

Gesellschaft für Informatik (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 the fields of

- Seminar
- Proceedings
- Dissertations
- Thematics

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

The volumes are published in German or English

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

ISSN 1614-3213  
ISBN 978-3-88579-440-0



Gesellschaft für Informatik (Hrsg.): Informatiktag 2008

# GI-Edition

## Lecture Notes in Informatics



Gesellschaft für Informatik (Hrsg.)

## Informatiktag 2008

## Fachwissenschaftlicher Informatik-Kongress

14. und 15. März 2008

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





Gesellschaft für Informatik e.V. (GI)

## **Informatiktage 2008**

**Fachwissenschaftlicher Informatik-Kongress**

**14. und 15. März 2008**

**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-6

ISBN 978-3-88579-440-0

ISSN 1614-3213

### **Volume Editor**

Gesellschaft für Informatik e.V.

Ahrstraße 45

53175 Bonn

E-Mail: [gs@gi-ev.de](mailto:gs@gi-ev.de)

Redaktion: Ludger Porada

E-Mail: [ludger.porada@gi-ev.de](mailto:ludger.porada@gi-ev.de)

### **Series Editorial Board**

Heinrich C. Mayr, Universität Klagenfurt, Austria (Chairman, [mayr@ifit.uni-klu.ac.at](mailto:mayr@ifit.uni-klu.ac.at))

Jörg Becker, Universität Münster, Germany

Hinrich Bonin, Leuphana-Universität Lüneburg, Germany

Dieter Fellner, Technische Universität Darmstadt, Germany

Ulrich Flegel, SAP Research, Germany

Johann-Christoph Freytag, Humboldt-Universität Berlin, Germany

Ulrich Furbach, Universität Koblenz, Germany

Michael Koch, TU München, Germany

Axel Lehmann, Universität der Bundeswehr München, Germany

Peter Liggesmeyer, TU Kaiserslautern und Fraunhofer IESE, Germany

Ernst W. Mayr, Technische Universität München, Germany

Heinrich Müller, Universität Dortmund, Germany

Sigrid Schubert, Universität Siegen, Germany

Martin Warnke, Leuphana-Universität Lüneburg Germany

### **Dissertations**

Dorothea Wagner, Universität Karlsruhe, Germany

### **Seminars**

Reinhard Wilhelm, Universität des Saarlandes, Germany

© Gesellschaft für Informatik, Bonn 2008

**printed by** Köllen Druck+Verlag GmbH, Bonn

## **Wissenschaftliche Tagungsleitung**

Prof. Dr. Gottfried Vossen, Universität Münster

## **Programmkomitee**

Prof. Dr. Herbert Göttler, Universität Mainz

Prof. Dr. Katharina Morik, Universität Dortmund

Prof. Dr. Otto Spaniol, RWTH Aachen

Prof. Dr. Alfred Zimmermann, Fachhochschule Reutlingen

## **Workshops - Themen und Partner**

Umgang mit Komplexität interdisziplinär gestalten, oder: Wie muss eine Software ihre Anwender (ver-)führen?

**Comma-Soft AG**

Design-Led Engineering – A Hands On Experience

**D-Labs GmbH**

Business Intelligence in Praxisprojekten

**Informationsfabrik (IN-FAB) GmbH**

OLIVANOVA - Die Programmiermaschine

**Integranova GmbH**

Web 2.0 - Technologien

**iteratec Gesellschaft für iterative Softwaretechnologien mbH**

Planen / Optimieren / Steuern

**IVU Traffic Technologies AG**

Performance Engineering

**PPI Financial Systems**

Serviceorientierte Architekturen (SOA)

**sd&m AG**

# **Übersicht über das Rahmenprogramm und Partner der Informatiktage 2008**

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 2007:** 14.03. 8:30 Uhr

**Absolventen-Workshop:** 14.03. 9:30 Uhr

Leitung: Prof. Dr. Alfred Zimmermann, Fachhochschule Reutlingen

**Hauptvortrag:** 15.03. 14:00 Uhr:

Prof. Dr. Thomas Hoeren, Universität Münster

Ein Pferd frisst keinen Gurkensalat - Überlegungen zum Verhältnis von Informatik und Recht

**Postersession:** 14.03. 15:15 Uhr und 15.03. 9:00 Uhr

## **Workshops in drei Sessions**

### **Moderation der Veranstaltung:**

Prof. Dr. Gottfried Vossen, Universität Münster

## **Partner und Sponsoren der Informatiktage:**

### **Partner:**

B-IT Bonn-Aachen Center

### **Sponsoren:**

Comma-Soft AG

D-Labs GmbH

Informationsfabrik (IN-FAB) GmbH

Integranova GmbH

iteratec Gesellschaft für iterative Softwaretechnologien mbH

IVU Traffic Technologies AG

Micorsoft Deutschland GmbH

PPI Financial Systems

sd&m AG

### **Medienpartner:**

COMPUTER ZEITUNG

## **Informatiktage 2008 – Vorwort zum Tagungsband**

"Software - the new driving force" titelte schon 1984 Newsweek und dieser Titel ist auch heute noch aktuell, ja, wahrscheinlich aktueller als je zuvor. Die Wertschöpfung der meisten technischen Produkte liegt im Einsatz von Software und die im internationalen Wettbewerb unverzichtbare Flexibilität und Anpassbarkeit von Produkten ist nur durch vermehrten Einsatz von Software nachhaltig zu erreichen.

Die frühzeitige Erkennung dieses Trends zur Informatisierung unserer Gesellschaft hat bereits in den 70er Jahren zur breitflächigen Etablierung von Informatik in unseren Ausbildungsgängen geführt und sie hat wesentlich dazu beigetragen, unser Land technologisch auf höchstem Niveau als "Exportweltmeister" zu konsolidieren. Aber auch andere Länder sind diesem Trend gefolgt und haben verstanden, dass Investitionen in Informatik langfristig lohnend sind und vor allem eine Investition in Köpfe erfordern. Daraus resultiert vermehrte Präsenz auf den internationalen Märkten und ein verschärfter Wettbewerb um Globalisierungsstandorte.

Aus dem Wettbewerb um Produkte ist längst ein Wettbewerb um Köpfe geworden und die Existenz optimal in Informatik ausgebildeter Menschen ist mehr denn je ein bestimmender Faktor bei der Standortentscheidung global agierender Firmen geworden.

Die GI hat sich und wird sich bei der Entwicklung der Ausbildung in Informatik positionieren und engagieren. Die Informatiktage sind dazu ein kleiner, aber dennoch wichtiger Beitrag, denn er ermöglicht den besten unserer Studierenden, sich fachlich auszutauschen und zu messen. Doch nicht allein die wissenschaftlich-fachlichen Voraussetzungen bestimmen den Stellenwert der Informatik, sondern auch die Befähigung, Erlerntes in der täglichen Unternehmenswirklichkeit umzusetzen. Deshalb stehen die Informatiktage traditionell unter dem Motto „Wissenschaft meets Praxis“. So können sich die beteiligten Unternehmen, bei denen ich mich für ihre aktive Mitarbeit in den Workshops ganz herzlich bedanke, auch ganz direkt von der Qualität der Hochschulausbildung in der Informatik überzeugen. Unsere Studierenden haben ihrerseits die Möglichkeit, Methoden und Prozesse aus erster Hand kennen zu lernen. Ich freue mich außerdem, dass das B-IT Center Bonn-Aachen Center for Information Technology nunmehr im dritten Jahr den perfekten Rahmen für die Veranstaltung bietet.



Prof. Dr. Stefan Jähnichen  
(GI-Präsident)



# Inhaltsverzeichnis

## Grundlagen der Informatik

<b>Klein D.</b> Solving Randomised Sabotage Games for Navigation in Networks	15
<b>Manz M.</b> Stochastisch balancierte persistente Suchstrukturen in der Praxis	19
<b>Neider D.</b> Learning Automata for Streaming XML Documents	23
<b>Mosenkis V.</b> Lösung nichtlinearer Gleichungssysteme mit dem Newton-Verfahren	27
<b>Redlich D., Drobek M.</b> Entwicklung einer Umgebung für die Parallelisierung von Algorithmen am Beispiel von evolutionären Algorithmen	31
<b>Schneider T.</b> Practical Secure Function Evaluation	37
<b>Sorg J.</b> Breitensuche in Prolog	41

## Künstliche Intelligenz

<b>Borth D., Ulges A., Schulze C., Breuel T.M.</b> Keyframe Extraction for Video Tagging & Summarization	45
<b>Koleski A.</b> Informationsverarbeitung in spikenden Neuronalen Netzen durch Synchronisationen	49
<b>Radkhah K., von Stryk O.</b> Absolute Measurement Systems Based on Industrial Robots	53

## Softwaretechnik

### **Bade D.**

Context-Dependent and Self-Responsible Migration of Software Agents in Heterogeneous Environments 57

### **Bartels P.**

Clustering Techniques on Transformation Systems 61

### **Bauer J.**

Entwicklung einer OSEK/VDX-kompatiblen Systemschnittstelle für Linux 65

### **Friedrichs S., Günther H., Mielentz O., Wegner M.**

Praktikum: Entwurf und Implementierung eines Softwareprojekts in einer funktionalen Sprache 69

### **Ellerweg R.**

Werkzeuggestützte Anforderungsanalyse für die benutzerzentrierte Entwicklung mobiler Unternehmenssoftware 73

### **Ladkau M.**

A Revival for Legacy Systems - Successful Reengineering of Software Systems 77

### **Masur D., Sevrin D.**

Beschreibung von agentenbasierten Anwendungen 81

### **Mlynarski M.**

Qualitätssicherung im Business Process Modeling durch automatische Ableitung visueller Kontrakte aus heterogenen UML-Modellen 87

### **Natelberg S.**

Constraint Based Transformation Theory 91

### **Nicolai J.**

Towards non-exploitable processes: Extending the least privilege access principle support in OpenSolarisTM 95

### **Pradel M.**

Ontology Composition using a Role Modeling Approach 99

### **Rau O.**

Define a Web 2.0 Portal Solution Based on a New Lightweight Java Environment used to Bridge Web SOA and Enterprise SOA 103

### **Roth S.**

Technologiebasierte Performanceunterschiede in serviceorientierten Architekturen 107

<b>Tetzlaff D.</b> Erweitertes Hyperblock-Scheduling für VLIW-Prozessoren	111
--	-----

## **Mensch-Computer-Interaktion**

<b>Cremers E.</b> Maschinelles Lernen der semantischen Auszeichnung medizinischer Textdokumente	115
--	-----

<b>Döring, T.</b> Gestaltung eines papierbasierten Tangible Interface zur Unterstützung kunsthistorischer Arbeitstechniken	119
---	-----

<b>Eger T, Eichel S., Kluck C., Liedtke P., Rackow B.</b> VRLab – Interaktion in virtueller Realität	123
---	-----

<b>Kurt S.</b> Interessenprofile in virtuellen Identitäten	129
---	-----

<b>Lilienthal S.</b> T-Prox: Vorstellung einer Lösung zur Proxy-gestützten Usabilityanalyse von webbasierten GUIs	133
--	-----

<b>Pahs M.</b> An Analysis Dashboard for Evaluating System Performance and User Behavior in Web Portals	137
--	-----

<b>Pietzsch R., Nauerz A.</b> Using Unstructured Data Analysis and Semantic Tagging for the Adaptation of Portal Environments	141
--	-----

<b>Wiesner M.</b> Shareclipse - Ein Eclipse-Plugin für kollaborative Softwareentwicklung auf Basis der P2P-Technologie JXTA	145
--	-----

## **Datenbanken und Informationssysteme**

<b>Bächle S.</b> Mehrbenutzersynchronisation für XQuery in nativen XML-Datenbankmanagement-Systemen	149
--	-----

<b>Bremes T., Scheuermann L.</b> DyOgeneS: Generierte Oberflächen in einem verteilten automatisierten System	153
---	-----

<b>Döbbelin S., Gül S</b> MeduSA: Umgang mit Metadaten in dynamischen Softwareumgebungen	157
---	-----



<b>Fischer T.</b> Laufzeitadaption in einem serviceorientierten Komponentenmodell	161
<b>Fitting T.</b> Redaktionsprozesse in einer Serviceorientierten Architektur	165
<b>Gerecke F.</b> Process based Metadata for Semantic Desktop Search	169
<b>Holder S.</b> Availability Guarantees for Data Items in Unstructured P2P Networks	173
<b>Marterer R., Moi M., Schauerte T.</b> KoMa: Koordination verketteter Webservice-Aufrufe	177
<b>Mletzko C., Schmidt B.</b> Automatische Extraktion von Metadaten aus verschiedenen Dateiformaten	181
<b>Nelkner T.</b> Inhalts- und Personennetze im E-Mail-Verkehr	185
<b>Ou Y.</b> Performance Analysis and Optimization of an XML Database System	189
<b>Schwab A., Podwyszynski M.</b> Platform independent Data-Synchronization of mobile .NET and J2ME applications using a Service Oriented Architecture	193
<b>Schmidt B., Reinhardt W., Mletzko C.</b> MetaXsA: Modulares Framework zur Metadatenextraktion durch semantische Dokumentanalyse auf unstrukturierten Textdokumenten	197
<b>Weiner A.M.</b> Anfrageauswertung in nativen XML-Datenbankverwaltungssystemen	201

