzprodukten der Cross-Selling-Bereitschaft vorgelagert (vgl. [Wit10, S. 40f.]). Mit Hilfe von Werkzeugen wie der Kundenzufriedenheitsanalyse oder zyklischen Berichten aus dem Beschwerdemanagement können Gründe für die mangelnde Bereitschaft eruiert werden. Durch das Etablieren von Serviceleistungen (z.B. kostenloser Finanzcheck) oder die aktive Kundenansprache zu besonderen Ereignissen (z.B. 18. Geburtstag) sind Mittel um den Kundenbedarf näher kennen zu lernen.

Informationen analysieren

Unterstellt man der Bereitschaft und dem Bedarf eine unmittelbare Beziehung, so ergibt sich eine „Cross-Selling-Potenzial-Matrix“ wie in Tabelle 1.

	Bedarf	Kein Bedarf
Bereitschaft	Starkes Potenzial	Vorschlag generieren
Keine Bereitschaft	Kundenanalyse	Schwaches Potenzial

Tabelle 1: Cross-Selling-Potenzial-Matrix

Starkes Cross-Selling-Potenzial

Durch den gezielten Einsatz von Ressourcen kann das Cross-Selling-Potenzial mit dem Angebot von Produktbündeln ausgeschöpft werden. Dabei werden Produkte kundenindividuell so kombiniert, dass sie einen besonders hohen Nutzen aufweisen oder preislich sehr attraktiv gestaltet sind. Durch dieses Vorgehen werden Skaleneffekte genutzt und Wettbewerber kostengünstig unter Druck gesetzt.

Vorschlag generieren

In diesem Anwendungsfall müssen Banken auf ressourcensparende und kostengünstige Methoden zurückgreifen. „Produktvorschlagssysteme“ diverser Hersteller sind an dieser Stelle denkbar um Cross-Selling-Anreize zu schaffen. Da diese Systeme einen lernenden Charakter aufweisen, verbessern sich die Vorschläge durch die Rückmeldungen der Bankberater. Kommt ein Kunde auf seinen Bankberater zu, kann dieser die generierten Vorschläge an den Kunden weiterleiten.

Schwaches Cross-Selling-Potenzial

Liegt ein schwaches Potenzial vor, ist Cross-Selling nur sehr schwer möglich. Entweder sind die nötigen Informationen nicht bekannt oder der Kunde ist stark abwanderungsgefährdet. Dieser Fall ist isoliert von allen anderen Kombinationen der Matrix zu betrachten, da die Problematik durch Cross-Selling nicht beeinflusst oder gelöst werden kann. Cross-Selling dient hier nur als Hilfe um unzufriedene oder unprofitable Kundengruppen identifizieren zu können. Eine Lösung obliegt dem Kundenmanagement bzw. Service.

Kundenanalyse

Auch in diesem Bereich ist die Abwanderungsgefahr sehr hoch. Zwar sind diese Kunden an konkreten Produkten interessiert, wollen jedoch die aktuelle Geschäftsbeziehung aus unbekanntem Gründen nicht weiter vertiefen. Primäres Ziel muss es an dieser Stelle sein, die Gründe des Kunden kennen zu lernen um ihn nicht an einen unmittelbaren Konkurrenten zu verlieren. Der Kunde ist dem Service zu übergeben und einem entsprechenden Kundenrückgewinnungsprogramm zuzuordnen.

Maßnahmenplanung

Da die Finanzsituation eines Kunden ein sensibles Thema ist, ist mit einer entsprechenden Sorgfalt zu arbeiten. Aus diesem Grunde wird zunächst der zuständige Berater bzw. das entsprechende Vertriebsteam ermittelt. In diesem Schritt wird vor allem dem CRM-Paradigma „one face to the customer“ Rechnung getragen. Der Berater erstellt in Abhängigkeit des Cross-Selling-Potenzials kundenindividuelle oder automatisch generierte Produktvorschläge („Produktvorschlagssystem“). Durch diese menschliche Kontrollinstanz wird sichergestellt, dass der Kunde nicht durch rein automatische Prozesse mit unpassenden Angeboten „überflutet“ wird. Lehnt der Berater einen generierten Produktvorschlag ab, so sendet er diese Information zurück in die Analysephase.

Umsetzung

Liegen alle relevanten Informationen vor, kann unter Berücksichtigung der Konditionssätze des entsprechenden Kundensegments (z.B. A-, B-, C-Kunde) die Kalkulation des Angebots im CRM-System erstellt werden („Pricing“). Dieses Angebot wird dem Kunden über seinen favorisierten Kommunikationskanal (z.B. Telefon oder persönlicher Termin) angeboten und auf seine Rückmeldung gewartet. Nimmt er das Angebot an, so mündet das Cross-Selling-Angebot im Verkaufsprozess (Vertragsabwicklung) der Bank. Abschließend ist das Kundenprofil bzgl. des erfolgreichen Cross-Sellings zu aktualisieren. Aus diesen Informationen lässt sich die Cross-Selling-Bereitschaft ableiten und verkürzt somit den gesamten Cross-Selling-Prozess um die Bereitschaftsermittlung. Lehnt der Kunde das Ange-

bot hingegen ab, wird das Kundenprofil ebenfalls aktualisiert und das abgelehnte Angebot im Kontrollprozess separat und detailliert betrachtet. Dabei wird vor allem der Frage nachgegangen, ob das Angebot den Bedarf des Kunden verfehlt oder als unattraktiv gilt.

Kontrolle

Grundsätzlich lässt sich der Kontrollprozess aus zwei Perspektiven betrachten. Zum einen kann der Erfolg eines Kundensegments („Top-Down“) anhand von betriebswirtschaftlichen Kennzahlen wie z.B. Ertragsquote oder Deckungsbeitrag ermittelt werden, zum anderen kann ein konkretes Cross-Selling-Angebot („Bottom-Up“) hinsichtlich des Erfolgs (Annahmequote) bewertet werden. Wird bspw. ein Produktbündel in unterschiedlichen Segmenten ohne Erfolg angeboten, muss über eine Rekonstruktion der Produktkombination oder gar des gesamten Produkts nachgedacht werden.

3 Schlussbetrachtung

Das Zusammenspiel von einzelnen Aktivitäten in den CRM-Kernprozessen Marketing, Vertrieb und Service bildet die Basis, Cross-Selling koordiniert und effizient durchzuführen. Cross-Selling stellt in diesem Zusammenhang kein isoliertes Werkzeug dar. Vielmehr kapselt es verschiedene auf den Kunden ausgerichtete Maßnahmen. Durch die gezeigte prozessuale Verankerung können Banken ihre Leistungen nachhaltig auf den Kunden abstimmen. Nach den ersten Cross-Selling-Erfahrungen stellt sich bankintern ein Lerneffekt ein, mit dem zukünftige Aktivitäten besser koordiniert werden können. Die daraus resultierende „Cross-Selling-Kultur“ eröffnet Bankinstituten neue Perspektiven im Vertrieb.

Literatur

- [BNP08] Dieter Bartmann, Marco Nirschl und Anja Peters. *Retail Banking*. Frankfurt School Verlag, Frankfurt am Main, 1. Auflage, 2008.
- [Sch02] Heiko Schäfer. *Die Erschließung von Kundenpotentialen durch Cross-Selling: Erfolgsfaktoren für ein produktübergreifendes Beziehungsmanagement*. Deutscher Universitäts-Verlag, Wiesbaden, 1. Auflage, 2002.
- [TRS06] Torsten Tomczak und Elisabeth Rudolf-Sipötz. Bestimmungsfaktoren des Kundenwertes: Ergebnisse einer branchenübergreifenden Studie. In Bernd Günter und Sabrina Helm, Hrsg., *Kundenwert: Grundlagen - Innovative Konzepte - Praktische Umsetzungen*, Seiten 127–181. Gabler Verlag, 3., überarb. u. erw. Auflage, 2006.
- [Wit10] Georg Wittmann. *Cross-Selling von Finanzdienstleistungen*. Universitätsverlag Regensburg, Regensburg, 1. Auflage, 2010.
- [WKS03] Klaus-Peter Wiedmann, Alexander Klee und Nadine Siemon. Erfolgsfaktoren eines ertragsorientierten Zielkundenmanagement im Retail Banking. In Wiedmann Klaus-Peter, Alexander Klee, Holger Buxel und Frank Buckler, Hrsg., *Zielkundenmanagement für Finanzdienstleister: Innovative Strategien - Konzepte - Tools*, Seiten 55–72. Gabler Verlag, Wiesbaden, 1. Auflage, 2003.

Pattern-Katalog zur Etablierung von Vertrauen in Cloud Computing Technologien aus Sicht der Nutzer

Daniel Gaebler, Bastian Maucher

Hochschule Reutlingen

Fakultät Informatik

Vorname.Nachname@student.reutlingen-university.de

Art der Arbeit: Master Thesis

Betreuer/in der Arbeit: Prof. Dr. rer. nat. Alfred Zimmermann

Abstract: Das Potential von Cloud Computing ist vielversprechend, da Ressourcen stets verfügbar sind und sich zudem Hardware-, Software- und Wartungskosten einsparen lassen. Allerdings wird die neue Technologie nur mit einer gewissen Zurückhaltung genutzt: Im 3. Quartal 2011 von 12,5 % der kleinen und mittelständischen Unternehmen [HP11]. Fehlendes Vertrauen der Nutzer in die Sicherheit, die Zuverlässigkeit sowie in die Provider hat eine gewisse Skepsis erzeugt. Um Unternehmen eine sichere Cloud-Nutzung gewährleisten zu können, müssen sich diese mit den Risiken des Cloud Computing beschäftigen und gewisse Voraussetzungen erfüllen. Eine Ist-Analyse kann durch einen Fragenkatalog erfolgen, der die derzeitige Cloud-Fähigkeit ermittelt. Für die Verbesserung der Cloud-Fähigkeit werden Handlungsempfehlungen anhand eines Pattern-Katalogs ausgesprochen.

1 Einleitung

Viele Analysten sind sich in Bezug auf das enorme Potential von Cloud Computing einig. Das Thema Cloud Computing ist längst nicht mehr ein rein wissenschaftliches Thema oder ein Thema mit welchem man sich nur innerhalb von IT-Abteilungen beschäftigt. Beispielhaft ist hierfür ein Artikel der FAZ, die sonst eher für nicht-Technologie Themen und Artikel bekannt ist:

„Es gibt kaum ein Thema in der IT-Branche, das in der letzten Zeit so viel Aufmerksamkeit erregt hat wie Cloud Computing. Analysten wie Unternehmen sind sich einig: Cloud ist der nächste Paradigmenwechsel in der IT [...]“ [FA10]. Die Vorteile von Cloud Computing liegen auf der Hand, führt die FAZ weiter aus:

- Geringere Kapitalbindung
- Höhere Effizienz
- Verbesserte Kostenkontrolle und Wahlfreiheit beim Outsourcing
- Mehr Flexibilität bei der Anpassung der IT an veränderte Anforderungen

Problematisch wird das Thema Cloud Computing spätestens dann, wenn geschäftskritische, sensible oder auch private Daten in der öffentlichen Cloud verarbeitet werden [MA11], [SA10]. Fragen der Verfügbarkeit, Sicherheit und insbesondere des Vertrauens stehen in Folge dessen im Mittelpunkt:

- Was muss passieren, bevor ein Unternehmen seine IT-Infrastruktur und Anwendungen aus der Cloud bezieht?
- Welche abgeleiteten Anforderungen gibt es an die Compliance und die Sicherheit?

Das Vertrauen in die Provider und vor allem in die Sicherheit des Cloud Computing ist gering. Vor diesem Hintergrund ist es verständlich, dass Unternehmen bei der Einführung von Cloud Computing verunsichert und demzufolge zurückhaltend sind und die hohen Eintrittshürden scheuen [BE11].

2 Etablierung von Vertrauen in Cloud Computing

Bevor ein Unternehmen Cloud-Dienste in Anspruch nimmt, muss es sich gezielt mit den Vorteilen, den Herausforderungen sowie den Risiken des Cloud Computing auseinandersetzen und Erwartungen, Anforderungen sowie Voraussetzungen für die Einführung definieren. Ein von uns entwickelter Ansatz zur Etablierung von Vertrauen in die noch neuen Cloud Computing Technologien beinhaltet die Anwendung eines Fragenkatalogs zur Ermittlung der derzeitigen Cloud-Fähigkeit¹. Handlungsempfehlungen anhand von Patterns helfen anschließend das durch den Fragenkatalog identifizierte Potential, in Bezug auf die Cloud-Fähigkeit, zu heben. Die Patterns sind in einem Pattern-Katalog zusammengefasst.

2.1 Fragenkatalog

Der Fragenkatalog sollte von mindestens fünf Personen (Geschäftsführung, IT-, Compliance- und Sicherheits-Abteilung) beantwortet werden und besteht aus den drei folgenden Schwerpunkten:

- Fragen zur Compliance
- Fragen zur Nutzer-Sicherheit
- Fragen zum Risikomanagement

¹ Unter Cloud-Fähigkeit verstehen wir die Fähigkeit eines Unternehmens Cloud-Dienste in Einklang mit Vorgaben der Governance, Compliance und Security nutzen zu können.

Falls die Mehrheit der befragten Personen eine Frage positiv beantwortet, ist das Unternehmen in diesem Punkt Cloud-fähig. Wird eine Frage negativ beantwortet, besteht ein Handlungsbedarf, der durch die Anwendung der Patterns gedeckt werden kann.

2.2 Pattern-Katalog

Der von den Autoren entworfene Pattern-Katalog bietet Handlungsempfehlungen, wie sich die Anforderungen und Voraussetzungen für die Nutzung von Cloud-Diensten erfüllen lassen. Der gesamte Pattern-Katalog wird prototypisch durch ein Wiki abgebildet, um die Navigation durch die Patterns zu erleichtern.

Die Patterns werden in einer reduzierten kanonischen Form dargestellt. Dies bedeutet eine Beschränkung auf die wesentlichen Elemente, da es primär darum geht, die wissenschaftliche Methodik der Patterns anzuwenden und für den Nutzer eine transparente und übersichtliche Darstellung zu gewährleisten.

Der Pattern-Katalog zur Absicherung des Vertrauens der Nutzer ist in insgesamt sechs Hauptbereiche unterteilt. Zu jedem Hauptbereich gibt es weitere Patterns. Es werden Empfehlungen ausgesprochen, welche Voraussetzungen Unternehmen vor dem Cloud-Eintritt erfüllen müssen und was die Unternehmen während der Cloud-Nutzung beachten sollten.

Patterns zur Absicherung des Vertrauens der Nutzer	
1. Anforderungen vor dem Eintritt in die Cloud	4. Vertragliche Gestaltung
2. Gesetzliche Regelungen	5. Nutzer-Sicherheit
3. Interne Verpflichtungen	6. Risikomanagement

3 Fazit und Ausblick

Aus den hier dargestellten Überlegungen leiten sich folgende Ideen ab:

- Wenn der Cloud-Nutzer die hier vorgestellten Werkzeuge und Methoden nutzt, dann erreicht er eine höhere Transparenz in Bezug auf die Einführung von Cloud Computing Technologien in seinem Unternehmen.
- Wenn eine höhere Transparenz gegeben ist, steigt das Vertrauen in die neue Technologie.
- Wenn eine höhere Transparenz und Vertrauen gegeben sind, fördert dies die Erfolgsaussichten für die Einführung von Cloud Computing.
- Wenn Cloud Computing erfolgreich eingeführt worden ist, kann dies zu einer Optimierung der Unternehmensarchitektur im Ganzen führen.

Durch den von uns entwickelten Ansatz der Handlungsempfehlungen in Form eines Pattern-Katalogs, können sich Unternehmen besser auf den Einstieg in die Cloud vorbereiten. Indem sich Unternehmen mit den Chancen und Risiken des Cloud Computing auseinandersetzen, können sie Defizite erkennen und diese beheben. Dadurch wird nicht nur das Vertrauen in die Cloud-Technologie, sondern auch die Sicherheit im eigenen Unternehmen, vor allem in Bezug auf die Cloud-Nutzung erhöht. Für die für dieses Paper zugrunde liegende Masterarbeit bot sich bei einem Forschungsseminar die Gelegenheit zu einer ersten Diskussion der Ergebnisse, in Bezug auf die Inhalte des Fragebogens als auch der Patterns.