## Technische Informatik

<b>Blumenthal S.</b> Optimierte Umsetzung eines Telekommunikation Protokolls aus dem Bereich Internet-Telefonie auf Mikrocontrollern	205
<b>Brulic S., Brulic Sa., Deppe P., Eschweiler P., Etemadi R., Herbold K., Konersmann C., Meiche R., Scholz C., Timm N., Uckermann A., Weber E., Zorn R.C.</b> CSI: PC2 - A High Performance Biometric System	209
<b>Dziengel N., Wittenburg G.</b> Verteilte Ereigniserkennung in Sensornetzen	213

<b>Erdmann J., Schäfer A., Kauhaus C., Fey D.</b> A Profitability Heuristic that Reduces the Parameter Dependence of Dynamic Load Balancing	217
<b>Renner J.</b> Introducing Measurable Path Selection Metrics to Anonymizing Overlay Networks	221
<b>Wenzky S.</b> Kooperation heterogener Internet-Anwendungen	225
<b>Westermann B.</b> Ein einfaches Anonymisierungsverfahren basierend auf offenen Standards	231
<b>Widyadharma S.</b> Integration von Routing und Service Discovery in drahtlosen Mesh Netzwerken	235
<b>Zimmerling M.</b> An Energy-Efficient Routing Protocol for LinearWireless Sensor Networks	239

## **Graphische Datenverarbeitung**

<b>Fabritius, G.</b> The Virtual Aachen Project	243
<b>Gänß S., Kischlat K., Kraushaar J.</b> Qualitätssicherung von 3D Stadtmodellen - Studenten entwickeln ein innovatives Programm	247
<b>Goetzmann J.-F.</b> Massively Parallel Contact Simulation on Graphics Hardware using NVIDIA CUDA	251
<b>Pietsch S.</b> Anwendergestützte Erkennung und Auswertung ausgewählter Holzstrukturen für die Dendrochronologie mit Methoden der digitalen Bildverarbeitung	255

## **Wirtschaftsinformatik**

**Arsenova E.**

Unterstützung der Prozessmodellierung im Notfallmanagement 259

**Beck H.**

Integration of content into enterprise applications – Analysis, conceptualization and prototypical implantation using Java EE and JCR 263

**Ciechanowicz D., Wolf K.**

MoVIS – Ein mobiles, RFID-gestütztes Besucherinformationssystem für Museen 267

**Enns P.**

Entwicklung eines hochverfügbaren Backup Storage Area Network 271

**Matzat M., Schmiedel D.**

Transformation von EPK nach BPEL 275

# Solving Randomised Sabotage Games for Navigation in Networks

Dominik Klein

Dominik.Klein@rwth-aachen.de

In the classical notion of fault tolerant computing and model-checking one usually considers that the states of a system may change over time; the underlying structure itself however is assumed to be static.

While this assumption yields simplifications in terms of algorithmic manageability, it is also quite a big restriction if one tries to cover real world scenarios.

To attack this problem, a certain game model is analyzed that is able to model a dynamic behaviour of the system itself. This game model is based on the model of (discrete) sabotage games introduced by Rohde [Roh05]; it is however modified in such a way that it allows a more realistic modeling.

In the case of (discrete) sabotage games two players play against each other, the first player navigating in a network to reach a certain destination, the second player, the saboteur, trying to prevent that by deleting edges in the network.

In the altered model considered in this thesis the second player, the saboteur, is omitted and replaced by random fault. So in this approach the first player does not play against a second one, rather he plays against nature. There are several scenarios that motivate this changed model:

1. One could for example think of the “Traveling Researcher Problem” as formulated by van Benthem [vB05] where a researcher tries to get from his current location to some conference overseas. Unlucky as he is, his flight-connections may fail or flights may be rescheduled spontaneously. One might ask if he can - even under those obstacles - still reach his destination - e.g. asking for reachability.

In this scenario the system itself (various cities in the world with travel connections) is not static and might change after each movement within the system.

2. Another motivating example in practice might be the analysis of a network where packets are routed from node  $a$  to node  $b$ . The network itself however might be subject to failure, and links between nodes might be canceled. One could think of a cell-phone network.

In both cases it is somewhat awkward to assume that there is an omniscient adversary that manipulates the world; one could go so far to say that this is a rather paranoid view of the world. The fault in this type of scenarios is much more likely happening randomly.

The thesis mainly deals with establishing complexity results for solving the reachability question (can Runner reach a final vertex with a certain probability  $p$ ). It is shown that deciding reachability usually becomes PSPACE-hard for infinitely many probabilities  $p \in (0, 1)$ .

**Definition 1.** A randomised sabotage game is played on a multigraph  $(V, \eta)$ , where  $V$  is a set of vertices. The vertices are connected by a set of edges, given by  $\eta : V \times V \rightarrow \mathbb{N}$ , the multiplicity function, which denotes the multiplicity of an edge. If  $\eta(u, v) = 0$  the vertices  $u, v \in V$  are not connected at all, if  $\eta(u, v) = 1$  the vertices are connected by one edge and so on. In the case of undirected graphs,  $\eta(u, v) = \eta(v, u)$  is required.

A randomised game is a tuple  $\mathcal{G} = (\mathcal{A}, v_{in})$  with  $\mathcal{A} = (V, \eta_{in})$ . A position of the game is a tuple  $(v, \eta)$ . In each turn of the game, the player - called Runner - chooses an (outgoing) edge  $a = (u, v)$  in vertex  $u$  of position  $(u, \eta)$  and moves to vertex  $v$ . Then a dice with  $|\{e | \eta(e) \neq 0\}|$  sides is thrown and the multiplicity  $\eta$  of the chosen edge is reduced by one (deleted in the case of  $\eta = 1$ ).

Clearly, since edges are only deleted and not added, the number of positions is finite and the outcome of a deletion depends only on the current position of a game (and the move of the player, of course). So the positions form a Markov decision process.

**Definition 2.** Given a game  $(\mathcal{A}, v_{in})$  with  $\mathcal{A} = (V, \eta_{in})$ , a non-empty set of vertices  $F \subseteq V$  and a probability  $0 < p \leq 1$ , the reachability problem for randomised sabotage games is the question if Runner is able to reach a vertex  $v_f \in F$  with probability  $p$ , if he starts in  $v_{in}$  and chooses an optimal strategy.

**Theorem 1.** The reachability problem for randomised sabotage games is decidable in PSPACE

Listing 1: PSPACE-MAX-RSG

```

input:  $\mathcal{G} = (\mathcal{A}, v_{in}), \mathcal{A} = (V, \eta_{in}), v_f, p_{in}$ 
output: TRUE/FALSE

global float p; // stores  $Pr^{max}(\Diamond v_f)$  f.a.  $v_f \in F$ 
root =  $(v_{in}, \eta_{in})$ 
state_p =  $\frac{1}{\eta_{in} > 0}$ 
DFS-TRAVERSE(root, F, state_p);
if (p==p_{in}) return TRUE;
return FALSE;

```

Listing 2: DFS-TRAVERSE

```

/* state_p stores the probability of
   the subtree of the current state */
input: state s, a set of vertices F, float state_p

```

```

output: - // operates on the global variable p

if ( $s \rightarrow v == v_f$  f. a  $v_f \in F$ ) // reached the desired vertex
{
  p += state_p;
}

/* are there outgoing edges available
   from the current state f. some  $v_i$ ? */
if (!  $\eta(\text{state} \rightarrow v, v_i) == 0$ )
{
  for each outgoing edge(action)
  {
    for each edge  $i$  in the game graph with  $\eta_i > 0$ 
    {
      state nextState = new(this.state);
      nextState  $\rightarrow v = v_i$  for edge  $(v, v_i) = i$ ;
       $\eta(i) --$ ;
      float nextState_p = this.state_p +  $\frac{1}{|\text{state} \rightarrow \eta_{>0}|}$ ;
      DFS-TRAVERSE(nextState_p, F, nextState_p);
    }
  }
}

```

One can use an adapted depth-first search approach to calculate the probability  $p$  of the maximum-reachability and compare against the input probability. The algorithm is depicted in listing 1 and 2.

The PSPACE-hardness (and thereby the PSPACE-completeness) can be established by a reduction from the well-known problem of quantified boolean formula.

**Theorem 2.** *The QBF formula  $\varphi$  has a satisfying assignment iff. Runner wins the game on constructed game arena  $\mathcal{G}_\varphi$  with probability  $p = \prod_{i=0}^{k-2} \frac{3k+n+1}{4k+n-i}$  for some  $k \in \mathbb{N}$  and  $n \geq \text{size}(\varphi)$ .*

The exact details of the proof are omitted here, though the general approach is inspired by the game ALGORITHM as introduced by [Sto74] and [Pap94]: For each existential and for each all-quantified variable of the QBF-formula a partial game-arena is constructed. These partial arenas are connected in series. In an existential part Runner chooses the assignment of the corresponding variable by moving in a certain direction. In the all-quantified parts the assignment is chosen randomly. Later on, this assignment is “verified” in such a way that if the chosen assignment of the variables was correct, Runner wins exactly with the probability  $p$  as defined above.

## References

- [Pap94] C. M. Papadimitriou. *Computational Complexity*. Addison Wesley, 1994.
- [Roh05] P. Rohde. *On Games and Logics over Dynamically Changing Structures*. PhD thesis, RWTH Aachen, 2005.
- [Sto74] Larry J. Stockmeyer. *The Complexity of Decision Problems in Automata Theory and Logic*. PhD thesis, MIT, 1974.
- [vB05] Johan van Benthem. *Essays in Honor of Jrg H. Siekmann on the Occasion of His 60th Birthday*. Springer-Verlag, Berlin, 2005.

# Stochastisch balancierte persistente Suchstrukturen in der Praxis

Michael Manz

manz@ira.uka.de

## Abstract:

Das sortierte Suchen in großen Datenmengen in logarithmischer Laufzeit wird über einen Index realisiert, der auf einem persistenten sekundären Speicher hinterlegt ist. Ein solcher Index wird im Allgemeinen als deterministisch balancierter Suchbaum (z.B. als B-Baum) organisiert. Die Balancierungsmechanismen beugen einer Entartung der Baumstruktur bei sequentiellen Einfüge- und Löschoperationen vor.

Die 1997 erfundenen Skip-Bäume [Mes97] erreichen ihre Ausgewogenheit durch Zufallsexperimente. Dieser Balancierungsalgorithmus ist einfach und besitzt eine vergleichbare theoretische Komplexität wie sein deterministisches Pendant. Durch die stochastische Balancierung ist ein paralleler Entwurf einfacher zu realisieren. Aufgrund des zufälligen Charakters ist die Knotenbreite des Skip-Baums jedoch unbeschränkt.

In dieser Arbeit wurde empirisch ermittelt, wie sich die Varianz der Knotenbreite im wirklichen Betrieb als persistente Datenstruktur auswirkt. Aus Indizes von Datenbanken im Echtbetrieb wurden realistische Zugriffsmuster entwickelt, die die Grundlage für die praxisnahe Untersuchung der unterschiedlichen Balancierungsstrategien bildeten.

Die empirischen Ergebnisse bestätigten die Annahmen, die sich aus der Theorie ableiten lassen. Beide Balancierungsstrategien erbrachten eine vergleichbare Leistung. Eines der Zugriffsmuster erzeugte bei der deterministischen Balancierungsstrategie einen hohen Balancierungsaufwand. Die stochastische Balancierung zeigte hingegen unabhängig vom Zugriffsmuster eine gleichmäßige Leistung. Aus den Ergebnissen wurden die entscheidenden Leistungsfaktoren bestimmt, die sich maßgeblich auf die Effizienz eines balancierten Suchbaums auswirken. Weiter wurde die Beziehung der Parameter der Balancierungsmechanismen bestimmt, die zwischen den Strategien bestehen muss, um bei beiden eine ähnliche Leistung zu erhalten.

## 1 Einleitung

Sortierte persistente Suchstrukturen verwalten eine große Zahl von Elementen auf einem nicht-flüchtigen Speicher. Die von R. Bayer entworfenen B-Bäume garantieren eine logarithmische Laufzeit für alle Operationen, die für die Organisation solcher Indizes nötig sind [BM02]. Der B-Baum ist ein deterministisch balancierter Suchbaum, dessen Mechanismen einer Entartung der Baumstruktur bei sequentiellen Einfüge- und Löschoperationen vorbeugen. Diese Mechanismen verursachen einen nicht unerheblichen zusätzlichen Aufwand. Da die Notwendigkeit zur Balancierung erst auf der Blattebene erkannt wird, lässt



sich ein effizienter paralleler Algorithmus nur durch aufwändige Synchronisierung verwirklichen [PPC04][Eil80].

Die im Jahre 1997 entdeckten Skip-Bäume [Mes97] haben an dieser Stelle bessere Eigenschaften. Die Ausgewogenheit des Skip-Baums wird durch Zufallsexperimente erreicht. Dadurch entfällt bei sequentiellen Operationen die aufwändige explizite Balancierung, der Aufwand für den Erhalt der Baumstruktur ist unabhängig von der Eingabemenge. Durch die stochastische Balancierung reduzieren sich bei einem parallelen Entwurf die Synchronisationsprobleme erheblich, da zu jedem Zeitpunkt der Wirkungskreis einer Operation bekannt ist.

Aufgrund der stochastischen Balancierung sind die Laufzeit und die Gestalt des Skip-Baums Zufallsvariablen. Wie sich die Varianz dieser Variablen im wirklichen Betrieb auswirkt, lässt sich aufgrund der Komplexität von Computersystemen nur durch umfangreiche Testreihen ermitteln. Deswegen wurden in dieser Arbeit ein stochastisch balancierter Suchbaum hinsichtlich seiner Leistung und Robustheit untersucht und mit einer deterministischen Balancierungsstrategie verglichen.

## 2 Zugriffsmuster

Für die Untersuchung der Leistungsfähigkeit wurden die Anforderungen ermittelt, die an Suchstrukturen in der Praxis gestellt werden. Aus Datenbankindizes, die in einer ERP-Software im Echtbetrieb eingesetzt werden, wurden zwei Zugriffsmuster ermittelt, die diese Anforderungen abbilden. Ein Zugriffsmuster besteht aus Such-, Einfüge- und Löschoperationen. Das zufällige Zugriffsmuster simuliert rein zufällige Zugriffe auf einen Index. Das fast sequentielle Zugriffsmuster entspricht Zugriffen auf kurzlebige Daten, die häufig neu erstellt und nach einer bestimmten Zeitspanne wieder gelöscht werden.

## 3 Ergebnisse

Zu den durchgeführten Tests zählten Messungen über die Zahl der Zugriffe auf den sekundären Speicher. Die Zugriffszahlen des deterministisch balancierten Suchbaums waren bei einem fast sequentiellen Zugriffsmuster deutlich durch den Balancierungsaufwand erhöht. Trotz geringer Vorteile des B-Baums bei rein zufälligen Zugriffen, war das Gesamtbild der stochastisch balancierten Suchstrukturen hier besser und unabhängig vom gewählten Zugriffsmuster.

Bei Laufzeittests wurde jedoch deutlich, dass für eine gute Laufzeit eines Suchbaums die Zahl der Datenträgerzugriffe nicht entscheidend ist. Durch den Zwischenspeicher des Betriebssystems entfällt selbst bei Zugriffsmustern mit geringer Lokalität die Großzahl der Zugriffe auf den sekundären Speicher. Erst bei sehr großen Bäumen und reduziertem Zwischenspeicher werden Zugriffe auf der Blattebene zu einem Leistungsfaktor. Die Höhe eines Suchbaums, die maßgeblich mitentscheidend für die Zahl der Datenträgerzugriffe

ist, verliert dadurch ihren Einfluss als Leistungsfaktor.

Für die hier verwendete Implementierung stellte sich ein anderer Faktor als leistungsentcheidend heraus. Das Kopieren und Schreiben von Indexelementen aus beziehungsweise in den Datenträgerzwischenspeicher. Dieser Bearbeitungsaufwand tritt bei schreibenden Zugriffen auf und ist proportional zur Elementanzahl, die bearbeitet werden muss.

Aufgrund dieses Leistungsfaktors wirkte sich ein Effekt deutlich auf die Leistung der Suchbäume aus. Bei einem zufälligen Zugriffsmuster werden größere Blattknoten öfter bearbeitet. Aufgrund der stark variierenden Knotenbreiten der stochastisch balancierten Bäume ist hier dieser Effekt besonders ausgeprägt und führte zu einer Verdoppelung des Bearbeitungsaufwands. Unter Berücksichtigung der Tatsache, dass die Baumhöhe nicht entscheidend für die Leistung ist, wurde ein besseres Verhältnis der Parameter der Balancierungsstrategien bestimmt.

Die Varianz der Knotenbreiten der stochastisch balancierten Suchstrukturen liegt im Widerspruch zur Charakteristik von persistenten Speichern, die in festen Blockgrößen organisiert sind. Wie bereits erwähnt, ist ein Ergebnis der Untersuchungen, dass leistungsbeeinflussende Zugriffe auf diese Speicher nur im Extremfall auftreten, etwa bei sehr großen Indizes und hoher Konkurrenz um den Zwischenspeicher des Betriebssystems. Der Leistungsabfall bei einem Überlauf des Zwischenspeichers ist jedoch bei allen untersuchten Suchstrukturen so gravierend, dass dies in der Praxis durch bessere Hardwareausstattung verhindert werden muss.

Berücksichtigt man die Ergebnisse zu den Leistungsfaktoren, erreichen alle getesteten Suchbäume beim Suchen, Einfügen und Löschen im Mittel dieselbe Leistung. Der Balancierungsaufwand und folglich die Laufzeit der deterministischen Strategie ist stark von der Ordnung der Eingabemenge abhängig. Dagegen zeichnet sich die stochastische Strategie durch gleichmäßige Laufzeiten bei jeder Art von Zugriffsmuster aus. Für Datenbanken, die unbeschränkt große Datensätze verwalten können, sind diese Suchbäume also durchaus eine Alternative für das Verwalten der sortierten Indizes.

Der Entwurf eines pragmatischen, massiv parallelen Algorithmus für den Skip-Baum steht noch aus und bleibt zukünftiger Forschung überlassen. Durch den reduzierten Synchronisationsaufwand könnte eine parallele, stochastische Strategie effizienter arbeiten als eine vergleichbare deterministische Balancierungsstrategie.

Eine effektivere Organisation der Baumknoten, die den Bearbeitungsaufwand bei Schreibzugriffen reduziert, würde bei allen untersuchten Suchbäumen die Leistung verbessern. Eine Möglichkeit wäre eine als Array gespeicherte Skip-Liste, in der neue Einträge ans örtliche Ende des Arrays abgelegt werden. Die Verzeigerung unter den Elementen der Skip-Liste wird durch Verweise auf die Array-Indizes ersetzt. In gewissen Abständen muss die Fragmentierung des Arrays, die durch das Löschen von Einträgen entsteht, beseitigt werden. Dennoch hätte die Datenstruktur einen amortisiert logarithmischen Aufwand bei allen Operationen.

## Literatur

- [BM02] Rudolf Bayer und E. McCreight. Organization and maintenance of large ordered indexes. Seiten 245–262, 2002.
- [Ell80] Carla Schlatter Ellis. Concurrent search and insertion in 2 3 trees. Seiten 63–86, 1980.
- [Mes97] Xavier Messeguer. Skip Trees, an Alternative Data Structure to Skip Lists in a Concurrent Approach. *Informatique Theorique et Applications*, 31(3):251–269, 1997.
- [PPC04] Heejin Park, Kunsoo Park und Yookun Cho. Deleting keys of B-trees in parallel. *J. Parallel Distrib. Comput.*, 64(9):1041–1050, 2004.

# Learning Automata for Streaming XML Documents

Daniel Neider

daniel.neider@rwth-aachen.de

Rheinisch-Westfälische Technische Hochschule Aachen

Lehrstuhl für Informatik 7

Logik und Theorie diskreter Systeme

## 1 Introduction

The *eXtensible Markup Language (XML)* and the *Document Type Definition (DTD)* have become popular standards in practice, eg. for data exchange in commerce and finance. However, besides their practical importance also some interesting theoretical issues arise.

Probably the most fundamental one is the check whether an XML document is valid with respect to a given DTD. In fact, this task is easy to achieve using pushdown automata, which use an amount of memory that is bounded by the “depth” of the XML document. However, validating an XML document against a DTD with a constant amount of memory is much more involved.

This topic was introduced by Segoufin and Vianu [SV02]. They use *deterministic finite automata (DFAs)* to formalize the constant memory constraint. However, since well-formedness, ie. the property that the input is a proper XML document, is not a regular property, they require that a finite automaton only distinguishes correctly valid and non-valid XML documents but may work arbitrary on inputs that are not well formed.

Until now there is no direct construction method known that takes a DTD as input and outputs a DFA capable of validating XML documents in the above sense. We address this issue and show how the framework for learning regular languages can be adapted and learning algorithms can be applied to solve this special task.

## 2 The learning framework

The most prominent framework for algorithmic learning of formal languages was introduced by Angluin [Ang87]. In this setting, a *learner*, who initially knows nothing about the target language, tries to learn the target language by means of querying a so-called *minimally adequate teacher*. This teacher is capable of answering two types of queries.

- On a *membership query* the teacher is provided with a word and has to check whether the word belongs to the target language or not.
- On an *equivalence query* the teacher is provided with a conjecture<sup>1</sup> and has to decide whether this conjecture is a correct description of the target language. In case that the conjecture is not a correct description, the teacher has to reply a counter-example.

This framework needs to be adapted slightly to fit into Segoufin and Vianu’s setting: Since the behavior of a DFA on non XML documents may be arbitrary, a teacher is allowed to answer “don’t care” to a membership query if the provided word is not an XML document.

To implement such a teacher for a given DTD, we use a deterministic (*visibly*) *pushdown automaton* (VPA)<sup>2</sup> that accepts exactly all valid XML documents and, moreover, is able to indicate non well-formed inputs. The construction follows the intuitive way exploiting each individual rule of the DTD and can be done efficiently. To answer a membership query, our teacher simulates the unique run of the VPA on the provided input. To answer an equivalence query, our teacher processes the configuration graph of the product of the VPA and the conjecture. By means of a reachability analysis<sup>3</sup>, the teacher searches for reachable configurations where either the VPA is in an accepting configuration and the conjecture is not or vice versa. A counter-example can then easily be computed if necessary.

### 3 The learning algorithm

Our algorithm for learning DFAs that validate XML documents is an adaptation of Angluin’s well know learning algorithm for regular languages [Ang87]. However, since in this new setting a DFA can have an arbitrary behavior on inputs that are no XML documents, there is no longer a unique canonical DFA to learn. Thus, a learner has to choose a reasonable behavior on “don’t cares” and can even alter it during the learning process. Because of this the learning is much more demanding than in Angluin’s case.

Our learning algorithm uses a so-called *observation table*  $O = (R, S, T)$  to organize the learned data. Thereby,  $R$  is a prefix closed set of *representatives*,  $S$  is a suffix-closed set of *samples* and a mapping  $T : (R \cup R \cdot \Sigma) \cdot S \rightarrow \{0, 1, ?\}$ , which describes the table entries ( $\Sigma$  is the fixed alphabet). For an entry  $w \in (R \cup R \cdot \Sigma) \cdot S$ , a membership query is asked and the result is stored in the table. A table entry  $T(w)$  is 0 if the teacher indicates that  $w$  does not belong to the target language, 1 if it belongs to the target language and ? if it is a “don’t care”.

The observation table is used to compute a conjecture DFA compatible with the data stored in the table: Each representative corresponds to a states of the DFA while the samples are used to distinguish different states. To produce a deterministic automaton, the learner has

<sup>1</sup>In the case of learning regular languages, this may be a DFA

<sup>2</sup>A visibly pushdown automaton is a pushdown automaton, whose stack operations are determined by the type of the current input symbol. Confer to eg. Bárány, Löding and Serre [BLS06] for a detailed definition

<sup>3</sup>Confer to [EHR00] for detailed information

to choose a reasonable behavior on ?-entries, which can be expressed as a *closed* and *consistent constraint* that the table has to fulfill. Since each ?-entry can be resolved to either 0 or 1, one can think of these entries as Boolean variables  $x_w$ . The closed and consistent constraint can be expressed as Boolean formulae  $\varphi_{cl}$  and  $\varphi_{co}$  over these variables. A model of this formulae yields a DFA that is compatible with the data stored in the table. This DFA is used as conjecture on an equivalence query. If this conjecture is not yet correct, a counter-example is returned and added to the table, which is thereby extended. Additionally, the learner has to handle situations where no model exists.

A model for both formulae can be computed by means of a SAT solver. Unfortunately, the *satisfiability problem* is well known to be *NP*-complete. However, Gold [Gol67] showed that finding a (minimal) DFA compatible with a given set of data is *NP*-complete. Therefore, computing a DFA by means of solving SAT instances is a reasonable approach.

Since the results of a SAT solver are determined by its implementation, it seems impossible to make justified assumptions about its results and so we were not yet able to formally prove the termination of our learning algorithm. However, our Java-based proof-of-concept implementation empirically shows that the algorithm works correctly on small DTDs. A comprehensive analysis of large-scale DTDs was not yet performed. This implementation also contains several improvements to the original algorithm as well as an alternative learning algorithm based on the *regular positive, negative inference algorithm* proposed by Graciá and Oncina [GO92].

## 4 Further approaches

On a step towards a formal proof of correctness, we consider a second linearization of trees. In contrast to XML documents, where each subtree is surrounded by two matching tags, the so-called *parenthesis word representation* uses only one type of parenthesis to surround a subtree. We show that in this case a DFA and a *visibly one-counter automaton with threshold 1 (1-VCA)*<sup>4</sup> can do the validation task equivalently and develop an efficient transformation from 1-VCA's recognizing a DTD into DFAs and vice versa. The advantage of using 1-VCA's is that these automata can check whether the input is well formed and, therefore, “don't cares” can be avoided in a learning process.