Der Fokus lag dabei gezielt auf den Folgeaktivitäten, die sich im Anschluss an die Arbeit durch weitere Abschlussarbeiten oder Projektgruppen bearbeiten lassen können. Konkret wurden folgende Punkte identifiziert:

- Problem- /Lösungsformulierungen schärfen: Vor allem wenn die Lösung aus mehreren Punkten besteht, ist dies ein erstes Indiz für Folge-Patterns.
- Skalen beim Fragebogen standardisieren: Im Sinne von Reifegraden müssen auch die Skalen des Fragebogens standardisiert werden.
- Rollen der Betroffenen bzw. der Beteiligten in kanonische Form bringen
- Verbindung zum ESARC: Die vorliegende Thesis kann als Ausgangspunkt für die Einführung einer „Cloud-Architektur“ innerhalb des ESARC dienen. Ziel ist hierbei zum einen die Schaffung solch einer Architektur mit ihren speziellen Domänen, zum anderen die Ausrichtung der Patterns und des Fragenkatalogs an diesen Domänen.
- Pattern Language Management System: Die hier erarbeiteten Patterns können in ein PLMS, ähnlich zu dem in [RE11] skizzierten, eingepflegt werden. Auf diese Weise sind sie einer größeren Gruppe von Personen zugänglich und es wird ein kollaboratives Arbeiten an den Patterns möglich.

Literatur

- [BE10] Beuchat, Alain; Ziegler, Pius (2011): Cloud Computing ermöglicht Paradigmenwechsel. Online verfügbar unter http://www.kpmg.com/CH/de/Library/KPMG-in-the-Media/Documents/pr_2011_Cloud-Computing_KPMG_Beuchat_Ziegler.pdf, zuletzt aktualisiert am 26.01.2011, zuletzt geprüft am 06.01.2012.
- [HP11] Perspektive Mittelstand (2011): HP Cloud Index. <http://www.perspektive-mittelstand.de/HP-Cloud-Index-Cloud-Computing-Nutzung-bei-KMU-nimmt-weiter-zu/management-wissen/4342.html>, abgerufen am 23.12.2011.
- [MA11] Dr. Materna, Winfried (2011): Cloud Computing - Paradigmenwechsel statt Hype. NetSkill AG. <http://www.competence-site.de/software-as-a-service-saas/Cloud-Computing-Paradigmenwechsel-statt-Hype>, abgerufen am 23.11.2011.
- [RE11] Reiners, R.; Astrova, I.; Zimmermann, A (2011): Introducing new Pattern Language Concepts and an Extended Pattern Structure for Ubiquitous Computing Application Design Support. In (IARIA Rom Hrsg.): Patterns 2011, The Third International Conferences on Pervasive Patterns and Applications, Rome, 2011; S. 61-66
- [SA10] SaaS-Magazin - Messe& Kongresse Verlags GmbH (2010): Cloud Computing: Paradigmenwechsel in der IT. <http://www.saasmagazin.de/saasondemandmarkt/unternehmen/muenchner-kreis120110.html>, abgerufen am 23.11.2011.

KPIs für die strategische Business/IT Alignment-Quantifizierung

Sebastian Grunow, Matthias Lenk und Sascha Roth

Technische Universität München
Software Engineering of Business Information Systems (sebis),
{grunow, lenk, sascha.roth}@in.tum.de

Abstract: Zahlreiche Studien belegen die kritische Bedeutung einer optimalen Abstimmung zwischen Business und IT als Schlüsselfaktor für ein effizientes, effektives und vor allen wettbewerbsfähiges Unternehmen. Grundvoraussetzung um dies zu erreichen sind objektive Messgrößen zur Gewährleistung eines zielgerichteten Vorgehens. In diesem Papier schlagen wir unter Verwendung von Enterprise Architecture Modellen als Informationsbasis ein Framework an KPIs zur objektiven Quantifizierung des derzeitigen Business/IT Alignment vor.

1 Einführung und Motivation

Eine erfolgreiche Abstimmung zwischen Business und IT wird häufig assoziiert mit einer Verbesserung der Effektivität und Effizienz des Geschäftes sowie mit einer erfolgreichen Erschließung des Potentials von IT-Investitionen [HV93], [SPE06], [SC01], [Woo93]. Trotz zahlreicher Vorteile sind Ansätze zur Erreichung eines solchen strategischen Business/IT Alignment (SBITA) häufig unklar oder nur bedingt praktisch einsetzbar [HH05]. Beginnend beim Strategic Alignment Model von Henderson et al. [HV93] finden sich in der Literatur eine Vielzahl an Frameworks, Methoden und Modelle, die dieses Problem adressieren. Alignment ist dabei jedoch nicht als endgültiger Zustand zu verstehen, sondern vielmehr als Evolutionsprozess [HH05], welcher objektive Messgrößen erfordert um ein zielgerichtetes Vorgehen zu ermöglichen. Bei den wenigen Ansätzen, die sich mit Messmechanismen auseinandersetzen, erfolgt die Quantifizierung i.d.R. anhand der Wahrnehmung der Mitarbeiter, welche in Umfragen und Interviews einerseits befragt werden und andererseits ausgewählte Bereiche wie die IT/Business Kommunikation numerisch oder qualitativ bewerten sollen [Ven89], [RB00], [WA04], [CKH02], [GOS08], [RHR08]. Ein solches Vorgehen ist durch einen hohen Grad an Subjektivität gekennzeichnet, welcher ein zeitliches als auch ein unternehmens- sowie geschäftsbereichübergreifendes in Beziehung setzen nur bedingt gewährleistet. In anderen Fällen sind die angegebenen Indikatoren nur bedingt objektiv messbar. So empfiehlt beispielsweise Cobit 4.1, ein Framework für IT Governance, folgende Messgröße: „*Degree of compliance with business governance requirements*“ [Itg07]. In diesem Papier schlagen wir ein Framework von Key Performance Indicators (KPIs) zur objektiven, unternehmensunabhängigen Quantifizierung des derzeitigen SBITA vor. Grundlage dafür bilden Unternehmensarchitekturen (Enterprise Architecture), welche eine holistische Sicht auf das Unternehmen einnehmen und die wesentlichen Business und IT

Aspekte beschreiben [Sch06]. Unter Verwendung von ArchiMate [Lan09], einer formalen, industrieübergreifend anerkannten Modellierungssprache, wird einerseits eine unternehmensübergreifende Verwendbarkeit gewährleistet und andererseits eine hohe Objektivität sichergestellt. Gleichzeitig ermöglicht der formale Charakter eine tool-gestützte Berechnung. Die KPIs gestatten nicht nur Fortschritte zu erkennen sondern auch Zukunftsszenarien hinsichtlich ihrer Auswirkungen auf SBITA zu bewerten.

2 Business/IT Alignment Framework

Eine anerkannte Aussage von DeMarco [Dem86] lautet „*you can't control what you can't measure*“ und wird häufig ergänzt mit „[...] *you cannot measure what you cannot define*“. Entsprechend wurde im Rahmen der Analyse im ersten Schritt eine formale Beschreibung von SBITA unter Verwendung vorhandener allgemein anerkannter Modelle entwickelt. Anhand der identifizierten Charakteristika wurden im zweiten Schritt geeignete KPIs abgeleitet.

Strategisches Business/IT Alignment: Die verschiedenen Ansätze in der Literatur befassen sich grob mit drei Eigenschaften von SBITA, welche Anhaltspunkte für eine mögliche Quantifizierung bilden: Einflussfaktoren auf SBITA [HH05], [RB00], der Zustand SBITA [HV93] und Auswirkungen einer erfolgreichen/fehlenden SBITA [SC01], [Luf00], [CHB97], [CKH02]. Abbildung 1 illustriert ein Modell, welches verschiedene Ansätze miteinander in Beziehung setzt und die Grundlage unserer KPI Ermittlung bildet.

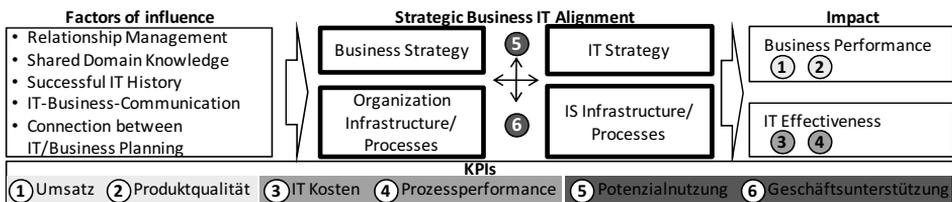


Abbildung 1:Strategic Business/IT Alignment Charakteristika und zugewiesene KPIs

Nach Reich et al. [RB00] und deren Erweiterungen ([HH05], [RHR08]) nehmen vor allem fünf Aspekte Einfluss auf SBITA: (1) das Beziehungsmanagement, (2) eine einheitliche Wissensbasis, (3) eine erfolgreiche IT-Historie, (4) eine optimale Kommunikation sowie (5) Planung zwischen beiden Seiten. Hierbei handelt es sich v.a. um soziale Aspekte, die einerseits schwer objektiv zu quantifizieren sind und andererseits nicht im Rahmen von ArchiMate modelliert werden und entsprechend für eine Quantifizierung nur bedingt geeignet sind. Das wohl anerkannteste Modell zur Beschreibung von SBITA findet sich bei Henderson et al. [HV93], welches Alignment als multilaterale Abstimmung von vier Domains auffasst: (1) *Business Strategy*, (2) *IT Strategy*, (3) *Organizational Infrastructure and Processes* und (4) *IS Infrastructure and Processes*. Während aufgrund des facettenreichen Charakters des Begriffs Strategie eine Messung der ersten beiden Domains nur eingeschränkt möglich ist [Ham80], [Tal07], ist

die formale Beschreibung der organisatorischen Infrastruktur und IS Infrastruktur einschließlich der Prozesse sowie deren Interaktion ein wesentlicher Aspekt des Unternehmensarchitektur Konzeptes und wird im Rahmen von ArchiMate ausführlich modelliert. Hinsichtlich der Auswirkungen von SBITA verweisen zahlreiche Studien auf das Potential im Hinblick auf die Geschäftsperformance wie Umsatz und Profitabilität [SC01], [CKH02] sowie auf die IT-Effektivität wie IT-Kosten und Durchlaufzeiten [CKH02], [CHB97]. Für beide Aspekte finden sich in der Literatur zahlreiche KPIs.

Key Performance Indicators: In Anbetracht der vorhergehenden Ergebnisse erfolgt eine Quantifizierung von SBITA einerseits direkt anhand der Bewertung der Abstimmung zwischen organisatorischer und IS Infrastruktur und Prozesse sowie indirekt über die Auswirkungen von SBITA auf die Business Performance als auch die IT-Effektivität. Um ein ganzheitliches Bild zu gewährleisten werden zu den einzelnen Bereichen jeweils mehrere KPIs definiert. Gleichzeitig liegen die Messgrößen auf verschiedenen Granularitätsebenen vor, um nicht nur auf Unternehmensebene eine Aussage über SBITA zu ermöglichen sondern auch für feingranulare Bestandteile des Unternehmens, z.B. Geschäftsbereiche. Chan et al. [SC01], [CHB97] zeigen anhand umfangreicher Studien eine Korrelation zwischen SBITA und mehreren Business Performance KPIs wovon sich *Umsatz* und *Produktqualität* anhand von ArchiMate bestimmen lassen. Jedem einzelnen ArchiMate *Produkt* wird dazu ein Attribut *generierter Umsatz* zugeordnet, welches unter Verwendung der Durchschnittsfunktion auf höhere Granularitätsebenen (Geschäftsbereiche und Unternehmensebene) propagiert werden kann (*KPI 1*). Ähnliches gilt für die zweite Größe, im Rahmen derer einem *Produkt* die *durchschnittliche Anzahl an Qualitätsmängel* zugewiesen wird (*KPI 2*). Im Hinblick auf die IT-Effektivität belegen [CKH02], [CHB97] eine positive Korrelation zwischen SBITA und *IT-Kosten* sowie der *Prozessperformance*. Um dies zu quantifizieren werden den *Geschäftsprozessen* einerseits die verbundenen *IT Kosten* (*KPI 3*) zugeordnet und andererseits die *Prozesszeit (Processing Time)* (*KPI 4*). Eine positive bzw. negative Entwicklung von SBITA führt zu einer Erhöhung/Verringerung der Größen. Der letzte Aspekt, die Abstimmung zwischen organisatorischen und IT Aspekten, entspricht im Rahmen von ArchiMate der Beziehung zwischen den Architekturen Business und Application. Während *KPI 5* hierbei die Potentialnutzung der Geschäftsseite anhand des Anteils verwendeter *IT Services im Verhältnis zur Gesamtanzahl an bereitgestellten Services* bestimmt, betrachtet *KPI 6* die Geschäftsunterstützung durch die IT. Letztere wird dabei anhand fehlerhafter *IT Transaktionen* im Rahmen von Geschäftsprozessen quantifiziert. Hierzu werden die Geschäftsprozessobjekte um das Attribut *Anzahl verbundener fehlerhafter IT Transaktionen* erweitert.

3. Zusammenfassung und Ausblick

Um eine zielgerichtete SBITA Verbesserung zu gewährleisten, wird im Rahmen des vorliegenden Papiers ein Framework an KPIs zur Quantifizierung von SBITA vorgeschlagen. Dabei werden verschiedene Aspekte des Konzeptes SBITA gemessen um ein ganzheitliches Bild zu gewährleisten. Derzeit werden die entwickelten Messgrößen in Zusammenarbeit mit einem IT Consulting Unternehmen praktisch evaluiert.

Literaturverzeichnis

- [CHB97] Chan, Y.E. et.al. Business Strategic Orientation, Information Systems Strategic Orientation, and Strategic Alignment, *Information Systems Research*, 1997, S. 125–150.
- [CKH02] Cragg, P. et.al.: IT alignment and firm performance in small manufacturing firms. In: *The Journal of Strategic Information Systems*. Elsevier, 2002; S. 109-132
- [Dem86] DeMarco, T.: *Controlling Software Projects: Management, Measurement, and Estimates*. New York, 1986.
- [Ham80] Hambrick, D. C.: Operationalizing the concept of business-level strategy in research. In: *Academy of Management Review*. Academy of Management, 1980; S. 567-575
- [GOS08] Gutierrez, A. et.al.: Using tactical and operational factors to assess strategic alignment: an SME study. In: *Information Systems Journal*. 2008; S. 1-10
- [HH05] Hu, Q.; Huang, D.: Aligning IT with Firm Business Strategies Using the Balance Scorecard System. In: *Hawaii International Conference on System Sciences*. Ieee, 2005; S. 1-10
- [HV93] Henderson, J. C.; Venkatraman, N.: Strategic alignment: Leveraging information technology for transforming organizations. In: *IBM Systems Journal*. IBM Corp., 1993; S. 472-484
- [Itg07] Itgi: *Governance Institute: Cobit 4.1*. 2007.
- [Lan09] Lankhorst, M.: *Enterprise Architecture at Work*. Springer, Berlin, 2009.
- [Luf00] Luftman, J.: Assessing business-IT alignment maturity. In: *Business*. Igi Global, 2000; S. 1-51
- [RB00] Reich, B. H.; Benbasat, I.: Factors that influence the social dimension of alignment between business and information technology objectives. In: *MIS Quarterly*. MIS Quarterly & The Society for Information Management, 2001; S. 81-113
- [RHR08] Weiss, J. W.; Anderson, D.: Strategic IT Alignment in Swedish Public Healthcare System. In: *Proceedings of the 1st world summit on The Knowledge Society: Emerging Technologies and Information Systems for the Knowledge Society*. Springer-Verlag, 2008; S. 105-113
- [SC01] Sabherwal, R.; Chan, Y. E.: Alignment Between Business and IS Strategies: A Study of Prospectors, Analyzers. In: *Information Systems Research*. MIS Quarterly & The Society for Information Management, 2001; S. 11-33
- [Sch06] Schekkerman, J.: *How to Survive in the Jungle of Enterprise Architecture Frameworks: Creating or Choosing an Enterprise Architecture Framework*. Bloomington: Trafford Publishing, 2006.
- [SPE06] Silva, E. et.al.: Strategic Business and IT Alignment: A Prioritized Theory Diagram. In: *2006 Technology Management for the Global Future PICMET 2006 Conference*. Ieee, 2006; S. 1-8
- [Tal07] Tallon, P. P.: A Process-Oriented Perspective on the Alignment of Information Technology and Business Strategy. In: *Journal of Management Information Systems*. M.E. Sharpe Inc., 2007; S. 227-268
- [WA04] Weiss, J. W.; Anderson, D.: Aligning technology and business strategy: issues & frameworks, a field study of 15 companies. In: *37th Annual Hawaii International Conference on System Sciences*. Ieee, 2004; S. 220-229
- [Woo93] Woolfe, R.: *The path to strategic alignment*. In: *Information Strategy*. Auerbach Publishers Inc., 1993; S. 13-23
- [Ven89] Venkatraman, N.: Strategic Orientation of Business Enterprises. In: *Management Science*. 1989; S. 942-962

Integration ereignisbasierter Middleware-Systeme in kontextbasierte Geschäftsprozesse

Richard Günther

Universität Hamburg
Department Informatik, Arbeitsbereich VSIS
riguenth@web.de

Art der Arbeit: Masterarbeit

Betreuer/in der Arbeit: Prof. Dr. Böhmann, Prof. Dr. Lamersdorf

Abstract: Die vorliegende Arbeit untersucht, wie sich die physische Umwelt in Form von Ereignisdaten in Geschäftsprozesse integrieren lässt. Im Gegensatz zu proprietären, speziell zugeschnittenen Systemen ermöglicht BPMN 2.0, dieses Vorhaben auf der Grundlage von Standards und unter Verwendung generischer Ereignisdaten durchzuführen. Diese Verbindung zwischen der Ebene der Informationssysteme und der Geschäftsprozesse ermöglicht, zeitnah auf relevante Geschäftsereignisse zu reagieren.