We present our polynomial runtime learning algorithm for *m*-VCA-acceptable languages and use a threshold reduction method developed by Bárány, Löding and Serre [BLS06] to reduce the resulting threshold to 1. Our learning algorithm exploits the fact that the *behavior graph*, ie. the graph induced by the Nerode congruence, of an *m*-VCA-acceptable language has a repeating structure and, thus, can completely be described by a finite initial fragment. Our learning algorithm learns an initial fragment with increasing size and once this fragment is big enough, the repeating structure can be discovered by a *parallel breadth-first search*. Finally, an *m*-VCA can be constructed from this initial fragment.

---

<sup>4</sup>A visibly one-counter automaton with threshold *m* is an automaton, which has access to a counter and can test this counter up to a threshold of *m*. Moreover, its stack operations are determined by the type of the current input symbol. Confer to eg. Bárány, Löding and Serre [BLS06] for a detailed definition

## References

- [Ang87] Dana Angluin. Learning Regular Sets from Queries and Counterexamples. *Inf. Comput.*, 75(2):87–106, 1987.
- [BLS06] Vince Bárány, Christof Löding, and Olivier Serre. Regularity Problems for Visibly Pushdown Languages. In *STACS*, pages 420–431, 2006.
- [EHRS00] Javier Esparza, David Hansel, Peter Rossmanith, and Stefan Schwoon. Efficient Algorithms for Model Checking Pushdown Systems. In *Computer Aided Verification*, pages 232–247, 2000.
- [GO92] P. Gracia and J. Oncina. Inferring regular languages in polynomial update time. In N. Pérez de la Blanca, A. Sanfeliu, and E. Vidal, editors, *Pattern Recognition and Image Analysis*, volume 1 of *Series in Machine Perception and Artificial Intelligence*, pages 49–61. World Scientific, Singapore, 1992.
- [Gol67] E. Mark Gold. Language identification in the limit. *Information and Control*, 10(5):447–474, 1967.
- [SV02] Luc Segoufin and Victor Vianu. Validating streaming XML documents. In *PODS '02: Proceedings of the twenty-first ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*, pages 53–64, New York, NY, USA, 2002. ACM Press.

# Lösung nichtlinearer Gleichungssysteme mit dem Newton-Verfahren

Viktor Mosenkis,  
LuFG Informatik 12, RWTH Aachen  
viktor.mossenkis@rwth-aachen.de

**Abstract:** Für das Lösen von nichtlinearen Gleichungssystemen stehen in der Numerik viele Verfahren zur Auswahl. Die meisten davon benötigen dazu auch die Ableitung der Funktion. Diese Diplomarbeit zeigt die Berechnung der Ableitung einer gegebenen Funktion mittels Automatischen Differenzierens und verdeutlicht dabei die Arbeitsweise dieses Verfahrens innerhalb des Newton-Algorithmus.

In der vorzustellenden Diplomarbeit sollte eine Software (Bibliothek) in Fortran 90 zur Lösung von nichtlinearen Gleichungssystemen, basierend auf dem Newton-Verfahren, entwickelt werden. Das heißt, es muss eine Nullstelle der Funktion  $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$  gefunden werden (zur Veranschaulichung der Vorgehensweise siehe Abb. 1). Die Bibliothek ist so aufgebaut, dass der Benutzer nur eine Prozedur bereitstellen muss, die für  $x \in \mathbb{R}^n$  den Wert der Funktion an dieser Stelle  $y = F(x)$  ausrechnet. Im Gegensatz zur Mehrzahl bekannter Bibliotheken, die die Vorgabe der Ableitung der Funktion durch den Anwender voraussetzen, wird diese Aufgabe von der Software übernommen, was einen der Schwerpunkte der Diplomarbeit bildet. Für die Berechnung der Ableitung stehen dem Anwender zwei Methoden zur Auswahl: Finite Differenzen und Vorwärtsmodus des Automatischen Differenzierens. Im Einzelnen galt es, die theoretischen Vorteile des Automatischen Differenzierens wie:

- Verbesserung der Stabilität (keine Berechnung des Differenzenquotientes)
- Bestimmung der exakten Ableitung
- Verkürzung der Laufzeit durch Compiler optimierten Code

zu untersuchen und zu verifizieren. Während der Tests fielen insbesondere die Auswirkungen der Tatsache, dass es sich im Fall des Automatischen Differenzierens um den exakten Wert und nicht Näherung der Ableitung handelt. Die Jacobi-Matrix verlor bei einigen Testläufen ihre symmetrische positive Definitheit, wenn sie mittels finiten Differenzen ausgerechnet wurde. Besonders häufig trat dieses Phänomen im matrix-freien Modus auf (siehe unten), mit Vorwärtsmodus des automatischen Differenzierens gab es dagegen keine Probleme dieser Art. In einigen Fällen ließe sich auch feststellen, dass die Anzahl der Newton-Iterationen geringer war, wenn die Ableitungen mittels automatischen Differenzierens ausgerechnet wurden.

Bei dem nächsten Schritt des Newton-Verfahrens, der Lösung des linearen Gleichungssystems, werden dem Anwender die folgenden Verfahren angeboten:



- direkte Verfahren
  - LU-Zerlegung bzw. Gauß-Elimination
  - Cholesky-Zerlegung
- iterative Verfahren
  - Verfahren der konjugierten Gradienten
  - GMRES
  - GMRES mit Restart.

Standardmäßig verwenden diese die komplette Jacobi-Matrix. Die oben erwähnten Ableitungsmethoden berechnen jedoch nur eine Richtungsableitung. Für die ganze Jacobi-Matrix benötigt man also Ableitungen in alle Richtungen der Standardbasis des  $\mathbb{R}^n$ . Unabhängig von der Art des zur Berechnung der Ableitung gewählten Verfahrens, muss es folglich  $n$ -mal ausgeführt werden. Dieses wirkt sich zum einen negativ auf die Laufzeit aus, zum anderen muss die Jacobi-Matrix im Speicher vorgehalten werden, was bei sehr großen Gleichungssystemen die Ressourcen der Hardware übersteigen kann. Die beiden Nachteile können durch die Verwendung der iterativen Verfahren im matrix-freien Modus umgangen werden, denn alle drei benötigen die Matrix selbst zu keinem Zeitpunkt, sondern nur ein Matrix-Vektor-Produkt, - eine Richtungsableitung. Die komplette Jacobi-Matrix wird nicht vor dem eigentlichen Verfahren ausgerechnet und als Parameter an die Prozedur übergeben, stattdessen wird innerhalb des Löser die gerade erforderliche Richtungsableitung berechnet. Da die Anzahl der Iterationen bei diesen drei Verfahren deutlich geringer als die Dimension des Gleichungssystems ist, verringert sich entsprechend die Anzahl der Funktionsaufrufe, und die Laufzeit sinkt. Zum Validieren jener realen Auswirkungen theoretisch günstiger Verlaufsmerkmale der matrix-freien Version, wurden innerhalb komparativer Tests Speicherbedarf und Laufzeit dokumentiert und ausgewertet. Deswegen erlaubt die Software die iterativen Verfahren sowohl im normalen als auch im matrix-freien Modus auszuführen.

Als eine weitere Option wurde inexaktes Newton-Verfahren implementiert, bei dem man die in der Regel höchst aufwendige exakte Lösung des Gleichungssystems vermeidet. Das lineare Gleichungssystem wird stattdessen mit vorgegebener Genauigkeit  $\eta_k \geq 0$  im  $k$ -ten Iterationsschritt gelöst. Die Lösung  $s_k \in \mathbb{R}^n$  der Newton-Gleichung

$$F'(x_k)s = -F(x_k) \tag{1}$$

muss also so genau sein, dass

$$\|F(x_k) + F'(x_k)s_k\|_2 \leq \eta_k \|F(x_k)\|_2$$

gilt. Die Größe  $\eta_k \geq 0$  misst somit den relativen Fehler in der Newton-Gleichung (1). Die Software legt den Parameter im  $k$ -ten Newton-Schritt durch

$$\eta_k := \min\{1/(k+1), \|F(x_k)\|_2\}$$

fest. Für diese Wahl des Parameters  $\eta_k$  konnte es mathematisch bewiesen werden, dass unter der Voraussetzung, die Funktion  $F$  sei zweimal stetig differenzierbar, die quadratische

Konvergenz des Newton-Verfahrens erhalten bleibt. Dieses wurde anhand verschiedener Testläufe überprüft. Außerdem wurde untersucht, inwiefern solches Vorgehen zur schnelleren Berechnung der Nullstelle der Funktion  $F$  führt.

Um das Globalisierungsproblem beim Newton-Verfahren zu umgehen, stehen dem Anwender zwei Verfahren zur Verfügung: die Armijo-Rule und die im folgenden skizzierte Methode (line search). Man versucht die Funktion

$$G : \begin{cases} \mathbb{R}^n \rightarrow \mathbb{R} \\ \lambda \mapsto \|F(x_{k+1})\|_2 = \|F(x_k + \lambda s_k)\|_2 \end{cases}$$

zu minimieren. Die Nullstelle der Funktion  $G'$  wird mit Hilfe des Newton-Verfahrens gefunden. Die zweite Ableitung wird durch das zweifache Anwenden des Vorwärtsmodus des Automatischen Differenzierens erhalten. Auch hier wurde die Übertragbarkeit der theoretischen Erkenntnisse in die Praxis durch Fallstudien überprüft.

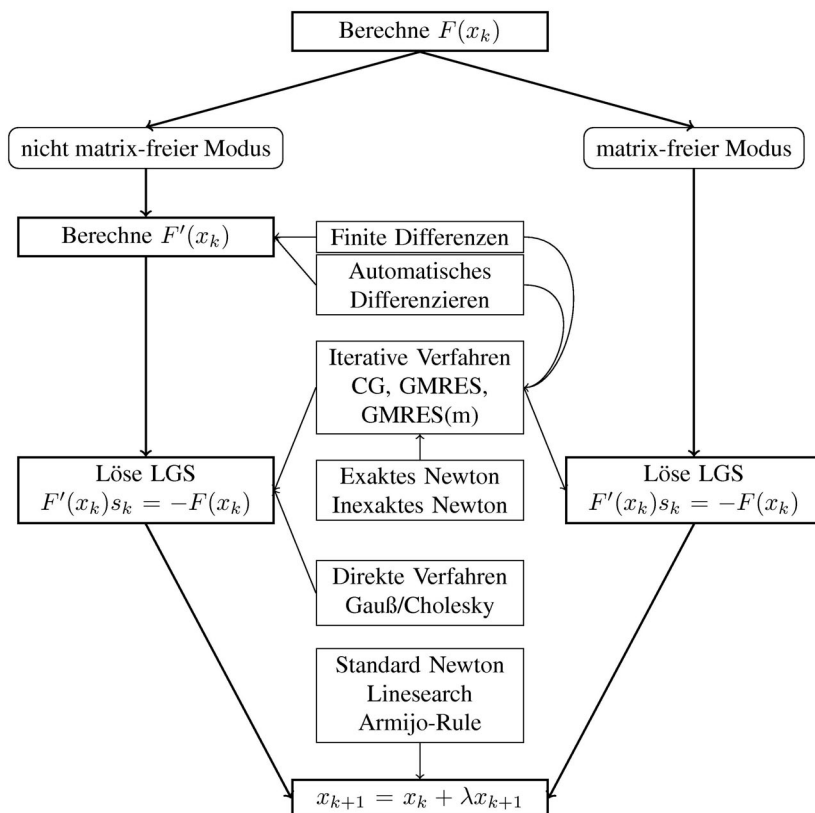


Abbildung 1: Aufbau der Diplomarbeit in Anlehnung an das Newton Verfahren

## Literatur

- [Gri00] A. Griewank. *Evaluating Derivatives. Principles and Techniques of Algorithmic Differentiation*. SIAM, Apr. 2000.
- [Kel03] C.T. Kelley. *Solving Nonlinear Equations with Newton's Method*. SIAM, Philadelphia, 2003.
- [MBG96] G. Corliss M. Berz, C. Bischof und A. Griewank. *Computational Differentiation: Techniques, Applications, and Tools*. Proceedings Series, SIAM, Philadelphia, 1996.
- [Mei05] Andreas Meister. *Numerik linearer Gleichungssysteme*. Vieweg, 2. Auflage, 2005.

# Entwicklung einer Umgebung für die Parallelisierung von Algorithmen am Beispiel von evolutionären Algorithmen

David Redlich, Marc Drobek

Hochschule für Technik, Wirtschaft und Kultur Leipzig  
Fachbereich für Informatik, Mathematik und Naturwissenschaften  
dredlich@imn.htwk-leipzig.de  
mdrobek@imn.htwk-leipzig.de

Art der Arbeit:           Abschlußarbeit zur Erlangung des Grades Bachelor of Science  
Betreuer der Arbeit:   Prof. Dr. J. Waldmann  
                              Prof. Dr. K. Weicker

**Abstract:** Durch die Arbeit wurde die Software UnionWork erstellt, mit deren Hilfe parallele Java-Algorithmen entworfen und ausgeführt werden können. Die Parallelisierung wird dabei, ähnlich wie bei einem Batchsystem, durch das Definieren und Zuweisen von einzelnen Aufträgen erreicht. Aufbauend auf der Sterntopologie gibt es viele Clientrechner, die als Rechenkraftwerke fungieren, einen Manager, der die Zuteilung der zu bearbeitenden Aufgaben verwaltet, sowie einen Operator, der den parallelen Algorithmus in das System einspeist und dessen Programmablauf überwacht. Im Speziellen wurde die Kombination zwischen der entstandenen Parallelisierungssoftware und einigen parallelen evolutionären Algorithmen untersucht.