1 Motivation

Sinkende Produktionskosten begünstigen den zunehmenden Einsatz von Sensoren, die Umgebungsinformationen (*Kontext*) digital erfassen und so die physische und virtuelle Welt verbinden – beispielsweise durch die Erfassung von Temperaturen, Wasserständen oder der Ankunft eines Containers. Im Zuge der steigenden Vernetzung sehen sich Unternehmen einer wachsenden Menge von geschäftsrelevanten Ereignissen ausgesetzt, auf die im Sinne der Wettbewerbsfähigkeit möglichst schnell zu reagieren ist [Luc01]. Derzeit kann die Integration von Ereignisdaten in Geschäftsprozesse jedoch nur mit speziell zugeschnittenen, proprietären Systemen erfolgen – die generische Bereitstellung von Ereignisdaten ist bislang nicht möglich. Beispielsweise führten Kiziltoprak et al. das Beispiel des *Smart Containers* vor, bei dem ein Container eigene Sensoren hat und diese Messwerte am Umschlagsplatz mitteilt [KSH+08]. So werden die zugrundeliegenden Prozesse beschleunigt – allerdings bislang auf Basis proprietärer, oft nur technisch dokumentierter Lösungen. Abb. 1 zeigt eine solche beispielhafte Architektur auf.

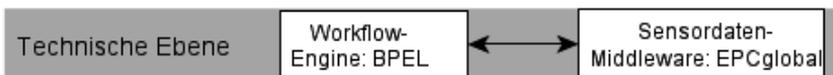


Abbildung 1: Technische, proprietäre Architektur zur Integration von Ereignisdaten in Prozesse

Mit *BPEL* [OA-1] stellt OASIS eine Sprache bereit, mit der sich Geschäftsprozesse beschreiben und ausführen lassen. Diese sind von der jeweiligen Middleware, hier z.B. EPCglobal, abhängig und darüber hinaus für eine abstrahierte Darstellung auf fachlicher Ebene zu detailliert. Im Rahmen des *Business-IT-Alignments* werden Geschäftsprozesse jedoch in immer stärkerem Maße zur Orchestrierung von IT-Systemen verwendet [GI08]. Mit Einführung von BPMN 2.0 durch die OMG [OMG] existiert für beide Ebenen eine gemeinsame, standardisierte Sprache, sodass nun zwei Ziele in greifbare Nähe rücken:

1. Fachlich abstrahierte Modellierung von Ereignisdaten
2. Unabhängigkeit zwischen Workflow-Engine und Sensordaten-Middleware

Dafür sind die in Abb. 2 gezeigten Integrationsaufgaben zu lösen.

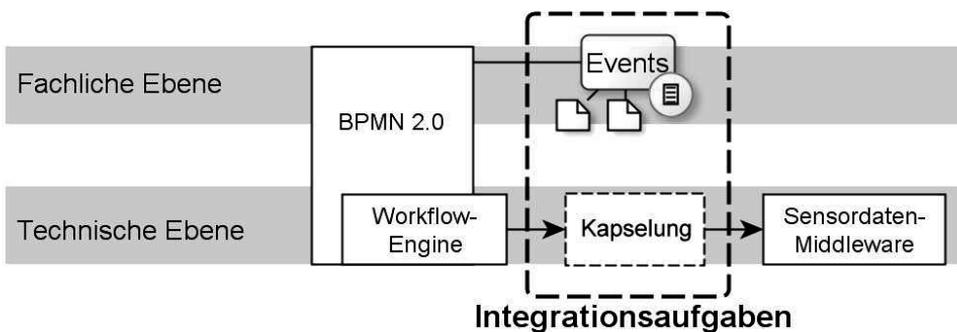


Abbildung 2: Integrationsaufgaben fachliche Modellierung und technische Systemunabhängigkeit
 Prozesse werden nun mit BPMN 2.0 modelliert und können sowohl auf fachlicher als auch technischer Ebene verwendet werden. Die Unabhängigkeit zwischen Workflow-Engine und Sensordaten-Middleware kann nur durch Kapselung der proprietären Schnittstellen gewährleistet werden. Im nachfolgenden Abschnitt wird eine Lösung für beide Integrationsaufgaben vorgestellt.

2 Integrationskonzept

Anhand typischer Geschäftsprozesse wurden fachliche Anforderungen an die Integration von Sensordaten ermittelt. Im Falle des Containers ist z.B. die Temperatur während der Reise von Interesse. Grundsätzlich lassen sich folgende, abstrakte Anforderungen identifizieren: 1) Benachrichtigen bei Eintritt von Ereignissen, 2) Abfragen von Ereignisdaten, 3) Auffinden von Sensoren und 4) das Steuern von Sensoren. Im Rahmen der Arbeit wurden nur die ersten beiden Punkte behandelt. Für BPMN mit seiner ablauforientierten Struktur fehlt bislang ein verhaltensbasierter Ansatz zur Beschreibung ereignisbasierter Systeme. Entsprechend erfolgt die Modellierung von Ereignisdaten semi-explizit, also ausschließlich unter Verwendung standardisierter Notationselemente. Ereignismuster werden beispielsweise gemäß der Relationenalgebra [Me04] mit drei

Textblöcken für die drei Select-, From- und Where-Klauseln modelliert. Das ist nötig, um proprietäre Erweiterungen der jeweiligen Workflow-Engine zu vermeiden.

Auf technischer Ebene wurde eine systemunabhängige Beschreibung in Form von *Meta*-Funktionen, -Protokollen und -Daten erarbeitet, indem für jedes der untersuchten Sensordaten-Middleware-Systeme [BL11, EPC08, BEJ11, EPC09] auf Grundlage des *Agents-and-Artefacts-Paradigma* (Ricci et al., [OPR+09, RVO08]) die jeweiligen *Funktionen*, *Protokolle* und *Daten* identifiziert wurden. Ein Proxy realisiert den generischen Zugriff auf die Sensordaten-Middleware auf Grundlage der *Meta*-Schnittstellen. Van der Aalst et al. [VTK+03, RTV+06] identifizierten Workflow-Pattern, auf deren Grundlage Modellierungskonstrukte zur Modellierung typischer Ereignisdaten erarbeitet wurden. Abb. 3 zeigt ein triviales Modellierungsbeispiel für den Abruf von Ereignisdaten.

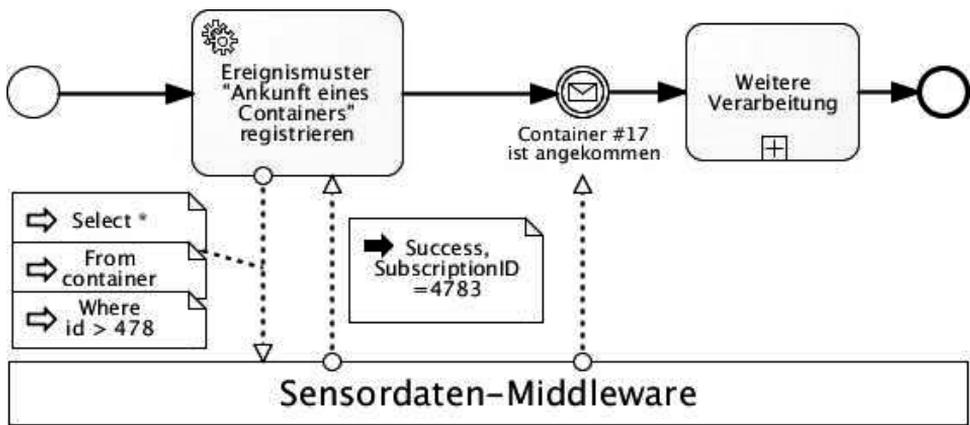


Abbildung 3: Triviales Modellierungsbeispiel für den Abruf von Ereignisdaten

Auf diese Art und Weise lassen sich alle identifizierten *Meta*-Funktionen modellieren. Zum Preis der semi-expliziten Modellierung wird die mit BPMN 2.0 eingeführte Interoperabilität der Prozesse sowie deren Unabhängigkeit von der ausführenden Workflow-Engine erhalten. Im letzten Abschnitt werden Risiken und Chancen der Lösung skizziert.

4 Fazit

Einschränkungen. Die vorgenommene Abstraktion der Funktionen, Protokolle und Daten der Sensordaten-Middleware-Systeme kann für spezielle Anwendungsfälle unpassend sein – z.B. wenn auf technischer Ebene eine Nähe zu Schnittstellen des *Sensor Web Enablements* (SWE) vorliegt. Desweiteren ist bislang offen, wie eine Verifizierung der Ereignisdaten erfolgen kann.

Chancen. Mit der vorgeschlagenen Lösung wurde die Interoperabilität zwischen Prozessen und den orchestrierten IT-Systemen auf Systeme zur Bereitstellung von

Ereignisdaten ausgeweitet. Im Ergebnis erscheint die Vision eines ganzheitlichen Geschäftsprozessmanagements, das die technische, fachliche und strategische Ebene umfasst, etwas realistischer [FRH10, Ge98].

Literaturverzeichnis

- [BEJ11] Bröring, A.; Echterhoff, J.; Jirka, S.: New Generation Sensor Web Enablement, *Sensors*, 2011, *11*, 2652-2699
- [BL11] Bade, D.; Lamersdorf, W.: An Agent-Based Event Processing Middleware for Sensor Networks and RFID Systems. *Comput. J.* 54(3): 321-331 (2011)
- [FRH10] Freund, J.; Rücker, B. & Henninger, T.: Praxishandbuch BPMN inklusive BPMN 2.0 *Hanser*, 2010
- [Ge98] Gehring, H.: Betriebliche Anwendungssysteme, Kurs 0825, Prozessorientierte Gestaltung von Informationssystemen, 1998
- [EPC08] EPCglobal: Application Level Events 1.1(ALE 1.1) Overview, 2008
- [EPC09] GSN: Global Sensors Networks, 2009
- [GI08] Business IT alignment: Trends im Software und Servicemarkt ; Tagungsband zur Teilkonferenz im Rahmen der Multi-Konferenz Wirtschaftsinformatik 2008, 26. Februar 2008 in München
- [KSH+08] Kiziltoprak, T.; Schumann, R.; Hahn, A. & Behrens, J. Kreowski, H.-J.; Scholz-Reiter, B. & Haasis, H.-D. (Eds.): Distributed Process Control by Smart Containers, *Dynamics in Logistics*, Springer, 2008, 321-328-328
- [Luc01] Luckham, D.: The Power of Events: An Introduction to Complex Event Processing in Distributed Enterprise Systems, *Addison-Wesley*, 2001
- [Me04] Meier, A.: Relationale Datenbanken: Leitfaden für die Praxis, *Springer*, 2004
- [OA-1] OASIS: BPEL 2.0-Spezifikation, 2007, URL: <http://docs.oasis-open.org/wsbpel/2.0/OS/wsbpel-v2.0-OS.html>
- [OMG] Object Management Group, Spezifikationsdokument zur BPMN 2.0, URL: <http://www.omg.org/spec/BPMN/2.0/PDF>
- [OPR+09] Omicini, A.; Piunit, M.; Ricci, A.; Viroli, M.: Agents, Intelligence and Tools, *Artificial Intelligence*, 2009, *LNAI 5640*, 157-173
- [RVO08] Ricci, A.; Viroli, M. & Omicini, A.: The A&A Programming Model and Technology for Developing Agent Environments in MAS, *ProMAS 2007*, 2008, *LNAI 4908*, 89-106
- [RTV+06] Russell, N.; Ter Hofstede, A. H. M.; Van Der Aalst, W. M. P. & Mulyar, N. Workflow control-flow patterns: A revised view, *BPM Center Report BPM0622* *BPMcenter org, Citeseer*, 2006, *2*, 06–22
- [VTK+03] Van Der Aalst, W. M. P.; Ter Hofstede, A. H. M.; Kiepuszewski, B.; Barros, A. P.: Workflow Patterns, *Distrib. Parallel Databases*, *Kluwer Academic Publishers*, 2003, *14*, 5-51

Analyse der Fertigungstiefe bei der Fachkonzeptmodellierung im Umfeld der Billinganwendung Abili-T

Daniel Roppelt

Hochschule Darmstadt
Fachbereich Informatik
Kooperativer Studiengang Informatik
daniel.roppelt@stud.h-da.de

Art der Arbeit: Praxisprojekt im Rahmen des Kooperativen Studiengangs Informatik
Betreuerin der Arbeit: Prof. Dr. Inge Schestag

Abstract: Die Studienarbeit beinhaltet die Analyse und Bewertung der Beschreibungstiefe einer vorhandenen Fachvorgabe zur Softwareerstellung im komplexen und unternehmenskritischen Bereich Rating & Billing. Mit der Einführung einer neuen Modellierungsmethodik muss sich der Betrieb mit neuen Prozessen, Vorgaben und einem neuen Modellierungswerkzeug auseinandersetzen. Neben der Abgrenzung zum technischen Konzept soll die Frage nach einem sinnvollen Schnitt (klar und verständlich für Fachseite, eindeutige Vorgabe für Realisierer) in Auslegung auf die neue Methodik für das System beantwortet werden.

1 Einführung

Die Fachvorgabe für das Billingssystem wurde nach einer im Unternehmensbereich entwickelten Methodik zur Softwareerstellung konzeptioniert. Die bewährte Methode prägt das Fachkonzept in seiner heutigen Form. Mit den Jahren gewachsen umfasst dieses weit über 100 Modelle und spezifiziert die komplexen Vorgänge der unternehmenskritischen Domäne Billing sehr tief. Die Vielseitigkeit der IT-Landschaft im Konzern sowie die gesammelten Erfahrungen bei der Beauftragung vergangener Software-Realisierungen sind Gründe für ein sehr tiefes und in Teilen auch sehr technisch beschriebenes Abstraktionsniveau der fachlichen Vorgabe. Durch Umstrukturierungen im betrieblichen Umfeld übernimmt der Bereich eine konzernweite Modellierungsmethode und muss sich künftig mit neuen Prozessen, Richtlinien sowie Vorgaben und einem neuen Modellierungswerkzeug auseinandersetzen. Diese rein fachliche Beschreibung der Anforderungen soll dem Realisierer mehr Verantwortung als bisher übertragen. Im Hinblick auf diese Veränderungen muss die derzeitige Fertigungstiefe für das Billingssystem betrachtet und für eine Anwendung der neuen Methode ebenso bewertet werden. Dabei müssen allerdings auch die unterschiedlichen Spezifika der einzelnen Unternehmensbereiche bei der konzernweiten Vorgehensweise beachtet werden – gerade besonders bei Rating & Billing.