## 1 Einleitung

In unserem Zeitalter der Informationstechnologie sind Probleme aus allen Bereichen der Naturwissenschaften um ein Vielfaches komplexer und aufwendiger, als es zu irgendeinem anderen Zeitpunkt der menschlichen Forschung der Fall war. Die Primzahlberechnung in der Mathematik, die Suche nach Impfstoffen in der Medizin, die Quantenmechanik in der Physik, die Klimasimulation in der Meteorologie und die Kryptografie in der Informatik sind Beispiele für heutige rechenintensive Forschungsgebiete. Komplexe Probleme dieser Art erfordern fortschrittliche Methoden zur Lösung. Ein solcher Ansatz ist die Parallelisierung. Im Gegensatz zu einer sequentiellen Abarbeitung werden mehrere Aufgaben zeitgleich von unterschiedlichen Instanzen verarbeitet. Das zieht in den meisten Fällen ei-

ne Senkung der Arbeitszeit nach sich. Je nach Größenordnung eines Problems verspricht eine Verteilung von Aufgaben mehr oder weniger Erfolg. Verbindet man diese Erkenntnis mit der Idee, das Internet und seine brachliegenden Leistungen für die Ausführung zu gebrauchen, entsteht ein vielversprechendes Konzept. Wie bei einer Ameisenkolonie steht ein Kollektiv von Arbeitern zur Verfügung, bei dem jeder einen kleinen Teil zur Lösung der Aufgabe beiträgt. Im Gegensatz zu jener Gemeinschaft beruht jedoch die Mitarbeit im Internet auf freiwilliger Basis. Um eine genügende Anzahl an Mitarbeitern für die Lösung seines Problems gewinnen zu können, sollte das Thema daher so interessant wie möglich gestaltet sein.

Eine derartige Eigenschaft weist der evolutionäre Algorithmus auf. Durch seine Nähe zur Biologie kann man sich dessen Arbeitsweise vorstellen. Sobald nicht nach optimalen - sondern lediglich nach hinreichend guten - Lösungen gesucht wird, kommt dieser Lösungsansatz in Frage. Obwohl bei der Verwendung mit enormen Zeiteinsparungen zu rechnen ist, existieren Probleme, die selbst dann noch jede zeitliche Beschränkung sprengen. Für solcherlei Aufgaben kann der evolutionäre Algorithmus mit einer Parallelisierung verbunden werden, um dem Problem Herr zu werden.

## 2 Parallelisierungsarchitektur

Das Grid-Computing, wie es heute praktiziert wird, bildet im Grunde einen Supercomputer nach dem Vorbild des Cluster-Computers nach, womit es als eine Untergruppe des Cluster-Computings gesehen werden kann. Allerdings besteht eine gravierende Einschränkung: Das Prozedere des Grid-Computing ist in heutigen Anwendungen meist starr und einfach. Das hat seinen Ursprung in der Nutzung des Internets und der damit verbundenen Voraussetzung einer hohen Robustheit. Die Kommunikation wird dabei auf ein Minimum gesenkt. Damit kann nur ein bestimmter Aufgabentypus die Vorteile des Grid-Computing nutzen, denn komplexere parallele Algorithmen sind auf einem solchen System keine umsetzbaren Anwendungen. Das ist uns zu wenig. Besonders im späteren Bezug auf evolutionäre Algorithmen reicht es nicht aus, immer nur vollkommen abgeschlossene Aufgaben auf einzelnen Rechnern ausführen zu lassen. Diesem Verlust algorithmischer Flexibilität will die Parallelisierungsbibliothek UnionWork entgegenwirken. Ziel ist, eine Annäherung an das ursprüngliche Cluster-Computing zu erreichen, in welchem eine höhere Parallelisierungsflexibilität erlaubt ist.

Wegen der eben erwähnten Einwände ist es nicht ganz einfach UnionWork korrekt in die Parallelisierungslandschaft einzuordnen. Grundsätzlich möchte sich UnionWork als Klassenbibliothek verstehen, die - ähnlich einem Batch-System - Aufgaben in einem Cluster-Computer verteilt. Dabei soll aber die Beschaffenheit und das Verbindungsnetzwerk eines Clusters keine Rolle spielen. Dies wird durch die Benutzung allgemeiner Standards bei der Kommunikation erreicht. Somit ist die UnionWork-Bibliothek sowohl für das Rechnen auf gewöhnlichen Clustern, als auch für das verteilte Rechnen über Internet (Grid-Computing) geeignet. Besonders *supercomputer at night* (SCAN) stellen ein sinnvolles Anwendungsgebiet dar. Eine gute Übersicht über die hier erwähnten Architekturen kann [BM06] entnommen werden.

Der grundsätzliche Aufbau des simulierten Clusters mit Manager, Operator und den Rechenknoten ist in Abb. 1 aufgezeigt. Dort ist deutlich die *Master-Slave*-Architektur herauszuerkennen.

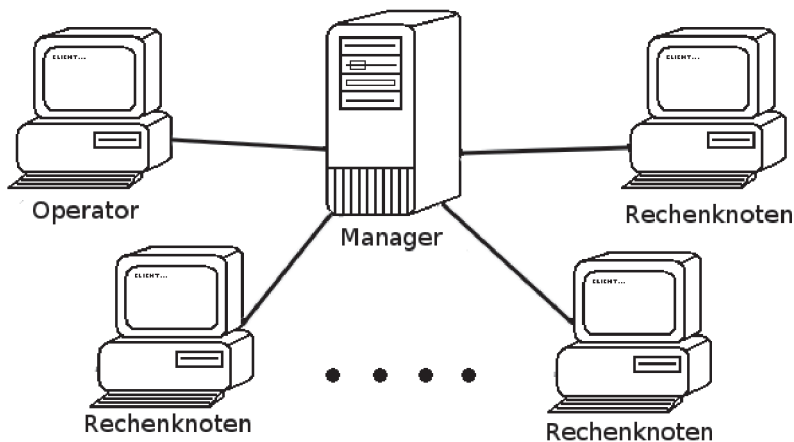


Abbildung 1: Simulierter Aufbau eines UnionWork-Clusters

### 3 Verwendung

Parallelisiert wird in UnionWork mit Hilfe von Aufträgen. Dazu muss je ein Ablaufplan für den Manager und einer für den Rechenknoten existieren. Ein Auftrag ist dabei ein entfernter Funktionsaufruf auf dem jeweils anderen Ablaufplan. Diese *Remote-Procedure-Call*-Technik wurde für eine sinnvolle Parallelisierung weiter abstrahiert. Durch die von der Bibliothek bereitgestellten Parallelisierungsfunktionen ist es möglich einem kompletten Rechner-Cluster eine Vielzahl von Aufträgen zuzuordnen. UnionWork übernimmt die Verteilung und Überwachung, sowie die Zuordnung eventueller Resultate zu den Aufträgen. Prinzipiell kann der Benutzer über viele Möglichkeiten Einfluss auf die Parallelisierung nehmen. Die wichtigsten werden kurz vorgestellt:

1. Es können beliebige Rechen-Cluster gebildet werden.
2. Der Benutzer hat die Möglichkeit sowohl bei Rechen-Clustern, als auch einzelnen Rechnern eine Auftragsmenge anzumelden.
3. Die Synchronisation des Algorithmus' wird durch den Aufruf synchroner Parallelisierungsfunktionen oder aber durch das Erstellen eines künstlichen Synchronisationspunktes hergestellt.
4. Der Algorithmus kann zentral durch den Manager-Ablaufplan und/oder dezentral durch den Rechenknoten-Ablaufplan gesteuert werden.

Besonders im Hinblick auf die Entwicklung hin zu Mehrkern-Prozessoren ist die Verwendung von Java als Programmiersprache sehr vorteilhaft, da mit ihr sehr einfach Threads erstellt und behandelt werden können. Dieser Fakt lässt eine überschaubare zusätzliche Parallelisierung innerhalb eines Rechenknotens zu. Desweiteren garantiert Java die Plattformunabhängigkeit, was eine Installation der Front-End-Software auf den meisten Systemen zulässt und damit das Bilden eines Grids bzw. Clusters aus verschiedensten Rechenknoten erlaubt.

## 4 Parallelisierung von evolutionären Algorithmen

In der heutigen Informatik treten vermehrt Probleme mit riesigen Suchräumen auf. Beispiele dafür sind graphentheoretische Aufgaben, wie etwa das Finden von Zyklen in Graphen oder die Suche nach dem kürzesten Weg. Das sogenannte „Handlungsreisendenproblem“ hat schon bei 101 Knotenpunkten einen Suchraum von über  $4,66 \cdot 10^{157}$  möglichen Kombinationen. Zur Handhabung dieser Lösungsmenge ist der evolutionäre Algorithmus ein effektives Mittel. Es werden nicht alle Möglichkeiten im Suchraum überprüft, sondern nur ein geringer Teil. Dadurch erhält man folglich eine kürzere Laufzeit im Vergleich zu *Brute-Force* Algorithmen, die den kompletten Suchraum durchlaufen. Bei der Suche nach einem Ergebnis mit dem evolutionären Algorithmus ist nie garantiert, dass die bestmögliche Lösung gefunden wird. Oftmals wird diese Bedingung in praktischer Hinsicht auch nicht vorausgesetzt, da man sich mit einem hinreichend guten Resultat zufrieden gibt. Der grundlegende Algorithmus arbeitet auf einer sogenannten *Population*, die eine Menge von Lösungen darstellt. Diese Menge wird vom Algorithmus zyklisch durch geeignete Operatoren verändert, so dass die eine Annäherung an bessere Lösungen erfolgt. Weiterführende Literatur zu den evolutionären Algorithmen bieten [Wei02] und [Bru96].

Der Hauptgrund für eine Parallelisierung von evolutionären Algorithmen liegt in der Verringerung der Wartezeit für den Anwender und der Möglichkeit, mehrere Evolutionsprozesse betrachten zu können. Natürlich bleibt die Gesamtrechenzeit des Algorithmus für einen sequentiellen Ablauf die gleiche, da jedoch Aufgaben gleichzeitig ausgeführt werden können, existiert ein Faktor  $x$ , um den sich die letztliche Wartezeit bis zum Erreichen einer hinreichend guten Lösung für den Anwender verkürzt. Dabei gibt es zwei wesentliche Parallelisierungsarten (siehe Abb. 2), die in der Praxis Verwendung finden und mit UnionWork umgesetzt wurden:

- Bei dem *Master-Slave* Prinzip existiert ein Rechner, ein sogenannter *Master*, auf dem zu Beginn des Evolutionsprozesses eine Population erstellt wird. Auf dem Master laufen auch alle evolutionären Operationen ab. Nach jeder Generation sendet der Master einen Teil der Population zu den *Slaves*, die ausschließlich zur Verfügung stehen, um die erhaltenen Individuen zu bewerten. Dieses Prinzip kommt zur Anwendung, wenn dem gewählten Problem eine komplexe Bewertungsfunktion für die Individuen zugrunde liegt.
- Die *grobkörnige Parallelisierung* unterscheidet sich grundlegend vom Master-Slave-Prinzip, da alle Slaves eine eigene Population generieren und darauf auch selbstständig

den Evolutionsprozess durchführen. Untereinander treten die Slaves durch das Senden von Individuen an andere Slaves in Verbindung. Dieser Prozess bedeutet immer auch eine Erweiterung des Genpools einer Population, da unter Umständen Individuen aus anderen Teilen des Suchraums in die Population einwandern. Eine weitere Eigenschaft, die die Güte und Rechendauer des evolutionären Algorithmus' beeinflusst, ist die Netztopologie. Sie gibt Auskunft darüber, in welche Populationen Individuen migrieren. Übliche Netztopologien sind dabei etwa vollständige Graphen oder Hypercubes. Im Regelfall liefert dieses Verfahren hinreichend gute Ergebnisse in akzeptabler Zeit. Aufgrund der aufgezählten Fakten kommt die grobkörnige Parallelisierungsstrategie evolutionärer Algorithmen der natürlichen Evolution am nächsten.