Aufgabe der Arbeit ist die Analyse der derzeitigen Vorgabentiefe sowie eine anschließende, mit dem Kunden abgestimmte, Bewertung der Differenzen. Eine Aussage hinsichtlich der Beschreibungstiefe einer Fachvorgabe mit der neuen Methode soll getroffen werden können, welche mit externen Quellen [Ba00, Oe06, Wi05, Sc07] ergänzt und belegt werden soll. Wo kann ein sinnvoller „Schnitt“ bei der Modellierung der fachlichen Anforderungen gefunden werden? Fest steht: Eine allgemeingültige Lösung für diese Fragestellung wird es nicht geben! Projektspezifisch sind somit die Konventionen beider Methodiken zu analysieren und zu bewerten.

2 Durchführung

Die Arbeit setzt sich umfangreich mit der Problemstellung auseinander und teilt sich in zwei grundlegende Kapitel. Als Ausgangsbasis für das Projekt ist eine weitreichende Einarbeitung und Auseinandersetzung mit der Thematik¹ ausschlaggebend. Im ersten Teil sollen die Richtlinien analysiert und Differenzen bei der Modellierung aufgezeigt werden. Eine Empfehlung zeigt Verbesserungspotential anhand der nutzbaren Elemente und Konzepte der UML [JRH04, Ru07] als eingesetzte Beschreibungssprache auf. Eine Konkretisierung der Methodik, die die Modellierungstiefe [FHR08] bis zur Übergabe zur Realisierung an den Lieferanten beschreibt, ist im zweiten Kapitel näher beschreiben.

2.1 Analyse der Modellierungsrichtlinien und Fertigungstiefe

Für eine solche Analyse und anschließender Beurteilung kommt man nicht umher, beide Modellierungsmethoden miteinander zu vergleichen und konkrete Differenzen in den Richtlinien aufzuzeigen. Bei der Betrachtung wird exemplarisch ein besonders gut geeignetes Modell herausgegriffen, anhand dessen die Analyse für das System durchgeführt wird. Eine komplette Bewertung des umfangreichen Konzeptes wäre im Projektzeitraum nicht umsetzbar.

Der erstellte Konventionenkatalog sammelt ca. 150 wichtige Richtlinien der neuen Methode und gruppiert diese in zehn Kapitel. Er bildet die Grundlage für die Analyse der Modellierungsrichtlinien und der Fertigungstiefe. In einem ersten Schritt werden die Abweichungen zwischen Fachkonzept und der neuen Methodik auf Modellierungsebene dokumentiert. Die vorgegebenen Konventionen erreichen dabei selbst einen hohen Detaillierungsgrad, so dass auch kleine Differenzen erkannt und anschließend bewertet werden können. Kurze Kommentare verbessern zusätzlich die Nachvollziehbarkeit der Einstufungen.

Richtlinien
Allgemeine Modellierung
Fachspezifische Modellierung
Befüllungsgrad
Benennen von Elementen
Layout von Diagrammelementen
Dokumentieren von Elementen
Schneiden von Elementen
Strukturieren von Elementen
Konsistenzsicherung
Verlinkung von Modellelementen

Abbildung 1: Inhaltsstruktur des Konventionenkatalogs

¹ Input liefern zahlreiche betriebsinterne Dokumentationen und Quellen

Ebenso wichtig ist es hierbei, die auftretenden „Baustellen“ bei der Modellierung genauer zu betrachten. Durch die neuen Prozesse und ein neues Werkzeug bei der UML-Modellierung ergeben sich zahlreiche Änderungen in der Vorgehensweise. Die Arbeit betrachtet bis zu einem gewissen Grad auch diese Aspekte.

2.1 Beurteilung und Konkretisierung der Methode

Im zweiten Teil der Arbeit werden die erkannten Differenzen in drei Kategorien untergliedert:

- allgemeiner Modellierungshinweis
- Unterschiede in der Methodik
- konkretes Verbesserungspotential

Hinweise haben hierbei eher allgemeingültigen Charakter und werden als Anmerkungen mit in die Empfehlung aufgenommen; wohingegen Methodikunterschiede zu bewerten sind und in Auslegung an das Billingsystem eine Empfehlung auszusprechen ist. Richtlinien, die abweichend zur Methode modelliert sind, werden als Verbesserungspotential markiert.

Die Empfehlung greift die bewerteten Differenzen auf und beschreibt anhand der nutzbaren Elemente und Konzepte der UML [JRH04, Ru07] geeignete Notationen, um vor allem der sehr ausführlichen textuellen Beschreibung sowie der Beschreibungssprache gegenzuwirken. Gerade diese beiden Punkte fallen bei der Analyse negativ auf.

Ein weiterer wichtiger Punkt sind die Veränderungen im IT-Bereich hinsichtlich Vorgehen und Prozesse. Die neuen Prozesse werden im Tool abgebildet, das über eine Exportfunktion eine automatisierte Dokumentengenerierung für eine Fachvorgabe gemäß der Methode unterstützt. Die ausgesprochene Empfehlung kann hierbei als eine Art „Hebel“ verstanden werden, um sich auf die neue Methodik im Unternehmensbereich auszurichten. Eine genauere Spezifikation der modellierungsrelevanten Kapitel der erstellen Dokumente sowie die Definition eines sinnvollen Schnitts im Prozessvorgang bilden die Abrundung des erarbeiteten Management Summarys, das die Ergebnisse (siehe Kapitel 3) zusammenfassend aufbereitet.

3 Ergebnisse & Ausblick

In Bezug auf die beiden Methoden werden die Modellierungsrichtlinien analysiert und konkrete Differenzen aufgezeigt. Diese Unterschiede werden anschließend, anhand eines mit den Kunden abgestimmten Kriterien-Set, bewertet. Externe Quellen aus der Literatur ergänzen und belegen eine Empfehlung, die im Bezug auf die aktuelle Methodenvorgabe, die konkreten Verbesserungspotentiale anhand der nutzbaren Konzepte und Elemente der Unified Modeling Language (UML) aufzeigt. Die Studienarbeit konkretisiert zudem die Methodik hinsichtlich der Modellierungstiefe bis zur Übergabe

zur Realisierung an einen Lieferanten und liefert hiermit einen wichtigen Beitrag zur Definition eines sinnvollen „Schnitts“ in Auslegung an das Billingsystem.

Zusammenfassend:

- ✓ geeignete fachliche Ebene als Basis für eine künftige Beauftragung des Lieferanten
- ✓ Handlungsempfehlung als Grundlage für eine Pilotierung im Umfeld des Abrechnungssystems
- ✓ konkretes Verbesserungspotential zur Prozessoptimierung

Die darüber hinaus erworbenen theoretischen und praktischen Inhalte im betrieblichen Kontext sowie die Vertiefung von Schlüsselkompetenzen machen das Projekt zu einem persönlichen Erfolg.

Für die anschließende Abschlussarbeit zum Erlangen des Titels Bachelor of Science bildet die praktische Arbeit einen bedeutenden Grundstein auf theoretischer Basis [Br05] und wirft etliche Fragenstellungen auf, die im Rahmen einer Bachelorarbeit thematisiert werden können. Der Zielsetzung aus betrieblicher sowie wissenschaftlicher Sicht wird die Arbeit also voll gerecht.

Literaturverzeichnis

- [Ba00] Balzert, Helmut: Lehrbuch der Software-Technik - Software-Entwicklung. Spektrum Akademischer Verlag GmbH, Heidelberg 2000.
- [Br05] Brugger, Ralph: IT-Projekte strukturiert realisieren. Vieweg & Sohn Verlag, Wiesbaden 2005.
- [FHR08] Fieber, Florian; Huhn, Michaela; Rumpe, Bernhard: Modellqualität als Indikator für Softwarequalität. Eine Taxonomie. Springer-Verlag, Heidelberg 2008.
- [JRH04] Jeckle, Mario; Rupp, Chris; Hahn, Jürgen; Zengler, Barbara; Queins, Stefan: UML 2 glasklar. Unified Modeling Language. Carl Hanser Verlag, München 2004.
- [Oe06] Österreich, Bernd: Analyse und Design mit UML 2.1: Objektorientierte Softwareentwicklung. Oldenbourg Wissenschaftsverlag, München 2006.
- [Ru07] Rupp, Chris; Queins, Stefan; Zengler, Barbara: UML 2 glasklar. Praxiswissen für die UML-Modellierung. Carl Hanser Verlag, München 2007.
- [Sc07] McLaughlin, Brett D; Pollice, Gary; West, David (dt. Übersetzung: Schulten, Lars): Objektorientierte Analyse und Design von Kopf bis Fuss. O'Reilly Verlag, Köln 2007.
- [Wi05] Winter, Mario: Methodische objektorientierte Softwareentwicklung - Eine Integration klassischer und moderner Entwicklungskonzepte. Dpunkt Verlag, Heidelberg 2005.

Geschäftsprozessmodellierung - automatisches Erzeugen von Benutzerschnittstellen mit der BPMN 2.0

Daniel Schlich

FH Aachen
Fachbereich Elektrotechnik und Informationstechnik
diplomarbeit@dschlich.de

Art der Arbeit: Diplomarbeit
Betreuer: Prof. Dr. rer. nat. Heinrich Faßbender

Abstract: Die Diplomarbeit beschäftigt sich im Rahmen des *Spider*-Projektes der FH Aachen mit dem Modellieren von Geschäftsprozessen mithilfe der BPMN-2.0. Es wird gezeigt, wie ein bestehender manueller Arbeitsablauf in einen EDV-basierten Geschäftsprozess umgewandelt und durch die BPM-Plattform *Activiti* ausgeführt werden kann. Dabei wird ein besonderes Augenmerk auf die Definition von Benutzerschnittstellen innerhalb von BPMN-2.0-Modellen sowie die Einbindung der BPM-Plattform in die bestehenden EDV-Systeme der FH Aachen gelegt.

1 Motivation

Als *Geschäftsprozess* bezeichnet man einen Prozess im organisatorischen Ablauf eines Unternehmens, der aus einer Folge von Tätigkeiten besteht, die von verschiedenen Aufgabenträgern ausgeführt werden. Auch an einer Universität gibt es eine Vielzahl von Geschäftsprozessen wie beispielsweise das Anmelden von Prüfungen, das Verwalten von erreichten Leistungen oder die Verteilung von Seminarplätzen. Zur Verwaltung dieser Geschäftsprozesse gibt es an der Fachhochschule (FH) Aachen eine große Anzahl unterschiedlicher Elektronischer Datenverarbeitungen (EDV)-Systeme, viele Prozesse werden komplett ohne oder nur mit geringer EDV-Unterstützung durchgeführt [Sch11].

Diese Diplomarbeit ist im Rahmen des *Spider*-Projekt von Herrn Prof. Dr. rer. nat. Faßbender entstanden. Das *Spider*-Projekt hat als Ziel die Optimierung der EDV-gestützten Verwaltung des Fachbereich (FB) 5 der FH Aachen durch Schaffung einer einheitlichen Kommunikationsplattform, auf der alle bestehenden Geschäftsprozesse elektronisch abgebildet, optimiert und soweit wie möglich automatisiert werden sollen [Bur11].

Die Kommunikationsplattform soll vorrangig folgende Merkmale erfüllen: Zum Erstellen und Ändern von Geschäftsprozessen wird kein technisches Wissen über die zu dessen Implementierung verwendeten Techniken benötigt, sondern nur noch Fachwissen über den Prozess selbst. Dies bedeutet, dass die jeweiligen Aufgabenträger ihre Prozesse selbst gestalten und anpassen können, ohne dazu spezielle Hilfe von Software-Entwicklern in

Anspruch nehmen zu müssen.

Außerdem soll das neue System einfach an die bestehenden EDV-Systeme der FH Aachen angebunden werden können. Diese Anbindung wird in der Regel über Web-Services erfolgen, allerdings soll auch die Verwendung andere Technologien möglich sein.

2 Modellierung von Geschäftsprozessen

Zur Modellierung von Geschäftsprozessen wurden in den vergangenen Jahren diverse Notationen entwickelt, wie z.B. *Ereignisgesteuerte Prozesskette*¹, *Petri-Netze*², *Unified Modeling Language*³ oder *Business Process Modeling Notation*⁴. Diese Notationen sind allerdings nur beschreibend, d.h. die so modellierten Geschäftsprozesse liegen nur in einer grafischen Darstellung vor und können nicht automatisch ausgewertet beziehungsweise *ausgeführt* werden.

Mit der Anfang des Jahres 2011 erschienenen Version 2.0 der Business Process Modeling Notation (BPMN) wurde erstmalig die Möglichkeit geschaffen, die in der BPMN beschriebenen Prozesse mittels einer Software (einer sogenannten *Process Engine*) auszuführen [Omg11]. Dies ermöglicht die Entwicklung eines EDV-basierten Geschäftsprozesses durch Zeichnen eines Prozessdiagramms in einem grafischen Editor (siehe Abbildung 1). Die einzelnen Arbeitsschritte können dabei beispielsweise durch Orchestrierung bereits verfügbarer oder neu zu entwickelnder Web-Services ausgeführt werden. Somit ist das Erstellen und Bearbeiten eines Geschäftsprozesses ohne Programmierkenntnisse möglich.

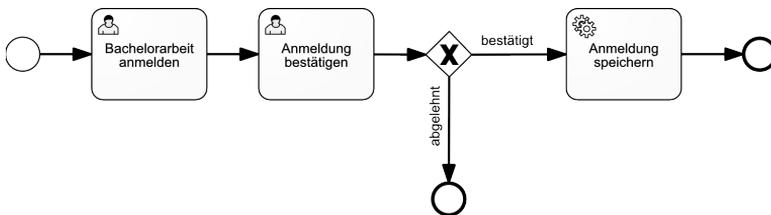


Abbildung 1: BPMN-Prozessdiagramm eines einfachen Beispielprozesses.

Eine in [Nol11] durchgeführte Bewertung und Gegenüberstellung verschiedener verfügbarer Business Process Model (BPM)-Lösungen empfiehlt den Einsatz der quelloffenen BPM-Plattform *Activiti*⁵ [PR10]. Diese enthält neben der *Process Engine* (*Activiti Engine*) einen grafischen Editor zum Bearbeiten von BPMN-Diagrammen (*Activiti Designer*) und weitere Module.

¹<http://www.epk-community.de/>

²<http://www.informatik.uni-hamburg.de/TGI/GI-Fachgruppe0.0.1/index.html>

³<http://www.omg.org/spec/UML/>

⁴<http://www.omg.org/spec/BPMN>

⁵<http://www.activiti.org>

3 Anbindung bestehender Systeme

Ein weiteres Hauptmerkmal der Kommunikationsplattform soll das problemlose Anbinden an die bestehenden EDV-Systeme der FH Aachen sein. Dafür bietet sich der Einsatz eines Enterprise Service Buss (ESBs) an: „Mit Enterprise Service Bus (ESB) bezeichnet man [...] eine Kategorie von Softwareprodukten, die die Integration verteilter Dienste (engl. service) in der Anwendungslandschaft eines Unternehmens (engl. enterprise) [...] unterstützen.“ [Wik11]

Da eine Entscheidung über den innerhalb des Spider-Projektes genutzten ESB noch nicht gefallen ist, wird hier exemplarisch der Einsatz des quelloffenen *Mule*-ESB⁶ beschrieben. Über diesen lassen sich mittels so genannter *Transport-Module* eine Vielzahl von EDV-Systemen integrieren, außerdem unterstützt er den Aufruf und das Bereitstellen von Web-Services und lässt sich einfach mit der BPM-Plattform *Activiti* verknüpfen.

4 Benutzerschnittstellen

Für die Interaktion der Aufgabenträger mit dem BPM-System ist eine Art von Benutzerschnittstelle notwendig, über die einerseits den Aufgabenträgern (den *Benutzern*) Informationen über den aktuellen Prozessstatus mitgeteilt werden und andererseits die Aufgabenträger weitere Informationen zu ihrer Aufgabe abrufen oder eingeben können.

Da die BPMN 2.0-Spezifikation keine Möglichkeit zur Definition von Benutzerschnittstellen beinhaltet, werden bestehende Methoden wie beispielsweise die *Activiti*-Erweiterungen der BPMN 2.0 zur Formulargestaltung vorgestellt. Zusätzlich werden neue Methoden entwickelt, unter anderem ein *Message-Processor* zum Generieren von Hypertext Markup Language (HTML)-Formularen im *Mule*-ESB und ein Plugin für die *Eclipse*-Entwicklungsumgebung⁷. Dieses kann HTML-Formulare sowohl aus den Definitionen von genutzten Web-Services als auch aus den *Activiti*-Erweiterungen der BPMN 2.0 zur Formulargestaltung erzeugen.

5 SpiderESB

Zum Überprüfen, ob die in der Motivation aufgestellten Ziele durch den Einsatz der BPMN 2.0 sowie eines ESBs erreicht werden können, wird eine Web-Anwendung mit dem Namen *SpiderEsb* erstellt, die als konzeptionelle Basis für die Entwicklung einer webbasierten Kommunikationsplattform dienen soll.

In dieser Web-Anwendung wird exemplarisch der Prozess *Anmelden der Bachelorarbeit im FB 5 an der FH Aachen* [FA10] implementiert. Dazu werden in einem ersten Schritt der bestehende Arbeitsablauf aufgenommen und in Zusammenarbeit mit den am Prozess

⁶<http://www.mulesoft.org/>

⁷<http://www.eclipse.org>

beteiligten Aufgabenträgern die durchgeführten Arbeitsschritte analysiert, optimiert und als BPM-Prozessdiagramm abgebildet.

Einer ausführlichen Anleitung zur Installation und Konfiguration der benötigten Softwarekomponente wie beispielsweise der BPM-Plattform *Activiti* und des *Mule*-ESBs folgt die Programmierung der Web-Anwendung *SpiderEsb*. Dabei wird der Prozess mithilfe des *Activiti Designers* modelliert und in *Activiti* bereitgestellt, die Entwicklung, Bereitstellung und Einbindung der benötigten Web-Services erläutert sowie die verschiedenen Methoden zur Definition von Benutzerschnittstellen angewendet.

Abschließend werden die *SpiderESB*-Anwendung und der erstellte Geschäftsprozess an den *Mule*-ESB angebunden.

6 Fazit

Diese Diplomarbeit zeigt, dass die Realisierung der Ziele des *Spider*-Projektes mithilfe der BPMN 2.0 sinnvoll ist. Durch den Einsatz einer BPM-Plattform können zukünftige Geschäftsprozesse schnell und einfach von den jeweiligen Prozessverantwortlichen definiert werden, ohne dass diese dazu Programmierkenntnisse besitzen oder die Hilfe von Softwareentwicklern in Anspruch nehmen müssten. Dies wird neue Wege zur Automatisierung der Verwaltung eröffnen, wodurch Ressourcen für die eigentliche Aufgabe der Fachhochschule - den Lehrbetrieb - freigesetzt werden.

Durch den Einsatz eines ESBs wird die Möglichkeit geschaffen, bestehende EDV-Systeme und Arbeitsabläufe an die neu zu erstellende BPM-Plattform anzubinden. So können neue EDV-basierte Geschäftsprozesse parallel zu bestehenden Arbeitsabläufen eingeführt werden und diese schrittweise ersetzen.

Literatur

- [Bur11] Vanessa Burkhardt. Geschäftsprozessoptimierung des Fachbereiches mittels BPMN 2.0 und ESB. Bachelorarbeit, FH Aachen, Januar 2011.
- [FA10] Rektor FH Aachen. Prüfungsordnung für den Bachelorstudiengang „Informatik“ und für den Bachelorstudiengang „Informatik mit Praxissemester“ im Fachbereich Elektrotechnik und Informationstechnik an der Fachhochschule Aachen, 6 2010.
- [Nol11] Otto Noll. Unterstützung des Geschäftsprozessmanagements durch IT Gesamtkonzepte. Diplomarbeit, FH Aachen, September 2011.
- [Omg11] Business Process Model and Notation (BPMN) Version 2.0, Januar 2011.
- [PR10] Niels Preuske und Bernd Rücker. *Activiti 5.0. Java Magazin*, 11:106–110, 2010.
- [Sch11] Sergei Schulz. Modellierung der IT-Systeme und Geschäftsprozesse im Fachbereich 5 mit BPMN2.0. Diplomarbeit, FH Aachen, März 2011.
- [Wik11] Wikipedia.de. Enterprise Service Bus. 17. März 2011.

Service Governance Patterns auf Basis des Open Group Referenzmodells

Benjamin Westphal

Hochschule Reutlingen
Fakultät Informatik

Benjamin.Westphal@student.reutlingen-university.de

Art der Arbeit: Masterthesis

Betreuer/in der Arbeit: Prof. Dr. rer. nat. Alfred Zimmermann

Abstract: SOA und Cloud Computing versprechen eine zeitnahe Adaption der IT-Möglichkeiten an die Bedürfnisse des Marktes [ST07]. Viele Organisationen scheitern jedoch oft hinsichtlich dieses Versprechens. Zurückzuführen ist dieses Problem meist auf eine fehlende oder unzureichende Lenkungsstruktur in Form einer Service Governance. Der hier vorgestellte Ansatz zeigt auf Basis des SOA Governance Frameworks der Open Group, in welcher Form eine solche Struktur aufgebaut werden kann. Eigens entwickelte Ansätze erweitern dieses Modell um beschreibende Patterns, die dazu dienen, Service Governance Lösungen zu generieren. Mithilfe dieser Modelle soll es jeder Organisation möglich sein, eine individuelle Service Governance aufzubauen, um die eigene Service Landschaft zu stärken.

1 Einleitung

Der Trend zu SOA ist kaum abgeklungen, da entstehen schon neue Ideen hinsichtlich des Cloud Computing und dessen Möglichkeiten. Laut BITKOM verspricht der dahinterstehende Markt eine Wachstumsrate von 26%. Eine Größe, die in keinem anderen IT-Sektor derzeit so hoch angesiedelt ist [BK09]. So wird mit Nachdruck an den verschiedensten Betriebsmodellen gearbeitet, um den hohen Erwartungen gerecht zu werden, denen Werbe- und Medienversprechen vorauseilten [WJ11]. So bleibt zu beobachten, wie nachhaltig Service-Landschaften entstehen, die unsere zukünftigen Geschäftsprozesse prägen. Dabei ist die Disziplin der Governance einer der wichtigen Knackpunkte, welcher beeinflusst, wie stabil und effizient solche Entwicklungen sind. Sie liefert den erforderlichen Rahmen, der durch Steuerungsprozesse begleitet wird, um beständig mit den Anforderungen und Systemen zu wachsen. Dies schließt nicht bisherige Governancemodelle, wie die der IT-Governance, aus, sondern erweitert es lediglich um etwaige Serviceaspekte und -kompetenzen, hier genannt Service Governance [WJ11]. Die in der Arbeit entstandenen Patterns beschreiben, wie einzelne Instanzen und Ausprägungen in Form von Prozessen, Rollen, Artefakten und Technologien zu generieren sind, deren Fundament das SOA Governance Framework der Open Group bildet.

2 Service Governance

Service Governance hilft der Organisation, nach individuellem Ermessen und Fokus ein eigenes SOA Governance Model zu definieren und anzuwenden. Dies nicht in einer Ad-hoc-Lösung, sondern mit ansteigender Auslieferung des Regelwerks, sodass die Organisation allen aktuellen Anforderungen gerecht wird, während sie ihre langfristigen Serviceziele anstrebt. [J08]. Es gibt kein einheitliches Modell einer guten Service Governance aufgrund der verschiedenen Gegebenheiten innerhalb einer Organisation. Aus diesem Grund formuliert jede Organisation ihr eigenes SOA-Governance-Modell und liefert Antworten auf folgende Fragen [SG09]:

- Welche Entscheidungen müssen getroffen werden, um eine effektive SOA Governance zu erhalten?
- Wer in der Organisation trifft diese Entscheidungen?
- Wie werden diese Entscheidungen umgesetzt und überwacht?
- Welche Strukturen, Prozesse und Tools sollten ausgeführt werden?
- Welche Metriken sind notwendig, um zu gewährleisten, dass SOA Implementationen im Sinne der strategischen Ziele sind?

3 SOA Governance Framework

Das SOA Governance Framework besteht aus dem SOA Governance Reference Model (SGRM) als Governance Fundament und der SOA Governance Vitality Method (SGVM), welche ein Definitions-, Verbesserungs- und Feedbackprozess ist, um den Fokus und das Customizing der SOA Governance zu definieren.

Das SGRM ist ein generisches Model, welches Informationen und den Rahmen bereitstellt, um ein SOA Regelwerk auf eine Organisation zuzuschneiden. Die gelieferten Best-Practice Beispiele dienen als Ausgangspunkt für Diskussionen, um übernommen oder ausgeweitet zu werden. Dabei lassen sich die Aspekte des SGRM wie folgt unterteilen: Guiding Principles, Governing Processes, Governed Processes, Roles & Responsibilities, Process Artifacts und Technology.

Die SOA Governance Vitality Method (SGVM) ist ein Prozess, welchem das SGRM zugrunde liegt. Dem Referenzmodell folgt eine Reihe von Aktivitäten, um das SGRM an die Organisation stetig anzupassen. Die Phasen des SGVM bilden eine kontinuierliche Verbesserungsschleife, bei welcher der Ablauf ständig gemessen wird und Kurskorrekturen und Aktualisierungen am SOA-Governance-Regelwerk und der SOA-Governance-Roadmap vorgenommen werden, sofern Bedarf besteht. Die Phasen des SGVM sind: Plan, Define, Implement und Monitor. [SG09]

4 Service Governance Patterns

Auf Grundlage der einzelnen Aspekte und Phasen des SOA Governance Frameworks entstanden im Rahmen dieser Arbeit zehn beschreibende Service Governance Patterns. Sie dienen als Hilfsmittel, um Governance-Lösungen zu generieren, die aus Organisationssicht und deren Stakeholder anzustreben sind. Daher liegt jedem Aspekt des Frameworks ein Pattern zugrunde, das beschreibt, wie sich das entsprechende Probleme darstellt und in welcher Form Lösungen zu generieren sind. Letztlich liefern sie eine Diskussionsgrundlage, da jede Organisation unterschiedliche Anforderungen und Prioritäten definiert. Best-Practice-Methoden aus dem Open-Group-Konsortium helfen bei der Orientierung und geben erste Anregungen für mögliche Lösungsbausteine, die entstehen können, da sie sich in der Praxis bewährt haben. In unserem angeführten Beispiel greifen wir ein Pattern heraus, das eine allgemeine Vorstellung über den Aufbau von Service Governance Patterns geben soll. Grundsätzlich gilt für alle Patterns, dass sie über den Namen der entsprechenden SOA Governance Framework Struktur zugeordnet werden können und immer ein individuelles Problem beschreiben, für das man Lösungen definieren sollte. Diese können Prozesse, Aktivitäten, Technologien, Artefakte oder auch Rollen sein. Wie eine solche Lösung beschrieben werden muss, wird durch das Pattern vorgegeben. Folgendes Beispiel zeigt anhand der Rollen und Verantwortlichkeiten, wie ein solches Pattern aussehen könnte. Für alle weiteren Arbeitsergebnisse verweisen wir auf die dazugehörige Masterarbeit.

Beispiel aus dem SOA Governance Reference Model / Roles and Responsibilities

Name	Roles and Responsibilities
Problem	How can I create roles and responsibilities to ensure the organization and service governance conventions?
Solution	Describes the team-, group- and domain structure of service governance
Role	Name of the role in the domain structure
Responsibilities	Description of the domain activities

Mögliches Lösungselement:

- Solution: SOA Steering Board
- Role: SOA Chief Architect; (...)
- Responsibilities: Define future SOA strategic direction and roadmap; (...)

Der Name beschreibt die Anordnung im SOA Governance Framework. Wie in jedem Pattern steht zentral ein Problem im Vordergrund, welches einen Anstoß auf entsprechende Lösungsergebnisse liefert. In diesem Beispiel ist die Lösung (grau dargestellt) durch eine Domain-Struktur, die Rolle in der Domain-Struktur und durch die Beschreibung der Aktivitäten dargestellt. Der Aufbau der Lösungsergebnisse ist in jedem Pattern anders zu formulieren, da sie vom Kontext des SOA Governance Frameworks abhängig sind.

5 Fazit und Ausblick

Der Open Group Standard des SOA Governance Frameworks liefert erste Strukturen und Best-Practice Ansätze, die im Rahmen dieser Arbeit durch eigene Service Governance Patterns erweitert wurden. Unter Berücksichtigung aller Aspekte und Phasen des SGRM und SGVM entstanden so zehn Lösungsschablonen, die dazu dienen, Service Governance Lösungen korrekt zu beschreiben, um sie effizient in laufende Service-Landschaften zu integrieren. Im Rahmen dieser Arbeit entstand so ein erster Entwurf, bei welchem alle Best-Practice Ansätze anhand der erstellten Patterns ausformuliert wurden. Die Ergebnisse konnten dann anhand eines entsprechenden IBM Modellierungstools in einem Prototyp zusammengeführt werden, der sich in Form eines Plug-Ins in Entwicklungsumgebungen integrieren lässt. So wäre der nächste Schritt das Implementieren der Service Governance in eine SOA- oder Cloud-Umgebung, welches im begrenzten Rahmen der Abschlussarbeit nicht mehr getestet werden konnte. Des Weiteren wäre zu beobachten, wie sich entsprechende Service-Modelle weiterentwickeln und in welcher Form die Service Governance hierzu neu strukturiert werden muss, um auch zukünftig einen stabilen Rahmen für neue Entwicklungen bereitzustellen.

Literatur

- [BK09] Weber, M.: Cloud Computing - Evolution in der Technik, Revolution im Business, BITKOM, 1. Auflage, Berlin, 2009
- [J08] Josuttis, N.: SOA in der Praxis - System-Design für verteilte Geschäftsprozesse, dpunkt Verlag, 1. Auflage, Heidelberg, 2008
- [ST07] Starke, G. & Tilkov, S.: SOA-Expertenwissen – Methoden, Konzepte und Praxis serviceorientierter Architekturen, dpunkt Verlag, 1. Auflage, Heidelberg, 2007
- [WJ11] Johannsen, W. & Goeken, M.: Referenzmodelle für IT-Governance, dpunkt Verlag, 1. Auflage, Heidelberg 2011
- [SG09] Open Group: SOA Governance Framework, 2009:
www.opengroup.org/projects/soa-governance, abgerufen am 04.09.2011

Gegenseitige Teilnehmerauthentifizierung in Telefonaten mit dem neuen Personalausweis

Steffen Brandt, Torsten Wiens und Michael Massoth

Hochschule Darmstadt
Fachbereich Informatik
sbrandt@fbihome.de, torsten.wiens@h-da.de
michael.massoth@h-da.de

Art der Arbeit: Masterarbeit
Betreuer/in der Arbeit: Prof. Dr. Michael Massoth

Abstract: Bei Telefonaten lassen sich Gesprächspartner an ihrer Stimme erkennen - vorausgesetzt, man kennt die Person, mit der man telefoniert. Ist dies jedoch nicht der Fall, muss man darauf vertrauen, dass sie die Wahrheit sagt oder durch technische Verfahren ein minimales Vertrauen schaffen. Diese Situation stellt tagtäglich in vielen Unternehmen ein Problem dar. Das Projekt Universal Call Authentication stellt eine Lösung für dieses Problem auf Basis des neuen Personalausweises und eines Smartphones bereit. Im Rahmen der Abschlussarbeit wurde eine Systemarchitektur für diese Lösung konzipiert.