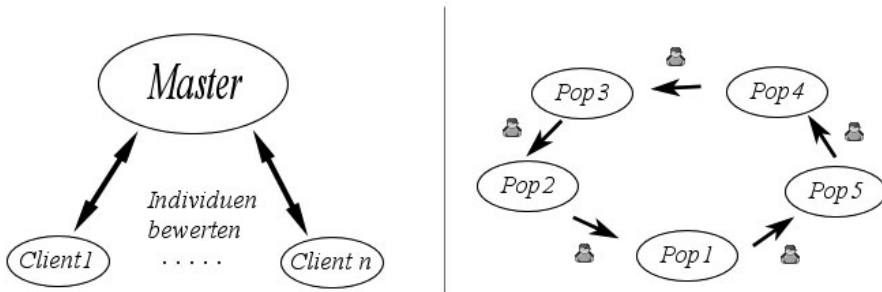


Abbildung 2: Master-Slave-Prinzip (links) und die grobkörnige Parallelisierung (rechts)

## 5 Fazit

Grundsätzlich ist für jedes Problem mit genügend Parallelisierungspotential eine Implementierung sinnvoll. Zukünftig kann sowohl eine Vertiefung der evolutionären Algorithmen, als auch eine Erweiterung des Frameworks um neue Themengebiete stattfinden. Das entstandene Framework ist für wissenschaftliche Forschungs- und Lehrzwecke frei.

## Literatur

- [BM06] Heiko Bauke und Stephan Mertens. *Cluster Computing*. Springer, Berlin [u.a.], 2006.
- [Bru96] R. Bruns. *Wissensbasierte genetische Algorithmen*. Dissertation, Carl von Ossietzky Universität Oldenburg, 1996.
- [Wei02] K. Weicker. *Evolutionäre Algorithmen*. 1.Auflage. B.G. Teubner GmbH, Stuttgart/Leipzig/Wiesbaden, 2002. 3-519-00362-7.





# Practical Secure Function Evaluation

Thomas Schneider\*

thomas.schneider@informatik.stud.uni-erlangen.de

**Abstract:** This thesis focuses on the practical aspects of general two-party Secure Function Evaluation (SFE). A new SFE protocol that allows free evaluation of XOR gates and is provably secure against semi-honest adversaries in the random oracle model is given. Furthermore, the extension of SFE to private functions (PF-SFE) using universal circuits is considered. Based on a new practical universal circuit construction, FairplayPF is implemented that extends Fairplay, a well known SFE system, with PF-SFE.

## Introduction

Consider the following situation. Several parties, each of which has a private input, wish to evaluate a function on their inputs. This need arises often indeed – almost any transaction or a communication over a network can be cast as evaluation of a function on the participant’s inputs. For example, in an online auction, the bidders and the auctioneer are players who wish to evaluate the auction function, whose value is equal to the ID of the highest bidder. Other natural and important examples include financial transactions, voting, distributed database mining, etc.

A lot of the time, parties’ inputs need to be hidden from the rest of the world. For example, in the case of the auction, unsuccessful bidders would want to preserve the privacy of their bids. Depending on the function, some information about the parties’ inputs might be easily derived from the output of the function. For example, the winning bid of an auction might necessarily be revealed. The goal of secure computation is to ensure that no other information is leaked during the computation. Clearly, efficient methods of secure evaluation of functions are of great interest. The problem is often referred to as Secure Function Evaluation, or SFE.

## Current Approaches and State of Knowledge

We first note that the general problem of SFE is a well-researched problem [LP, Yao82, Yao86]. That is, there exist well-defined protocols that allow secure computation of any

---

\*This Diplomarbeit (masters thesis) is supervised by Prof. Volker Strehl (Department of Computer Science, Universität Erlangen-Nürnberg, Germany) and Vladimir Kolesnikov, PhD (Bell Labs, Murray Hill, NJ, USA).

function, based on certain reasonable physical or complexity assumptions.

A natural and efficient method is the so-called *Yao's garbled circuit* approach (see [LP] for an excellent presentation of the subject). There, the evaluated function is viewed as a binary circuit. During the evaluation, the signals on all wires of the circuit, except for the output wire, are garbled, and thus the evaluator (one of the participants of the computation) is limited to obtaining only the output of the computation.

This known general solution is often too inefficient to be applied in practice. Nevertheless, for a large class of useful functions (especially those with efficient circuit representation), this offers the most efficient solutions. Such problems include number comparison, auctions, evaluation of conditional statements, etc. [NPS99].

Continuing advances in available computational power and communication have made secure computation of many useful functions affordable. Several recent works approach the problem of general SFE from the practical angle, discuss and fine-tune the implementation details [KJGB06, MNPS04].

## Contributions of this thesis<sup>1</sup>

This thesis focuses on the practical aspects of general two-party SFE. The currently best known approach for general SFE of Fairplay [MNPS04] is combined with the information-theoretically secure approach of Gate Evaluation Secret Sharing (GESS) [Kol05] to a new, practical method for general SFE. This new method results in substantial performance improvements of 50% for many important circuit structures like addition or number comparison. A proof of security in the semi-honest model is given that is based on the same assumptions as Fairplay, namely the existence of random oracles (RO).

In practice, there is often a need to not only protect the inputs, but the function being evaluated as well. One example is checking a passenger against the no-fly list (or, more generally, no-fly function of passenger's data). Here, a compromise of the function weakens the security of the system significantly. Other examples include credit checking or background- and medical history checking functions.

This well-known problem is called SFE of private functions (PF-SFE) and addressed by a large amount of work like [FAZ05, CCKM00, SY99, Pin02]. In PF-SFE, the evaluated function is known only by one party and needs to be kept secret (i.e. everything besides the size, the number of inputs and the number of outputs is hidden from the other party). Full or even partial revelation of these functions opens vulnerabilities in the corresponding process, exploitable by dishonest participants (e.g. credit applicants), and should be prevented.

The problem of PF-SFE can be reduced to the “regular” SFE by parties evaluating a *Universal Circuit* (UC) instead of a circuit defining the evaluated function [SY99, Pin02].

---

<sup>1</sup>Parts of this thesis will be published by Vladimir Kolesnikov and Thomas Schneider on two international conferences: “A Practical Universal Circuit Construction and Secure Evaluation of Private Functions”, Financial Cryptography and Data Security (FC08) and presumably ICALP08.

UC can be thought of as a “program execution circuit”, capable of simulating any circuit  $C$  of certain size, given the description of  $C$  as input. Therefore, disclosing the UC does not reveal anything about  $C$ , except its size. At the same time, the SFE computes output correctly and  $C$  remains private, since the player holding  $C$  simply treats description of  $C$  as additional (private) input to SFE. This reduction is the most common (and often the most efficient) way of securely evaluating private functions [SY99, Pin02].

We improve previous PF-SFE constructions by giving a new simple and efficient UC construction. Our universal circuit for simulating  $k$  gates has size  $\sim 1.5k \log^2 k$  and depth  $\sim k \log k$ . It is up to 50% smaller than the best UC (of Valiant [Val76], of size  $\sim 19k \log k$ ) for practical circuit sizes of up to  $\approx 5000$  gates. This improvement results in corresponding performance improvement of SFE of (small) private functions. Since, due to cost, only small circuits (i.e.  $< 5000$  gates) are practical for PF-SFE, our construction appears to be the best fit for many practical PF-SFE.

General PF-SFE is implemented based on this UC construction and the Fairplay SFE system [KS].

When using the improved SFE protocol of this thesis to evaluate a universal circuit, PF-SFE can be improved to approximately 25% of the previously best known solution using Fairplay as underlying SFE protocol.

The results of this thesis substantially improve general SFE for many practical circuits and demonstrate practicability of general PF-SFE for “small” functions.

## References

- [CCKM00] Christian Cachin, Jan Camenisch, Joe Kilian, and Joy Müller. One-Round Secure Computation and Secure Autonomous Mobile Agents. In *ICALP '00*, pages 512–523, London, UK, 2000. Springer-Verlag.
- [FAZ05] Keith Frikken, Mikhail Atallah, and Chen Zhang. Privacy-preserving credit checking. In *EC '05: Proceedings of the 6th ACM conference on Electronic commerce*, pages 147–154, New York, USA, 2005. ACM Press.
- [KJGB06] Louis Kruger, Somesh Jha, Eu-Jin Goh, and Dan Boneh. Secure function evaluation with ordered binary decision diagrams. In *CCS*, pages 410–420. ACM Press, 2006.
- [Kol05] Vladimir Kolesnikov. Gate Evaluation Secret Sharing and Secure One-Round Two-Party Computation. In *Advances in Cryptology – ASIACRYPT 2005*, volume 3788 of *LNCS*, pages 136–155. Springer, 2005.
- [KS] Vladimir Kolesnikov and Thomas Schneider. FairplayPF. <http://thomaschneider.de/FairplayPF>.
- [LP] Yehuda Lindell and Benny Pinkas. A proof of Yao’s protocol for secure two-party computation. Cryptology ePrint Archive, Report 2004/175.
- [MNPS04] D. Malkhi, N. Nisan, B. Pinkas, and Y. Sella. Fairplay — a secure two-party computation system. In *USENIX*, 2004.

- [NPS99] Moni Naor, Benny Pinkas, and Reuben Sumner. Privacy Preserving Auctions and Mechanism Design. In *EC'99*, pages 129–139, 1999.
- [Pin02] Benny Pinkas. Cryptographic techniques for privacy-preserving data mining. *SIGKDD Explor. Newsl.*, 4(2):12–19, 2002.
- [SYY99] Tomas Sander, Adam Young, and Moti Yung. Non-Interactive CryptoComputing for  $NC^1$ . In *Proc. 40th IEEE Symp. on Foundations of Comp. Science*, pages 554–566, New York, 1999. IEEE.
- [Val76] Leslie G. Valiant. Universal circuits (Preliminary Report). In *Proc. 8th ACM Symp. on Theory of Computing*, pages 196–203, NY, USA, 1976.
- [Yao82] Andrew C. Yao. Protocols for Secure Computations. In *Proc. 23rd IEEE Symp. on Foundations of Comp. Science*, pages 160–164, Chicago, 1982.
- [Yao86] Andrew C. Yao. How to Generate and Exchange Secrets. In *Proc. 27th IEEE Symp. on Foundat. of CS*, pages 162–167, Toronto, 1986.

# Breitensuche in Prolog

Jürgen Sorg

Fachhochschule Aachen  
Elektrotechnik und Informationstechnik  
juergensorg@gmx.net

Art der Arbeit: Diplomarbeit

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

**Abstract:** Kapitel 1 gibt eine kurze Einführung in das Thema Prolog. Herkömmliche Prolog-Implementierungen beruhen auf der Warren Abstract Machine. Da auch die hier vorgestellte Implementierung von der WAM ausgeht, werden deren wichtigsten Komponenten in Kapitel 2 nochmal kurz vorgestellt, um dann in Kapitel 3 auf die Erweiterungen für die Implementierung der Breitensuche einzugehen. Kapitel 4 beschreibt dann einige Kernpunkte der Übersetzung, bevor in Kapitel 5 auf die Implementierung als Java-Klasse eingegangen wird.

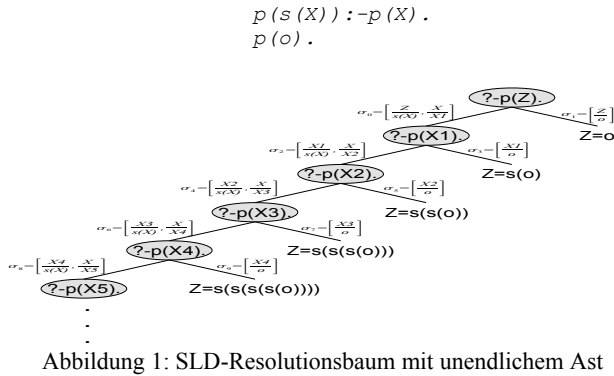
## 1 Einleitung

Prolog ist eine logische Programmiersprache und zählt zu den deklarativen Sprachen. Die Grundprinzipien von Prolog beruhen auf der Prädikatenlogik erster Stufe (PL1). In Prolog geschriebene Programme sind Sammlungen von Hornklauseln, die eine Untermenge der Prädikatenlogik erster Stufe darstellen. Das durch Hornklauseln formulierte Programm wird als Datenbasis bezeichnet, an das der Anwender Anfragen stellen kann.

```
grossvater(X,Y) :- vater(X,Z), vater(Z,Y).
```

Eine (Horn) Klausel besteht aus einem Kopf und einem optionalem Rumpf. Falls die Klausel keinen Rumpf besitzt wird sie auch Faktum oder Fakt genannt. Das Prolog-System versucht mit Hilfe der Hornklauseln die Anfragen zu beweisen bzw. abzuleiten. Beim Ableiten einer Anfrage wird ein Baum erzeugt, der SLD-Resolutionsbaum genannt wird. Eine Lösung für eine Anfrage, die dem System gestellt wurde, entspricht einem Pfad von der Wurzel bis zu einem Blatt. Beim Ableiten müssen Terme unifiziert werden, was mit Hilfe eines Unifikationsalgorithmus geschieht. Dabei werden Variablen gebunden oder Terme auf Gleichheit geprüft. Der oben beschriebene Ableitungsbaum wird von den gängigen Prolog-Compilern mit einem Tiefensuchlauf durchlaufen um eine Ableitung zu finden. Da es auf Grund der Unentscheidbarkeit der Hornklauseln im SLD-Resolutionsbaum zu einem unendlichen Ast kommen kann, ist es für diese Compiler

nicht möglich alle Lösungen zu finden. Die Tiefensuche in einem SLD-Resolutionsbaum ist somit keine vollständige Beweisstrategie. In dem vorliegenden Abstract wird gezeigt wie ein Prolog-System entwickelt werden kann, dass als Beweisstrategie die Breitensuche benutzt. Die Breitensuche in einem SLD-Resolutionsbaum ist eine vollständige Beweisstrategie für ein Prolog-System, somit werden mit dieser Strategie theoretisch alle endlichen Ableitungen für ein Ziel gefunden. In Abbildung1 ist ein Beispiel für einen Ableitungsbaum mit einem unendlichen Ast angegeben. Das Prädikat ist eine induktive Definition der natürlichen Zahlen in symbolischer Darstellung in Prolog.