1 Motivation

In der Vergangenheit konnte eine Zunahme von Diensten, die über das Internet abruf- und nutzbar sind, festgestellt werden. Ebenso hat die Verbreitung von Callcentern in den vergangenen drei Jahrzehnten zugenommen und dadurch den Service für den Kunden einfacher und zeitgemäßer gestaltet. Bei diesen Tele-Dienstleistungen, bei denen der Kunde nicht mehr physisch anwesend sein muss, ist es sehr wichtig geworden, sichere Authentifizierungsmechanismen einzusetzen, die die Identität des Kunden zweifelsfrei bestätigen können. Im Besonderen wäre es wünschenswert, kein zusätzliches Gerät mit sich führen zu müssen, um eine sichere Authentifikation zu ermöglichen.

Im Projekt *Universal Call Authentication* [UC12] wird eine Lösung dieses Problems entwickelt. Den universellen Vertrauensanker bildet hierbei der seit November 2010 verfügbare neue Personalausweis. Er verfügt über einen RFID-Chip für das kontaktlose Auslesen der im Ausweis gespeicherten Daten des Inhabers, welche als *electronic Identity* genutzt werden können. Diese erlaubt es jedem Bundesbürger, seine Identität online nachzuweisen. Aufbauend auf diese Funktion wurde eine Lösung entwickelt, die es ermöglicht, innerhalb von Telefongesprächen die Identität der Gesprächspartner sicher nachzuweisen.

2 Grundlagen – Neuer Personalausweis und VoIP

Die beiden für dieses Konzept wichtigsten Technologien sind die Internettelefonie (VoIP) [Tr07] und der neue Personalausweis (nPA) [PA10].

Die Telekommunikationslandschaft unterliegt heutzutage einem substanziellen Wandel. Wurden Telefonate in der Vergangenheit immer über ein geschlossenes und leitungsvermitteltertes Netz getätigt, wird sich dies in Zukunft ändern. Viele Betreiber stellen ihr Telekommunikationsnetz auf konvergente, paketvermittelte Netze um (sog. Next Generation Networks, NGN). Aufgrund dessen ist es möglich, beispielsweise einfache Mehrwertdienste anzubieten oder eine Verschlüsselung der Sprachdaten vorzunehmen.

Der neue elektronische Personalausweis wurde in der Bundesrepublik zum 1. November 2010 eingeführt [BSI10]. Die zentralen Funktionalitäten sind:

1. **Hoheitliches Dokument** als Ausweis gegenüber Behörden, mit Unterstützung biometrischer Daten.
2. **Elektronische Identität (eID)**. Diese Funktion ermöglicht es dem Inhaber des nPA, sich gegenüber Dritten im Internet zu authentifizieren. Dazu wird eine Sicherheitsinfrastruktur verwendet, die auf Inhaberseite die Verwendung der *AusweisApp* erfordert.
3. Weiterhin kann eine **Qualifizierte Elektronische Signatur** enthalten sein.

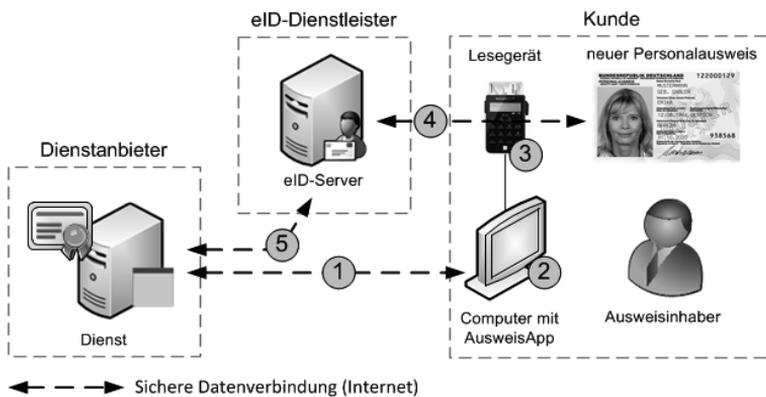


Abbildung 1: Authentifikation mit dem neuen Personalausweis

Im Folgenden soll die eID-Funktion erläutert werden (Abbildung 1). Zu Beginn des Ablaufs stellt der Ausweisinhaber eine Anfrage für eine Authentifikation an den Dienstanbieter, woraufhin dieser sein Berechtigungszertifikat über eine sichere Internetverbindung an den Kunden sendet (1). Das Zertifikat wird innerhalb der *Ausweis.App* auf dem Computer des Kunden angezeigt (2). Nach einer visuellen Prüfung und Bestätigung der Datenschutzbestimmungen und der ausgelesenen Felder des nPA leitet der Ausweisinhaber den Datenaustausch mit der Eingabe seiner PIN auf dem Tastenfeld des Lesegeräts ein (3). Nachdem eine Absicherung der Verbindungen zwischen den einzelnen Komponenten abgeschlossen ist, werden die Daten ausgetauscht (4). Sobald diese beim eID-Server vorliegen, können sie vom Dienstanbieter ausgelesen werden (5).

3 Systemüberblick

Bei *Universal Call Authentication* werden die klassischen Schutzziele der IT-Sicherheit durch verschiedene Technologien abgedeckt. Die *Vertraulichkeit* des Gesprächs wird durch die Verschlüsselung der Sprachdaten erreicht. Die ist allerdings nur bei Telefongesprächen über VoIP möglich. Bei diesen wird die Verschlüsselung über SRTP [RFC04] genutzt. Der Schutz der *Integrität* der Sprachdaten wird bei beiden Arten der Telefonie (VoIP und leitungsvermittelte Telefonie) vom System unterstützt. In herkömmlichen Netzen schützt der *Speaker Matching Algorithm* (SMART) die Integrität. In paketvermittelten Netzen wird diese durch die SRTP-Verschlüsselung garantiert.

Die *Authentizität* wird bei VoIP-Telefonie durch ein Challenge-Response-Verfahren gesichert. Dieses wird durchgeführt, bevor der Sprachkanal als sicher markiert wird. Die Lösung für herkömmliche Netze sichert dieses Schutzziel durch Anwendung des SMART. Bei einem Telefongespräch ist die Verbindlichkeit des gesprochenen Wortes von grosser Bedeutung. Dieses Schutzziel wird in diesem Projekt durch den Schutz der Integrität der Sprachdaten und durch die Anwendung des neuen Personalausweises erreicht.

Um vertrauliche Telefonie zu ermöglichen wurde in [PI09] ein Ansatz entwickelt, der den neuen Personalausweis dazu verwendet, die Gesprächsteilnehmer gegenseitig zu authentifizieren. Die Lösung entstand jedoch ausschließlich für die Voice-over-IP-Kommunikation. In [Wi10] wurde eine Erweiterung entwickelt, die den unsicheren Kanal, der leitungsvermittelten Telefonie absichert. Mit diesem Ansatz wird sichergestellt, dass keine dritte Instanz manipulativ in das Telefonat eingreifen kann.

Im Rahmen der Abschlussarbeit des Autors erfolgte die Konzeption eines Systemdesigns, das diese beiden Ansätze miteinander verbindet. Hierzu mussten die beiden Konzepte in einem ersten Schritt erfasst und in einem zweiten Schritt kompatibel zueinander gestaltet werden. Nachdem dies abgeschlossen war, konnte eine Gesamtarchitektur entworfen werden, die danach durch ein studentisches Entwicklerteam implementiert wurde.

Eine vereinfachte Darstellung der Architektur wird in Abbildung 2 enthalten.

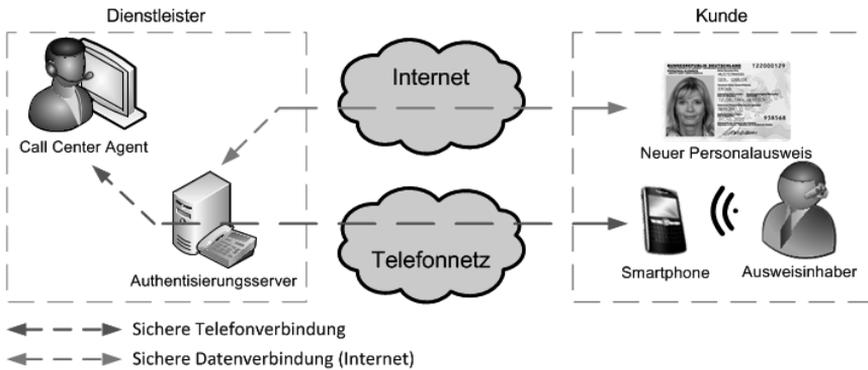


Abbildung 2: Vereinfachte Darstellung des Universal Call Authentication-Systems

Ausblick

Gegenwärtig existiert ein Prototyp des Universal Call Authentication-Systems, der auf der CeBIT 2012 vorgestellt wird. Der Prototyp wird im weiteren Verlauf des Projekts zu einem marktreifen Produkt entwickelt. Der Vorteil gegenüber Konkurrenzprodukten besteht darin, dass keine weiteren Komponenten außer jenen nötig sind, die ein Teilnehmer bei sich tragen kann (neuer Personalausweis, Smartphone).

Literaturverzeichnis

- [BSI10] Bundesamt für Sicherheit in der Informationstechnik: Der neue Personalausweis. http://www.personalausweisportal.de/cln_093/DE/Home/home_node.html, abgerufen 21.09.2011
- [PA10] Personalausweisportal. <http://www.personalausweisportal.de/>, abgerufen 13.01.2012
- [PI09] Plies, Andreas: Konzept zur Realisierung vertraulicher Kommunikation im eBusiness und eGovernment Umfeld unter Einsatz des neuen elektronischen Personalausweises. Hochschule Darmstadt, 2009
- [RFC04] Baugher et al.: The Secure Real-time Transport Protocol (SRTP). <http://tools.ietf.org/html/rfc3711>, erstellt März 2004, abgerufen 07.01.2012
- [Tr07] Trick, Ulrich; Weber, Frank: SIP, TCIP/IP und Telekommunikationsnetze. Oldenbourg Verlag, 3. Auflage, 2007
- [UC12] Universal Call Authentication Projektseite. <http://u-can.h-da.de>
- [Wi10] Wiens, Torsten: Anruferauthentifizierung in Telefondiensten auf Basis der eID-Funktion des elektronischen Personalausweises. Hochschule Darmstadt, 2010

Gesture-based User Authentication for Mobile Devices using Accelerometer and Gyroscope

Dennis Guse, Benjamin Müller

Quality and Usability Lab, Telekom Innovation Laboratories,
Technische Universität Berlin
{dennis.guse, benjamin.mueller}@campus.tu-berlin.de

Art der Arbeit: Master-Thesis in Informatik

Betreuer/in der Arbeit: Prof. Dr.-Ing. Sebastian Möller und Prof. Dr. Michael Rohs

Abstract: In this paper a user authentication mechanism for handheld mobile devices using hand gestures is presented. To authenticate a prior chosen hand gesture needs to be repeated. The mobile device uses a 3D-accelerometer and a 3D-gyroscope to measure the resulting device movement. With a user study it was shown, that the presented approach is feasible and has advantages with regard to usability over widespread knowledge-based authentication mechanisms like PINs and passwords on mobile devices. In a second study a realistic attack vector using video recordings was simulated. The results show that repeating an observed gesture similar enough is a very challenging task.

1 Introduction

Mobile devices offer lots of possibilities and a feeling of freedom to their users as they can be used almost everywhere and at any time. However, this freedom comes along with new security threats as mobile devices store a lot of sensitive information and provide access to important services [BE01]. In fact, unauthorized persons can gain access to mobile devices as for example they can get lost or stolen. To provide security, it needs to be verified that the current user is authorized to use the device. For mobile devices user authentication mechanisms based upon knowledge are widespread like PINs and passwords. Those are easy to implement and well understood by users. However, knowledge-based mechanisms require memorizing a non-guessable secret over a long period of time and everyone knowing this secret can successfully authenticate. On mobile devices those mechanisms are further restricted due to limited text input capabilities. In fact, implementing authentication mechanisms does not necessarily enhance security as genuine users may not use or circumvent unusable mechanisms [Re05]. To be successful an authentication mechanism needs to be fast, reliable and comfortable to use for the genuine users [AS99].

In this paper a user authentication mechanism based upon personalized hand gestures for handheld mobile devices is presented, which is introduced in the following section. Section 3 discusses the implementation and the approach to evaluate the mechanism. This paper closes with the results of the evaluation and gives an outlook on future work.

2 Gesture-based User Authentication

Gestures are expressive body movements and naturally used in human communication. User interfaces using gestures are promising as they allow creating intuitive and natural interfaces for human computer interaction. In prior work it was shown that user movements are an alternative for pairing multiple electronic handheld devices [PPA04, MG09]. For authentication on mobile devices interface designed hand gestures have been found useful [CM09]. However, such mechanisms should only be used in a secure environment as the gestures are potentially easy to observe and to mimic. Farella et al. presented a gesture-based identification mechanism for mobile devices using personalized gestures [Fe06]. A personalized authentication mechanism was also presented by Okumura et al. [Ok06] and refined by Matsuo et al. [Ma07].

The presented gesture-based mechanism is designed for mobile devices, which are mainly used one-handed. In the enrollment process the genuine user chooses freely a hand gesture, which he likes to use for authentication. A valid gesture is made holding the device in one hand. It needs to be repeatable. In the authentication process the current user is requested to repeat the prior chosen gesture and authentication succeeds, if the gesture could be reproduced. The usage of gestures for authentication on mobile devices has some advantages over knowledge-based mechanisms. The way how to move is memorized by humans implicitly during training in the motoric cortex, which is not prone to information overload and forgetting [KFR10]. In addition, a gesture cannot be written down or disclosed to other persons easily as it is a movement over time. Also, the available input space is vastly larger and authentication can be done eyes-free. However, humans cannot repeat a gesture completely similarly and the authentication mechanism must allow some variances including path, rotation as well as timing.

3 Implementation and Evaluation

The presented gesture-based authentication mechanism [Gu11] is implemented using a 3D-accelerometer and 3D-gyroscope. These sensors measure the motion indirectly, so the absolute path and orientation cannot be derived, but require little energy and are already built-in in modern mobile devices. To avoid computation and make the mechanism more robust a manual segmentation approach using a push-to-gesture button is used. Continuous First-order Hidden Markov Models (HMM) [Ra89] and Dynamic Time Warping (DTW) [SC78], two machine learning techniques, are used in the authentication process to decide, if the recorded measurements of a movement are similar to the chosen genuine gesture. In addition, the *Length Constraint* is introduced to limit the global variances in timing. To train the machine learning model and calculate the Length Constraint the genuine user provides samples of his genuine gesture.

The presented gesture-based authentication mechanism was evaluated in two user studies using an iPhone 4. In the first study feasibility and usability with 15 participants, which simulated genuine users, were explored. Each subject provided 25 samples of his individual interpretation of six pre-defined gestures, which were presented as a 2D image of the general path and a short description.

The goal of the second study was to prove resistance against skilled attacks [BLM07]. 10 participants acted as forgers using video recordings of the first user study. Three types of attack were studied: *Naïve Forgery*, *Semi-naïve Forgery*, and *Visual Forgery*¹. Naïve Forgeries are samples created in the first user study based upon another pre-defined gesture. Semi-naïve Forgeries are based upon the same pre-defined gesture, but not the same subject. Visual Forgeries were created by the forgers using the video recordings only. The quality of the forgeries should increase from Naïve to Visual Forgeries as more details about the used interpretation of the gesture are available. The collected data was evaluated offline using 5 samples for enrollment and the rest as forgeries.

4 Results and Future Work

The results of the user studies showed that gesture-based authentication is perceived as a usable alternative to PINs and passwords on mobile devices. The genuine participants found the mechanism not exhausting or annoying. On the one hand, two thirds of them would use such a mechanism in public places. On the other hand, the forger participants assumed that they can learn the gestures quite easily and reproduce them precisely. In contrast to this assumption, the collected data of both user studies shows that the implementation of the gesture-based authentication mechanism is feasible and robust against attacks. Figure 1 shows the Receiver Operating Characteristic (ROC), that is the tradeoff between the False-Acceptance-Rate and the False-Rejection-Rate, for Semi-naïve and Visual Forgeries depending on the allowed variance. The data shows that DTW performs better for Naïve and Semi-naïve Forgery, but HMM outperforms DTW for Visual Forgeries, i.e. realistic attacks.

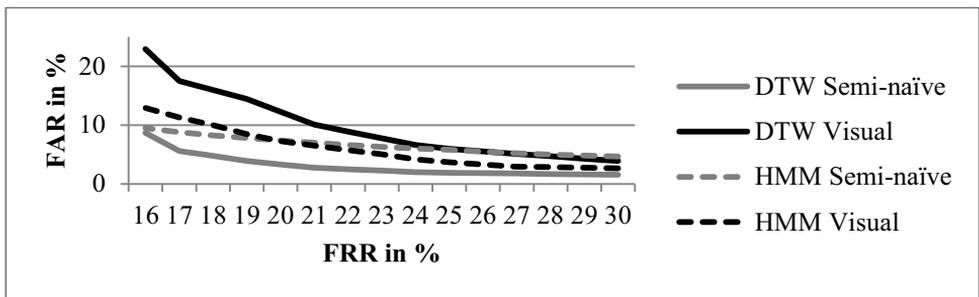


Figure 1: ROC diagram of DTW and HMM for Semi-naïve and Visual Forgery.

The results of the user studies show that the gestures-based authentication is feasibly as well as usable on mobile devices with built-in 3D-accelerometer and 3D-gyroscope. Overall, gesture-based authentication is a very promising alternative to established authentication mechanisms although the performance is lower than for standard PIN authentication. Future work is required to refine the applied algorithms and study the long term stability of memorized gestures. Also, social aspects have to be considered and evaluated. Gesture-based authentication may be a promising alternative to widespread knowledge-based authentication mechanisms.

¹ Based upon the classification of [BLM07].

References

- [AS99] Adams, A.; Sasse, M. A.: Users are not the Enemy. *Communications of the ACM*. December 1999, Vol. 12, No. 42, pp. 41-46.
- [BLM07] Ballard, L.; Lopresti, D.; Monrose, F. Forgery Quality and its Implications for Behavioral Biometric Security. *IEEE Transactions on Systems, Man, and Cybernetics*. October 2007, Vol. 37, No. 4, pp. 1107-1118.
- [BE01] Baumgarten, U.; Eckert, C.: Mobile, but Nevertheless secure? it+ti - Informationstechnik und Technische Informatik. 2001, Vol. 43, No. 5, pp. 254-263.
- [CM09] Chong, M. K.; Marsden, G.: Exploring the Use of Discrete Gestures for Authentication. Tom Gross et al. (Eds.). *Lecture Notes in Computer Science: Human-Computer Interaction – INTERACT '09*. Heidelberg, Germany : Springer, 2009, Vol. 5727, pp. 205-213.
- [Fe06] Farella, E., et al.: Gesture Signature for Ambient Intelligence: A Feasibility Study. Kenneth Fishkin, et al. (Eds.). *Lecture Notes in Computer Science: Pervasive Computing*. Heidelberg, Germany : Springer, 2006, Vol. 3968, pp. 288-304.
- [Gu11] Guse, D.: Gesture-based User Authentication on Mobile Devices using Accelerometer and Gyroscope, Master-Thesis, Technische Universität Berlin, May, 2011.
- [KHT06] Klemmer, S. R.; Hartmann, B.; Takayama, L.: How Bodies Matter: Five Themes for Interaction Design. *DIS '06 Proc. of the 6th Conf. on Designing Interactive Systems*. June 26-28, 2006.
- [Ma07] Matsuo, K., et al.: Arm Swing Identification Method with Template Update for Long Term Stability. Seong-Whan Lee and Stan Li. (Eds.). *Lecture Notes in Computer Science: Advances in Biometrics*. Heidelberg, Germany : Springer, 2007, Vol. 4642, pp. 211-221.
- [Ok06] Okumura, F. et al.: A Study on Biometric Authentication based on Arm Sweep Action with Acceleration. *ISPACS '06. Int. Symposium on Intelligent Signal Processing and Communications*. December 12-15, 2006, pp. 219-222.
- [MG09] Mayrhofer, R.; Gellersen, H.: Shake Well Before Use: Intuitive and Secure Pairing of Mobile Devices. *IEEE Transactions on Mobile Computing*. 8, 2009, Vol. 8, 6, pp. 792-806.
- [PPA04] Patel, S. N.; Pierce, J. S. and Abowd, G. D.: A Gesture-based Authentication Scheme for Untrusted Public Terminals. *UIST '04 Proc. of the 17th annual ACM Symposium on User Interface Software and Technology*. October 24-27, 2004.
- [Ra89] Rabiner, L. R.: A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *Proc. of the IEEE*. 1989, Vol. 77, 2, pp. 257-286.
- [Re05] Renaud, K.: Evaluating User Authentication Mechanisms. Lorrie, Faith Cranor and Simson Garfinkel (Eds.). *Security and Usability. Designing Secure Systems that People Can Use* Editors, O'Reilly, 2005, pp. 103-128.
- [SC78] Sakoe, H. and Chiba, S.: Dynamic Programming Algorithm Optimization for Spoken Word Recognition. *IEEE Transactions on Acoustics, Speech and Signal Processing*. 1978, Vol. 26, No. 1, pp. 43-49.

Entwicklung eines Authentifikationssystems mit Fingerfotos von Smartphone-Kameras

Chris Stein

Hochschule Darmstadt¹
- Fachbereich Informatik -
chris.stein@stud.h-da.de

Art der Arbeit: Masterarbeit

Betreuer: Prof. Dr. Christoph Busch, Prof. Dr. Michael Massoth, Claudia Nickel

Abstract: Diese Arbeit befasst sich mit der Authentifizierung von Personen an Smartphones mithilfe der Fingerabdruckerkennung. Hierbei werden Fingerbilder verwendet, die mit der eingebauten Smartphone-Kamera aufgenommen werden. Die vorgeschlagene Authentifizierungsmethode wird in Hinblick auf die Umsetzbarkeit analysiert und in einem Prototyp für das Android-Betriebssystem umgesetzt. Es werden Algorithmen für den Aufnahmeprozess entwickelt, die ein Mindestmaß an Qualität der aufgenommenen Bilder sicherstellen, um eine zuverlässige Fingerabdruckerkennung zu ermöglichen. Um die Erkennungsleistung zu optimieren, werden Verfahren zur Vorverarbeitung der Bilder analysiert und performante Lösungen für die Auswertung der Bilder entwickelt.

1 Motivation

Smartphones sind unter den Nutzern sehr beliebt. Der Smartphone-Markt wächst stetig und wird die klassischen Handys verdrängen. Durch die hohe Funktionsvielfalt der Smartphones werden auch mehr sensible persönliche Daten wie E-Mails, Fotos oder Account-Daten gespeichert. Diese Daten müssen vor Unbefugten geschützt werden. Klassische Authentifizierungsverfahren, wie Passwörter oder PINs, bieten hier nur begrenzten Schutz, weil diese vergessen, verloren oder gestohlen werden können. Insbesondere neigt der Mensch dazu, einfache kurze Passwörter zu benutzen, da diese leichter zu merken sind. Dieser Sachverhalt verschärft das Sicherheitsrisiko zusätzlich. Umfragen haben gezeigt, dass viele Nutzer die PIN-Eingabe sogar als störend empfinden und diese gar nicht nutzen [CF05]. Einen besseren Schutz gegenüber den herkömmlichen Authentifizierungsverfahren bietet die Authentifizierung über biometrische Merkmale [SK08]. In dieser Arbeit wird die Fingerabdruckerkennung mittels Fingerbilder durchgeführt. Jedes moderne Smartphone ist mit mindestens einer integrierten räumlich hochauflösenden Kamera ausgestattet, die in der Lage ist, feine Strukturen des Fingers zu erfassen, wie dies bei der Fingerabdruckerkennung erforderlich ist. Somit kann theoretisch auf jedem Smartphone eine Fingerabdruckerkennung ohne Zusatzgeräte durchgeführt werden. Ein solcher Ansatz

¹ Diese Arbeit wird in Zusammenarbeit mit dem CASED (www.cased.de) durchgeführt

wurde von Derawi et al. [DYB11] durchgeführt und lieferte vielversprechende Erkennungsraten. Die Ergebnisse wurden jedoch teilweise unter Laborbedingungen erzielt. Eine Untersuchung unter Alltagsbedingungen ist nicht bekannt und wird erstmals in dieser Arbeit durchgeführt. Um dem Benutzer mehrere Authentifizierungsmethoden anbieten zu können, wird die Lösung als Modul für MBASSy² [WN10] realisiert.

2 Herausforderungen und Probleme

Die Kamera des Smartphones muss in der Lage sein, auf möglichst nahe Objekte fokussieren zu können, um die Merkmale des Fingers erfassen zu können. Leider hat ein Test mit verschiedenen Modellen gezeigt, dass nicht alle getesteten Smartphones in der Lage sind, in den notwendig nahen Bereich fokussieren zu können. Solche Modelle sind somit nicht geeignet, um eine Fingerabdruckerkennung durchzuführen. Es ist aber davon auszugehen, dass in zukünftigen Modellen eine Verbesserung der Kamera mit dem Fortschritt der Technik stattfinden wird.

Weiterhin kommt erschwerend hinzu, dass aufgrund der kompakten Bauweise der Smartphones die Sensoren der Kameras klein ausfallen und die Bildqualität der Bilder trotz hoher Auflösung gering ist und durch Rauschen beeinträchtigt wird. Ein weiteres Problem sind die wenigen Konfigurationsmöglichkeiten der Kamera.

Außer dem Aufnahmegerät bietet auch der Aufnahmeprozess viele Herausforderungen, die es zu lösen gilt. Es existieren viele Parameter und Freiheitsgrade, die bei der Aufnahme sowie der Verarbeitung der Bilder beachtet werden müssen und die Erkennungsleistung beeinflussen: unterschiedliche Präsentationen des Fingers (Rotation, Skalierung, Translation), verschiedene Hintergründe, variable Beleuchtung, Struktur und Beschaffenheit des Fingers (Wölbung, Schmutz, Abrieb), Verwackelungen während der Aufnahme und Fokussierung. Zusätzlich muss auf eine ressourcenschonende Implementierung geachtet werden, da Rechenleistung und Arbeitsspeicher (insbesondere aufgrund der limitierenden maximalen Größe des Heap³ vom Betriebssystem) nur begrenzt auf dem Smartphone vorhanden sind.

3 Aufnahmeprozess

Während des Aufnahmeprozesses werden die Vorschaubilder der Kamera in einer grafischen Benutzeroberfläche (GUI) angezeigt. Die GUI liefert Informationen zum Aufnahmeprozess und ermöglicht dem Benutzer die Kamera auf die nahe möglichste Fokussierung zu stellen und den Aufnahmemodus zu starten. Sobald der Aufnahmemodus gestartet wurde, werden kontinuierlich die Vorschaubilder der Kamera mit dem Fingerbilderkennung- und Qualitätssicherungsalgorithmus ausgewertet (siehe Abschnitt 4.1 und 4.2). Sind die Kriterien erfüllt, wird automatisch eine Aufnahme gemacht. Die finale Aufnahme muss daraufhin nochmals auf ihre Qualität überprüft werden, da während der Auslöseverzögerung durch Bewegungen von Finger oder Kamera das Bild unscharf sein kann. Sollten alle Voraussetzungen bezüglich der

² MBASSy = Modular Biometric Authentication Service System

³ Menge an Arbeitsspeicher, die Android einer Applikation zuweist und nutzen darf

Qualität erfüllt sein, wird das Bild weiterverarbeitet und ein Template daraus erstellt. Je nach Modus wird eine Verifikation oder ein Enrollment mit dem erzeugten Template durchgeführt. Sollte das Bild nicht alle Qualitätsmerkmale erfüllen, wird das aufgenommene Bild verworfen und der Benutzer aufgefordert eine erneute Aufnahme zu machen. Die zu erfüllenden Qualitätsmerkmale werden in Abschnitt 4.2 näher erläutert.

4 Verarbeitungsprozesse

4.1 Fingerbildererkennung

Als erster Schritt muss der Finger im Vorschaubild der Kamera erkannt werden. Es wird sichergestellt, dass ein Finger möglichst zentral im Vorschaubild zu sehen ist und den Bildschirmrand nicht berührt. Beim Aufnahmeprozess wird die eingebaute LED des Smartphones aktiviert und leuchtet permanent. Dies schafft zum einen einheitlichere und robustere Beleuchtungsverhältnisse. Zum anderem erleichtert es die Unterscheidung zwischen Finger und Hintergrund, da durch die hohe Helligkeit die Kamera abblendet. Dadurch erscheint der nahe, von der LED beleuchtete Finger heller und der Hintergrund dunkler. Voraussetzung hierfür ist, dass der Finger einen gewissen Abstand zum Hintergrund hat. Aufgrund von Performance-Gründen und da die menschliche Haut in der Regel als Hauptfarbe den größten Farbanteil in Rot enthält, wird für alle nachfolgende Operationen nur der rote Farbkanal des Bildes ausgewertet. Für die Bestimmung der Region of Interest (ROI) wird ein farbbasierter Ansatz gewählt. Es wird ein Schwellwert für die Farbe Rot definiert und angenommen, dass Pixel, deren Rot-Anteil über den Schwellwert liegt, zum Finger gehören und Pixel, deren Rot-Anteil unter dem Schwellwert liegt, dem Hintergrund angehören. Ausgehend von den Rändern des Bildes werden die Pixel auf deren Farbwerte überprüft und bei Überschreitung des Schwellwerts die Grenzen der ROI gesetzt. Diese Vorgehensweise funktioniert zuverlässig unter der Annahme, dass sich keine Lichtquellen im Hintergrund des Bildes befinden. Um eine einheitliche Orientierung zu gewährleisten, die später für die Erstellung der Templates wichtig ist, wird zudem ein Korrekturwinkel berechnet, um den das Bild rotiert wird. Dies geschieht über die Bestimmung des Quadranten über die Lage der ROI und über trigonometrische Funktionen.

4.2 Qualitätssicherung

Nachdem die ROI erkannt wurde, muss festgestellt werden, wann eine für die Aufnahme geeignete Bildqualität vorliegt. Hierzu wird angenommen, dass eine gute Qualität vorliegt, wenn das Bild scharf ist. Um die Schärfe des Bildes festzustellen, wird der Sobel-Kantendetektor verwendet. Da bei der Aufnahme sichergestellt wurde, dass sich der Finger im Zentrum des Bildes befindet, ist es ausreichend, den Kantendetektor lediglich auf einen kleinen Teil im Zentrum des Bildes anzuwenden. Aus dem Ergebnis wird die Kantendichte (Summe der Kantenstärke der Pixel / Anzahl der Pixel) ermittelt. Sollte die berechnete Kantendichte den vorgegebenen Kantendichteschwellwert überschreiten und alle Pixel im Zentrum den Schwellwert für die Farbe Rot überschreiten, wird eine Aufnahme gemacht. Die finale Aufnahme unterzieht sich nochmals dieser Kontrolle. Nach Berechnung und Drehung der ROI wird zusätzlich die Plausibilität ihrer Dimensionen geprüft, bevor die Aufnahme akzeptiert wird.

4.3 Vorverarbeitung der Fingerbilder

Nach der Aufnahme des Bildes, bedarf es einer Vorverarbeitung bevor die Merkmale des Fingers aus dem Bild extrahiert werden können. Nach ausgiebigen Tests erwies sich die nachfolgend beschriebene Vorverarbeitungskette als effektiv und performant. Aufgrund der hohen Pixelanzahl des Bildes, wird das Bild auf halbe Breite und halbe Höhe skaliert, da unter Umständen sonst der Heap für die nachfolgenden Bildoperationen nicht ausreicht. Danach wird die ROI, wie in Abschnitt 4.1 beschrieben, berechnet, das Bild zugeschnitten und der zuvor berechnete Ausrichtungswinkel für die Rotation auf das Bild angewendet. Zudem wird das Bild segmentiert. Hierbei werden Pixel, deren Rot-Anteil unterhalb des definierten Schwellwerts ist, auf schwarz gesetzt. Um das Rauschen des Bildes zu reduzieren, wird ein Median-Blur-Filter angewendet. Anschließend wird das Bild mithilfe eines lokalen Binarisations-Filters binarisiert.