## 2 Warren's Abstract Machine (WAM)

Die Warren Abstract Machine (WAM) wurde 1983 von David H. D. Warren spezifiziert. Sie ist das gängige Standardausführungsmodell (vgl. [BOGAN]) für Prolog. Die WAM ist keine reale Maschine, sie wird als Software implementiert und wird deshalb auch als virtuelle Maschine bezeichnet. Die WAM besteht aus 6 Speicherbereichen und 9 Zeigerregistern, welche in Abbildung2 dargestellt sind. Die Instruktionen, die das übersetzte Prologprogramm darstellen, sind im CODE Segment angesiedelt. Im HEAP werden die Terme der Anfragen (queries) und der Regelrümpfe erzeugt, die dann mit Regelköpfen, respektive Fakten, unifiziert werden. Der STACK wird benutzt um Variablen zu sichern, die in einer Regel mehrmals vorkommen und dadurch überschrieben werden könnten. Außerdem werden im STACK Werte der Zeigerregister gehalten, die nach Abarbeitung eines Literals wieder zurück in die Zeigerregister geschrieben werden müssen. Der Speicherbereich TRAIL wird für das Backtracking-Konzept benutzt, welches für die Breitensuche im SLD-Resolutionsbaum nicht relevant ist. Der Speicherbereich X-REGISTER setzt sich aus mehreren Registern zusammen, die dazu benutzt werden Argumente an aufzurufende Klauseln bzw. Fakten zu übergeben. Dies entspricht ungefähr dem Prinzip der Parameterübergabe aus herkömmlichen Programmiersprachen (JAVA, C, ...), wobei jedoch ein bidirektionaler Unifikations-Algorithmus verwendet werden muss, da die Queries üblicherweise ebenfalls Variablen enthalten. Der PDL ist eine einfache LIFO-Datenstruktur welche für den

Unifikationsalgorithmus benutzt wird. Zu den Speicherbereichen gesellen sich noch 9 Zeigerregister, die jeweils Zeiger in die verschiedenen Bereiche darstellen.

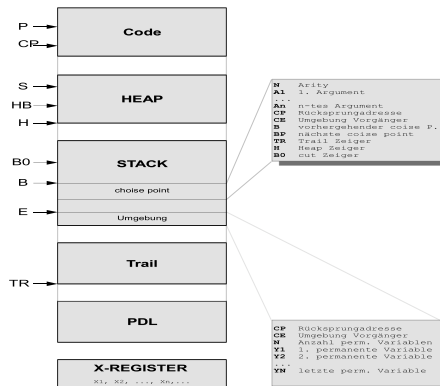


Abbildung 2: Speicherlayout WAM

### 3 WAM für die Breitensuche (JWAM)

Für die Breitensuche im Ableitungsbaum wurden die Konzepte der WAM modifiziert bzw. erweitert und die daraus entstandene Maschine wird mit JWAM bezeichnet und ist in Abbildung3 zu sehen. Die Instruktionen (CODE) der WAM wurden so erweitert das der Baum in der Breite durchlaufen wird. Die Speicherbereiche STACK, HEAP und PDL wurden nahezu identisch aus der WAM übernommen, mussten allerdings wegen den parallel existierenden Variablensubstitutionen geklont werden. Da der Unifikationsalgorithmus übernommen wurde, ist auch der PDL identisch mit dem der WAM.

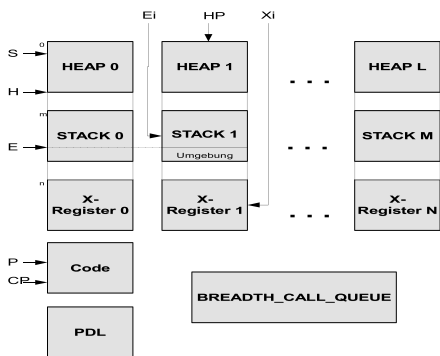


Abbildung 3: Speicherlayout JWAM

Nachfolgend ist der Algorithmus für die Breitensuche im Ableitungsbaum angegeben:

- (1) Schreibe alle direkten Nachfolger in die **BREADTH\_CALL\_QUEUE**.
- (2) Falls die **BREADTH\_CALL\_QUEUE** nicht leer ist:  
 Merke den ersten Eintrag in der **BREADTH\_CALL\_QUEUE**  
 Sonst: Fertig



- (3) Lösche den ersten Eintrag in der BREADTH\_CALL\_QUEUE.
- (4) Springe zum gemerkten Knoten.

## 4 Übersetzung (vgl. [BOGAN])

Für die Übersetzung eines Prolog Programms ist es sinnvoll die Tatsache auszunutzen, dass ein Prolog-Programm als Baumstruktur dargestellt werden kann. Die Baumdarstellung ermöglicht es die richtigen Instruktionen, in der richtigen Reihenfolge zu ermitteln, indem der Baum traversiert wird. Erzeugt wird der Baum indem der Quellcode einer syntaktischen Analyse unterzogen wird und in wohl definierten Situationen (in Grammatikregeln), sukzessive der Baum entwickelt wird. Der so erzeugte Baum wird dann traversiert und die benötigten Instruktionen werden erzeugt. Es ist bei dieser Vorgehensweise nicht relevant ob die erzeugten Befehle direkt in die WAM, respektive JWAM, geladen werden oder zunächst in einer Datei gespeichert werden. Damit kann mit obig beschriebenem Vorgehen sowohl ein Compiler als auch ein Interpreter für Prolog formuliert werden.

## 5 Vom JWAM-Code zu einer ausführbaren JAVA-Klasse

Teil dieser Arbeit ist eine Referenzimplementierung (*prolovis*), welche die JWAM, sowie den Compiler und den Interpreter in JAVA implementiert. Um eine Ausführbare JAVA Klasse zu erstellen, werden die JWAM-Instruktionen in einer Klasse gekapselt. Jede Instruktion ist als JAVA-Methode implementiert. In der gekapselten Klasse werden die Instruktionen (Methoden) in einem switch-case Konstrukt in einer Endlosschleife durchlaufen. Der so erzeugte Quellcode der Klasse wird im Dateisystem gespeichert und der JAVA Compiler *javac* wird aufgerufen um die Klasse zu kompilieren. Nachdem die *class*-Datei erstellt wurde, wird sie mit Hilfe der JAVA-REFLECTION API in die JWAM geladen.

## Literaturverzeichnis

- [HANUS] Prof. Dr. Michael Hanus. Skript zur Vorlesung Implementierung logischer Programmiersprachen, 1997, RWTH Aachen Lehr- und Forschungsgebiet Informatik II
- [KACI] Hassan Aït-Kaci. Warren's Abstract Machine / A Tutorial Reconstruction, 1991. MIT Press Cambridge, Massachusetts London, England
- [BOGAN] Sebastian Bogan Diplomarbeit, Entwicklung eines Compilers für eine Prolog-Erweiterung zur Programmierung paralleler Algorithmen, 2003. Technische Universität Berlin, Fakultät IV Elektrotechnik und Informatik, Fachgebiet Wissensbasierte Systeme

# Keyframe Extraction for Video Tagging & Summarization

Damian Borth, Adrian Ulges, Christian Schulze, Thomas M. Breuel  
damian.borth@dfki.de  
German Research Center for Artificial Intelligence (DFKI)  
and University of Kaiserslautern

**Abstract:** Currently, online video distributed via online video platforms like YouTube experiences more and more popularity. We propose an approach of keyframe extraction based on unsupervised learning for video retrieval and video summarization. Our approach uses shot boundary detection to segment the video into shots and the k-means algorithm to determine cluster representatives for each shot that are used as keyframes. Furthermore we performed an additional clustering on the extracted keyframes to provide a video summarization. To test our methods we used a database of videos downloaded from YouTube where our results show (1) an improvement of retrieval and (2) compact summarization examples.

GI-Topic: KI-BV (artificial intelligence - image understanding)

## 1 Introduction

Searching video media on the web often means to use the available search interfaces provided by online video portals. YouTube as the market leader provides keyword search based on manually generated meta information and tags. Unfortunately, meta data is limited in its ability of representing the content of a video, and tags are subjective labels that might be misleading in their semantics. This is why content-based video retrieval (CBVR) can improve video search. One key component of CBVR is the extraction of keyframes which then can be analyzed with known image processing algorithm. The resulting features can be indexed and used for further retrieval. As we will show the quality of keyframe extraction influences the overall performance of retrieval systems. Additionally, keyframes can be used to summarize search results and help the user to quickly evaluate the relevance of a video [BMN<sup>+</sup>03] and for video browsing like in [ROS04], where keyframes are organized along a temporal and visual plane to each other.

## 2 Our Approach

Video is by its nature a temporally structured media and the segmentation into its basic temporal units (shots) is usually the first step in the processing pipe of content-based retrieval systems. Based on the detected shots there exist several keyframe extraction approaches which then are used as representatives of the shot. A straightforward approach is

to take the first, middle or last frames of a shot as keyframes [O’C91], which might result in information loss within long shots. To overcome this problem Hammoud and Mohr [HM00] proposed unsupervised learning techniques that adapt to the content to the shot. Our approach is a combination of shot boundary detection and an intra-shot clustering of frames to find an adequate number of representative keyframes for the given shot with respect to its visual complexity.

**Shot Boundary Detection** Among the many different approaches that exist in temporal segmentation [KC01], we use the MPEG-7 Color Layout Descriptor (CLD) as a feature [MOVY01] for each frame and compute differences between consecutive frames. The detection of shot boundaries follows an adaptive threshold technique as stated in [Lie01]. This method works very well for hard cuts but leaks in performance for long dissolves and fades.

**Intra-Shot Clustering** Modelling every shot as a mixture of Gaussian densities we use the k-means algorithm [McQ67] for keyframe extraction on the CLD feature vectors extracted from every frame of a shot. After clustering we define the nearest frame to the mean of every cluster as a representative for the shot. To estimate the number of clusters and therefore the number of keyframes we use the Bayesian Information Criterion (BIC) [Sch78] for cluster validity measurement. According to the visual complexity of the shot this might lead to multiple keyframes. Figure 1 shows an example with a feature space representation of a complex shot in the center (for visualization purpose reduced to 2d). Visually similar frames are clustered together and their representatives are shown on the left and right side of the image.

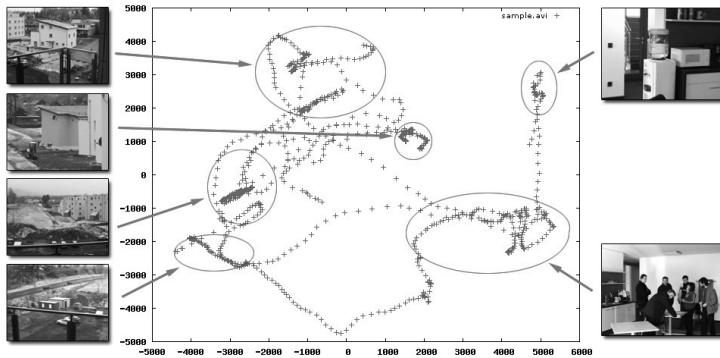


Figure 1: Feature space representation of a complex shot (center). For visualization purpose the feature space was reduced to 2d by PCA. A clustering of the frames leads to keyframes displayed on the left and right side.

**Keyframe "Meta-Clustering"** A shortcoming of this two step process is the duplication of the same content by multiple keyframes due to the missing inter-shot reasoning. This occurs often when actors have a dialog, or in music video clips where the artist is shown in different shots throughout the video. This tradeoff is addressed in the context of video summarization where we propose a "meta-clustering" on the keyframes extracted in the previous steps. Similar to [BMN<sup>+</sup>03] we use the idea of grouping keyframe together to create a summary of the video. But in contrast to creating a hierarchical structure defined by groups and scenes, we directly use the already extracted keyframes and cluster them using the k-means algorithm.

### 3 Experiments & Results

**Tagging** The keyframe extraction method was used in an experiment for automatic tagging of video material on a database of 2200 videos (total duration 194 hrs.) downloaded from YouTube. In this experiment a video tagger<sup>1</sup> learned concepts like *cat*, *sailing*, *desert* out of a training set of already available tags from YouTube and reached a Mean Average Precision (MAP) of 34.2% on unknown test videos. As features for the experiment we used different combination of color histograms, Tamura features and Visual Words extracted from the given keywords and motion descriptors extracted on shot level. To investigate the impact of the keyframe extraction for a successful tagging we performed the tagging experiment with two additional keyframe extraction methods: [first], where we take the first frame of a shot as representative and [regular], where we regularly sample keyframes within an interval of 7 sec. Based on the chosen combination of used features the results show that our approach increases performance by 2-9% compared to [first] and was as performant as [regular] but extracting less keyframes.