4.4 Minutien-Extraktor und Template-Comparator

Sobald das Bild vorverarbeitet wurde, können die Merkmale des Fingers extrahiert werden. Zum Einsatz kommt der Open Source Minutien-Extraktor „FingerJetFX OSE“ von DigitalPersona [Web1]. Dieser erstellt Templates nach dem ISO-Standard ISO/IEC FDIS 19794-2 [Web2]. Für die Verifikation wurde ein eigener Template-Comparator entwickelt, der zwei Templates nach dem genannten Standard miteinander vergleichen kann. Hierzu werden lokale Minutienvergleiche durchgeführt, wobei die Übereinstimmungs-Toleranz für Abstand und Winkel der Minutien sowie die Anzahl der zu übereinstimmenden Minutien zur positiven Authentifizierung konfiguriert werden kann.

5 Ergebnisse und Ausblick

Die hier beschriebene Lösung wurde mithilfe des Android SDK, Android NDK sowie der Grafikbibliothek OpenCV [Web3] umgesetzt und auf dem „Samsung Nexus S“ getestet. Ein erster Benutzertest mit wenigen Personen (277 Imposter- und 23 Genuine-Vergleiche) haben bei optimierten Parametern eine False Match Rate von 4,3% und eine False Non-Match Rate von 26,1% erzielt. Weitere Nachforschungen bzgl. der Auswirkungen der Parameter mit anschließenden Optimierungen an den Algorithmen und Benutzertests mit einer größeren Anzahl an Personen sind in Vorbereitung.

Literaturverzeichnis

- [CF05] Clarke N.L.; Furnell S.M.: Authentication of users on mobile telephones - A survey of attitudes and practices. In: Computers Security, S. 519-527, 2005.
- [DYB11] Derawi M. O.; Yang B.; Busch C.: Fingerprint Recognition with Embedded Cameras on Mobile Phones. In: MobiSec 2011, 3rd International ICST Conference, 2011.
- [SK08] Scheuermann D.; Kniess T.: Biometrie vs. Passwort - Advantage Biometrie. In: IT-Sicherheit (2008), Nr. 2, S. 46-48, 2008.
- [WN10] Witte H.; Nickel C.: Modular Biometric Authentication Service System (MBASSy). In: BIOSIG 2010, S. 115-120, 2010.
- [Web1] DigitalPersona: <http://www.digitalpersona.com/fingerjetfx> (Zuletzt besucht: 12.01.2012)
- [Web2] ISO: <http://www.iso.org> (Zuletzt besucht: 12.01.2012)
- [Web3] OpenCV: <http://opencv.willowgarage.com> (Zuletzt besucht: 12.01.2012)

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 Magenheimer (Hrsg.): Informatikunterricht und Medienbildung, INFOS'2001.
- P-9 Jan von Knop, Wilhelm Haverkamp (Hrsg.): Innovative Anwendungen in Kommunikationsnetzen, 15. DFN Arbeitstagung.
- P-10 Mirjam Minor, Steffen Staab (Hrsg.): 1st German Workshop on Experience Management: Sharing Experiences about the Sharing Experience.
- P-11 Michael Weber, Frank Kargl (Hrsg.): Mobile Ad-Hoc Netzwerke, WMAN 2002.
- P-12 Martin Glinz, Günther Müller-Luschnat (Hrsg.): Modellierung 2002.
- P-13 Jan von Knop, Peter Schirmbacher and Viljan Mahni_ (Hrsg.): The Changing Universities – The Role of Technology.
- P-14 Robert Tolksdorf, Rainer Eckstein (Hrsg.): XML-Technologien für das Semantic Web – XSW 2002.
- P-15 Hans-Bernd Bludau, Andreas Koop (Hrsg.): Mobile Computing in Medicine.
- P-16 J. Felix Hampe, Gerhard Schwabe (Hrsg.): Mobile and Collaborative Business 2002.
- P-17 Jan von Knop, Wilhelm Haverkamp (Hrsg.): Zukunft der Netze –Die Verletzbarkeit meistern, 16. DFN Arbeitstagung.
- P-18 Elmar J. Sinz, Markus Plaha (Hrsg.): Modellierung betrieblicher Informationssysteme – MobIS 2002.
- P-19 Sigrid Schubert, Bernd Reusch, Norbert Jesse (Hrsg.): Informatik bewegt – Informatik 2002 – 32. Jahrestagung der Gesellschaft für Informatik e.V. (GI) 30.Sept.-3. Okt. 2002 in Dortmund.
- P-20 Sigrid Schubert, Bernd Reusch, Norbert Jesse (Hrsg.): Informatik bewegt – Informatik 2002 – 32. Jahrestagung der Gesellschaft für Informatik e.V. (GI) 30.Sept.-3. Okt. 2002 in Dortmund (Ergänzungsband).
- P-21 Jörg Desel, Mathias Weske (Hrsg.): Promise 2002: Prozessorientierte Methoden und Werkzeuge für die Entwicklung von Informationssystemen.
- P-22 Sigrid Schubert, Johannes Magenheimer, 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 Rannenber, Wolfgang Wahlster (Hrsg.): Informatik 2003 – Innovative Informatikanwendungen (Band 1)
- P-35 Klaus Dittrich, Wolfgang König, Andreas Oberweis, Kai Rannenber, Wolfgang Wahlster (Hrsg.): Informatik 2003 – Innovative Informatikanwendungen (Band 2)
- P-36 Rüdiger Grimm, Hubert B. Keller, Kai Rannenber (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.): Persistence, 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, Wolfgang 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 Ad-hoc-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 Rannenber, Klaus Turowski (Hrsg.): Mobile Business – Processes, Platforms, Payments
- P-60 Steffen Friedrich (Hrsg.): Unterrichtskonzepte für informatische 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 its Applications

- P-64 Peter Liggesmeyer, Klaus Pohl, Michael Goedicke (Hrsg.): Software Engineering 2005
- P-65 Gottfried Vossen, Frank Leymann, Peter Lockemann, Wolfried 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 Kowalczyk, 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. Morgens-tern, 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 Röbling, 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 Management & 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 Magenheimer, 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 management 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 health 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 Special 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 Gimmich, 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 Kommunikationstechnologien 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 Brömmme, Christoph Busch, Detlef Hühnlein (Eds.)
BIOSIG 2008: Biometrics and Electronic Signatures
- P-138 Barbara Dinter, Robert Winter, Peter Chamoni, Norbert Gronau, Klaus Turowski (Hrsg.)
Synergien durch Integration und Informationslogistik
Proceedings zur DW2008
- P-139 Georg Herzwurm, Martin Mikusz (Hrsg.)
Industrialisierung des Software-Managements
Fachtagung des GI-Fachausschusses Management der Anwendungsentwicklung und -wartung im Fachbereich Wirtschaftsinformatik
- P-140 Oliver Göbel, Sandra Frings, Detlef Günther, Jens Nedon, Dirk Schadt (Eds.)
IMF 2008 - IT Incident Management & IT Forensics
- P-141 Peter Loos, Markus Nüttgens, Klaus Turowski, Dirk Werth (Hrsg.)
Modellierung betrieblicher Informationssysteme (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
I²CS 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 Mending, 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 Brömmе, 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 Rannenberг (Hrsg.)
Mobile und Ubiquitäre Informationssysteme
Technologien, Anwendungen und
Dienste zur Unterstützung von mobiler
Kollaboration
- P-164 Arslan Brömmе, 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 (I²CS) – 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 und E-Voting.CC
- P-168 Ira Diethelm, Christina Dörge, Claudia Hildebrandt, Carsten Schulte (Hrsg.)
Didaktik der Informatik
Möglichkeiten empirischer Forschungsmethoden und Perspektiven der Fachdidaktik
- P-169 Michael Kerres, Nadine Ojstersek, Ulrik Schroeder, Ulrich Hoppe (Hrsg.)
DeLFI 2010 - 8. Tagung der Fachgruppe E-Learning der Gesellschaft für Informatik e.V.
- P-170 Felix C. Freiling (Hrsg.)
Sicherheit 2010
Sicherheit, Schutz und Zuverlässigkeit
- P-171 Werner Esswein, Klaus Turowski, Martin Juhrisch (Hrsg.)
Modellierung betrieblicher Informationssysteme (MobIS 2010)
Modellgestütztes Management
- P-172 Stefan Klink, Agnes Koschmider, Marco Mevius, Andreas Oberweis (Hrsg.)
EMISA 2010
Einflussfaktoren auf die Entwicklung flexibler, integrierter Informationssysteme
Beiträge des Workshops der GI-Fachgruppe EMISA (Entwicklungsmethoden für Informationssysteme und deren Anwendung)
- P-173 Dietmar Schomburg, Andreas Grote (Eds.)
German Conference on Bioinformatics 2010
- P-174 Arslan Brömme, Torsten Eymann, Detlef Hühnlein, Heiko Roßnagel, Paul Schmücker (Hrsg.)
perspeGKtive 2010
Workshop „Innovative und sichere Informationstechnologie für das Gesundheitswesen von morgen“
- P-175 Klaus-Peter Fähnrich, Bogdan Franczyk (Hrsg.)
INFORMATIK 2010
Service Science – Neue Perspektiven für die Informatik
Band 1
- P-176 Klaus-Peter Fähnrich, Bogdan Franczyk (Hrsg.)
INFORMATIK 2010
Service Science – Neue Perspektiven für die Informatik
Band 2
- P-177 Witold Abramowicz, Rainer Alt, Klaus-Peter Fähnrich, Bogdan Franczyk, Leszek A. Maciaszek (Eds.)
INFORMATIK 2010
Business Process and Service Science – Proceedings of ISSS and BPSC
- P-178 Wolfram Pietsch, Benedikt Krams (Hrsg.)
Vom Projekt zum Produkt
Fachtagung des GI-Fachausschusses Management der Anwendungsentwicklung und -wartung im Fachbereich Wirtschaftsinformatik (WI-MAW), Aachen, 2010
- P-179 Stefan Gruner, Bernhard Rumpe (Eds.)
FM+AM'2010
Second International Workshop on Formal Methods and Agile Methods
- P-180 Theo Härder, Wolfgang Lehner, Bernhard Mitschang, Harald Schöning, Holger Schwarz (Hrsg.)
Datenbanksysteme für Business, Technologie und Web (BTW) 14. Fachtagung des GI-Fachbereichs „Datenbanken und Informationssysteme“ (DBIS)
- P-181 Michael Clasen, Otto Schätzel, Brigitte Theuvsen (Hrsg.)
Qualität und Effizienz durch informationsgestützte Landwirtschaft, Fokus: Moderne Weinwirtschaft
- P-182 Ronald Maier (Hrsg.)
6th Conference on Professional Knowledge Management
From Knowledge to Action
- P-183 Ralf Reussner, Matthias Grund, Andreas Oberweis, Walter Tichy (Hrsg.)
Software Engineering 2011
Fachtagung des GI-Fachbereichs Softwaretechnik
- P-184 Ralf Reussner, Alexander Pretschner, Stefan Jähnichen (Hrsg.)
Software Engineering 2011
Workshopband
(inkl. Doktorandensymposium)

- P-185 Hagen Höpfner, Günther Specht, Thomas Ritz, Christian Bunse (Hrsg.)
MMS 2011: Mobile und ubiquitäre Informationssysteme Proceedings zur 6. Konferenz Mobile und Ubiquitäre Informationssysteme (MMS 2011)
- P-186 Gerald Eichler, Axel Küpper, Volkmar Schau, Hacène Fouchal, Herwig Unger (Eds.)
11th International Conference on Innovative Internet Community Systems (I²CS)
- P-187 Paul Müller, Bernhard Neumair, Gabi Dreo Rodosek (Hrsg.)
4. DFN-Forum Kommunikationstechnologien, Beiträge der Fachtagung 20. Juni bis 21. Juni 2011 Bonn
- P-188 Holger Rohland, Andrea Kienle, Steffen Friedrich (Hrsg.)
DeLFI 2011 – Die 9. e-Learning Fachtagung Informatik der Gesellschaft für Informatik e.V. 5.–8. September 2011, Dresden
- P-189 Thomas, Marco (Hrsg.)
Informatik in Bildung und Beruf INFOS 2011
14. GI-Fachtagung Informatik und Schule
- P-190 Markus Nüttgens, Oliver Thomas, Barbara Weber (Eds.)
Enterprise Modelling and Information Systems Architectures (EMISA 2011)
- P-191 Arslan Brömme, Christoph Busch (Eds.)
BIOSIG 2011
International Conference of the Biometrics Special Interest Group
- P-192 Hans-Ulrich Heiß, Peter Pepper, Holger Schlingloff, Jörg Schneider (Hrsg.)
INFORMATIK 2011
Informatik schafft Communities
- P-193 Wolfgang Lehner, Gunther Piller (Hrsg.)
IMDM 2011
- P-194 M. Clasen, G. Fröhlich, H. Bernhardt, K. Hildebrand, B. Theuvsen (Hrsg.)
Informationstechnologie für eine nachhaltige Landbewirtschaftung
Fokus Forstwirtschaft
- P-195 Neeraj Suri, Michael Waidner (Hrsg.)
Sicherheit 2012
Sicherheit, Schutz und Zuverlässigkeit
Beiträge der 6. Jahrestagung des Fachbereichs Sicherheit der Gesellschaft für Informatik e.V. (GI)
- P-197 Jörn von Lucke, Christian P. Geiger, Siegfried Kaiser, Erich Schweighofer, Maria A. Wimmer (Hrsg.)
Auf dem Weg zu einer offenen, smarten und vernetzten Verwaltungskultur
Gemeinsame Fachtagung Verwaltungsinformatik (FTVI) und Fachtagung Rechtsinformatik (FTRI) 2012
- P-198 Stefan Jähnichen, Axel Küpper, Sahin Albayrak (Hrsg.)
Software Engineering 2012
Fachtagung des GI-Fachbereichs Softwaretechnik
- P-199 Stefan Jähnichen, Bernhard Rumpe, Holger Schlingloff (Hrsg.)
Software Engineering 2012
Workshopband
- P-200 Gero Mühl, Jan Richling, Andreas Herkersdorf (Hrsg.)
ARCS 2012 Workshops
- P-201 Elmar J. Sinz Andy Schürr (Hrsg.)
Modellierung 2012
- P-202 Andrea Back, Markus Bick, Martin Breunig, Key Pousttchi, Frédéric Thiesse (Hrsg.)
MMS 2012: Mobile und Ubiquitäre Informationssysteme

GI-Edition Lecture Notes in Informatics – Seminars

- S-1 Johannes Magenheimer, Sigrid Schubert
(Eds.):
Informatics and Student Assessment
Concepts of Empirical Research and
Standardisation of Measurement in the
Area of Didactics of Informatics
- S-2 Gesellschaft für Informatik (Hrsg.)
Informationstage 2005
Fachwissenschaftlicher Informatik-
Kongress
- S-3 Gesellschaft für Informatik (Hrsg.)
Informationstage 2006
Fachwissenschaftlicher Informatik-
Kongress
- S-4 Hans Hagen, Andreas Kerren, Peter
Dannenmann (Eds.)
Visualization of Large and Unstructured
Data Sets
First workshop of the DFG's International
Research Training Group "Visualization
of Large and Unstructured Data Sets –
Applications in Geospatial Planning,
Modeling and Engineering"
- S-5 Gesellschaft für Informatik (Hrsg.)
Informationstage 2007
Fachwissenschaftlicher Informatik-
Kongress
- S-6 Gesellschaft für Informatik (Hrsg.)
Informationstage 2008
Fachwissenschaftlicher Informatik-
Kongress
- S-7 Hans Hagen, Martin Hering-Bertram,
Christoph Garth (Eds.)
Visualization of Large and Unstructured
Data Sets
- S-8 Gesellschaft für Informatik (Hrsg.)
Informatiktage 2009
Fachwissenschaftlicher Informatik-
Kongress
- S-9 Gesellschaft für Informatik (Hrsg.)
Informatiktage 2010
Fachwissenschaftlicher Informatik-
Kongress
- S-10 Gesellschaft für Informatik (Hrsg.)
Informatiktage 2011
Fachwissenschaftlicher Informatik-
Kongress
- S-11 Gesellschaft für Informatik (Hrsg.)
Informatiktage 2012
Fachwissenschaftlicher Informatik-
Kongress

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