**Summarization** We performed "meta-clustering" on several music video clips, material that is characterized by a high amount of cuts and redundant shots. We obtained compact summarizations by grouping duplicate keyframe together. Pure clustering on the entire video without shot distinction was leading to roughly the same amounts of keyframes but seems to miss some content. Figure 2 shows an example of a music video clip where duplicates were successfully clustered together. The first row of the Figure 2 displays the result of our "meta-clustering" based on the previous extracted keyframes, which are shown in the third row. The second row displays the keyframe extracted by a pure clustering based on the entire video. We obtained comparable results for other videos.

---

<sup>1</sup>described in more detail in [USKB07]

Summarization via Meta-Clustering on Keyframes



Clustering without Shot Boundary Detection



Shot Boundary Detection + Clustering



Figure 2: Result of summarization for a music video clip. First row displays "meta-clustering". Second row displays the pure clustering. Third row displays keyframes extracted by our basic method

## References

- [BMN<sup>+</sup>03] W. Bailer, H. Mayer, H. Neuschmied, W. Haas, M. Lux, and W. Klieber. Content-based video retrieval and summarization using MPEG-7. *Internet Imaging V. Edited by Santini, Simone; Schettini, Raimondo. Proceedings of the SPIE*, 5304:1–12, 2003.
- [HM00] R. Hammoud and R. Mohr. A Probabilistic Framework of Selecting Effective Key-Frames for Video Browsing and Indexing. *Intern. Worksh. on Real-Time Img. Seq. Anal.*, pages 79–88, 2000.
- [KC01] I. Koprinska and S. Carrato. Temporal Video Segmentation: A Survey. *Signal Processing Image Communication*, 16:477–500, 2001.
- [Lie01] R. Lienhart. Reliable Transition Detection in Videos: A Survey and Practitioner's Guide. *International Journal of Image and Graphics*, 1(3):469–486, 2001.
- [McQ67] J.B. McQueen. Some methods for classification and analysis of multivariate observations. *In Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, pages 281–297, 1967.
- [MOVY01] B.S. Manjunath, J.R. Ohm, V.V. Vasudevan, and A. Yamada. Color and Texture Descriptors. *IEEE Trans. on Circuits Syst. for Video Techn.*, 11(6), 2001.
- [O'C91] Brain O'Connor. Selecting Key Frames of Moving Image Documents: A Digital Environment for Analysis and Navigation. *Microcomputers for Information Management*, 8(2):119–33, 1991.
- [ROS04] M. Rautiainen, T. Ojala, and T. Seppanen. Cluster-temporal browsing of large news video databases. *IEEE Int. Conference on Multimedia and Expo*, 2:751–754, 2004.
- [Sch78] G. Schwarz. Estimating the dimension of a model. *Annals of Statistics*, 6(2):461–464, 1978.
- [USKB07] A. Ulges, C. Schulze, D. Keysers, and T.M. Breuel. Content-Based Video Tagging for Online Video Portals. *In MUSCLE/Image-CLEF Workshop*, 2007.

# Informationsverarbeitung in spikenden Neuronalen Netzen durch Synchronisationen

Aleksander Koleski

Universität Hamburg, MIN Fakultät  
Department Informatik, Department Biologie  
koleski@biokybernetik.uni-hamburg.de

Art der Arbeit: Diplomarbeit

Betreuer/in der Arbeit: Prof. Dr.-Ing. Bernd Page (Department Informatik), Prof. Dr. Andreas G. Fleischer (Department Biologie)

**Abstract:** Informationsverarbeitung in neuronalen Systemen erfolgt generell durch viele parallel arbeitende Neurone. Zusätzlich befindet sich in biologischen neuronalen Systemen eine hohe Anzahl von Aktionspotentialen (Spikes), die sowohl örtlich als auch zeitlich verteilt auftreten können. Der vorliegende Beitrag untersucht die Gesetzmäßigkeiten, nach denen parallel wirkende biologische Neurone Informationen durch Aktionspotentiale verarbeiten können. Dabei spielt das synchrone Auftreten der Aktionspotentiale eine entscheidende Rolle in der Informationsverarbeitung. Für die Untersuchung wurden biologisch motivierte rekurrente neuronale Netze verwendet, mit denen verrauschte Muster erkannt und Mustersequenzen reproduziert wurden.

## 1 Einführung

Das Gehirn ist eine der komplexesten Strukturen. Dabei entsteht die Komplexität nicht nur durch die hohe Anzahl der beteiligten Neurone (ca.  $10^{10}$ ) oder durch die noch höhere Anzahl an Synapsen (ca.  $10^{14}$ ) [Sc98], sondern auch durch die hohe Dynamik der örtlich und zeitlich verteilten Spikes. Der Begriff „Spike“ ist eine andere Bezeichnung für Aktionspotential, welches die Potentialveränderung an der Zellmembran einer Zelle bezeichnet. Spikes werden durch Neurone erzeugt und dienen der Informationsübertragung zwischen den Neuronen. Die zeitliche Ausdehnung eines Spikes beträgt ca. 1 ms und die maximale Spikefrequenz je nach Zelltyp zwischen 50-500 Hz [Sc98]. Die Informationsverarbeitung im Gehirn geschieht parallel, durch parallel arbeitende Neurone und verteilt, durch die verteilte Verarbeitung in unterschiedlichen Gehirnarealen. Ein Grund für die Art der Verarbeitung ist, dass das Neuron ein einfaches Element ist, welches die Eingabesignale aufsummiert und bei Erreichen einer bestimmten Schwelle ein Aktionspotential erzeugt. Gerade wegen der Einfachheit eines Elementes muss das Netzwerk bestimmte Verarbeitungsverfahren aufweisen, damit insgesamt eine komplexe und fehlertolerante Informationsverarbeitung stattfinden kann. Als Verfahren der Informations-

verarbeitung, die auch in neuronalen Systemen stattfinden kann, wurden die Mustererkennung und die kontextabhängige Aktivierung von Mustersequenzen gewählt [HE05]. Die erste Hypothese besagt, dass die Mustererkennung in spikenden neuronalen Netzen durch die Synchronisation von Spikes stattfinden kann. Die zweite Hypothese besagt, dass die Aktivierung einer Mustersequenz durch synchronisierte Spikes erfolgen kann.

## **2 Methoden**

### **2.1 Neuronen-Modelle**

Alle Simulationen wurden mit Simulink-Modellen durchgeführt und mit MATLAB-Skripten analysiert. Als Modell für ein Neuron, welches ein Aktionspotential erzeugt, wurde ein Modell von Wilson [Wi99] verwendet. Das Wilson-Modell ist eine Vereinfachung des Hodgkin-Huxley-Modells [HH52], wobei das Wilson-Modell die Eigenschaft der Ionenkanäle, sich nur in Abhängigkeit der Membranspannung zu öffnen, beibehält. Durch die Vereinfachung des Wilson-Modells konnte die Geschwindigkeit der Simulation mindestens um den Faktor 2 gesteigert werden, was bei einer Anzahl von 256 simulierten Neuronen nicht zu vernachlässigen ist.

### **2.2 Gewichtsmatrix**

Die zu reproduzierende Information für die Mustererkennung und die Mustersequenzreproduktion wurde vor der Simulation in der Gewichtsmatrix gespeichert. Für die Mustererkennung entstand die Gewichtsmatrix durch das äußere Produkt jedes Musters mit sich selbst und der anschließenden Addition und Normierung der daraus entstandenen Matrizen. Dies entspricht der Autoassoziation der vier Muster. Die Gewichtsmatrix für die Mustersequenzreproduktion entstand durch das äußere Produkt jeweils zwei aufeinanderfolgender Muster und der anschließenden Addition und Normierung. Dies entspricht der Kreuzassoziation der vier Muster. Das Prinzip der Erzeugung der Gewichtsmatrix entspricht dabei der Verstärkung der synaptischen Verbindung zwischen zwei Neurone und damit dem Lernen von neuen Informationen. Denn nach Hebb [He06] wird eine Synapse verstärkt, wenn das präsynaptische und postsynaptische Neuron gleichzeitig aktiv sind. Inzwischen ist bekannt, dass eine Verstärkung der synaptischen Verbindung nur stattfindet, wenn das Aktionspotential des präsynaptischen Neurons die Synapse erreicht, bevor das postsynaptische Neuron ein Aktionspotential erzeugt [GK02]. Die Verbindung wird jedoch abgeschwächt, wenn das präsynaptische Neuron ein Aktionspotential erzeugt, nachdem das postsynaptische Neuron ein Aktionspotential erzeugte.

### **2.3 Rekurrente Neuronen-Population**

Für die Simulationen wurde ein rekurrentes Netz verwendet, welches dem Hopfield-Modell ähnelt. Dadurch ergab sich der Vorteil, dass bei Einhaltung bestimmter Regeln sichergestellt werden konnte, dass die Netzaktivität immer zu einem lokalen Minimum

konvergiert. Dabei wird der stabile Endzustand Attraktor genannt. Außerdem weist ein rekurrentes Netz eine Ähnlichkeit mit dem Hippocampus auf, einer Struktur im Gehirn, so dass die Ergebnisse verglichen werden können mit denen aus der Hippocampus-Forschung [HE05].

## **2.4 Synaptische Übertragung**

Da in der Simulation der Neuronenpopulation möglichst viele physiologische Strukturen und Mechanismen umgesetzt werden sollten, wurde für die postsynaptischen Potentiale ein Modell verwendet, welches Rall [Ra67] aufstellte und das bekannt ist unter der Bezeichnung „ $\alpha$ -Funktion“. Die  $\alpha$ -Funktion entspricht einem Tiefpass 2. Ordnung.

## **2.5 Spike-Rauschen**

Für die Erzeugung des Hintergrundrauschens wurde eine Neuronenschicht simuliert, die zufällig verteilte Aktionspotentiale mit einer definierbaren Häufigkeit erzeugte. Um die Gesamtaktivität der Neuronenpopulation im Arbeitsbereich zu halten, befand sich eine inhibitorische Neuronenschicht im Populationsmodell, welche an den Synapsen inhibitorische postsynaptische Potentiale (IPSP's) erzeugte.

## **3 Ergebnisse**

In den Simulationen der Mustererkennung konnte gezeigt werden, dass die Population nach dem kurzzeitigen Anlegen eines verrauschten Eingabemusters das Zielmuster, das in der Gewichtsmatrix kodiert war, aktivierte. Dabei wurde das Zielmuster durch die Synchronizität aller Spikes repräsentiert. Trafen allerdings zu viele, durch das Rauschen erzeugte, zufällig verteilte Spikes in einer gewissen Häufigkeit auf die Neuronenschicht, führte dies zu einer immer stärker werdenden Desynchronisation der Spikes, so dass das Muster nicht mehr repräsentiert werden konnte. Das Netz verließ dabei den stabilen Zustand, was dem Verlassen des Attraktors gleichkommt. Während der Simulation, in der eine Mustersequenz wiedergegeben werden sollte, konnten alle Muster korrekt aktiviert und in der richtigen Reihenfolge wiedergegeben werden. Das Anlegen des Spike-Rauschens führte hier ebenfalls zur Desynchronisation der Spikes, wodurch weder die Muster aktiviert noch die Mustersequenz korrekt wiedergegeben werden konnte.

## **4 Diskussion**

Die in der Einleitung erwähnten Hypothesen, konnten durch die Ergebnisse erfolgreich überprüft werden. Dabei konnte gezeigt werden, dass durch die Synchronisation der Spikes sowohl eine Mustererkennung als auch eine Mustersequenzreproduktion möglich ist. Sobald die Spikes nicht mehr synchron waren, konnte keine der beiden Informationsverarbeitungsprinzipien beobachtet werden.



Das synchrone Auftreten von Spikes ist aber nicht die einzige Möglichkeit, Informationen zu repräsentieren. Die Spikeratenkodierung spielt z.B. eine wichtige Rolle im sensorischen und motorischen System [GR03]. Allerdings ist die Kodierung von Informationen durch unterschiedliche Spikeraten nicht geeignet, um das sogenannte Bindungsproblem zu erklären [Si98]. Denn das beobachtete Phänomen der Aktivitätssynchronisation zweier Kortexareale kann nicht durch die Spikeratenkodierung erklärt werden.

Spikes dienen aber nicht nur zur Repräsentation von Informationen oder zur Aktivierung der nächsten Information in einer Sequenz, sondern auch zur Stabilisierung der Spikeaktivität [GK05]. Außerdem ermöglicht die Synchronisation von Spikes das Lernen an den Synapsen [He06].

## 5 Ausblick

Diese Arbeit konnte zeigen, dass durch die Synchronisation von Spikes Informationen sowohl kodiert, als auch aktiviert werden können. Dafür wurde eine Neuronenpopulation nach dem Verschaltungsprinzip des Hippocampus verwendet. Das Gehirn ist außerdem aus vielen Arealen aufgebaut, wobei Aktivitätssynchronisationen zwischen verschiedenen Arealen beobachtet werden. In einem nächsten Schritt könnte durch die Simulation mehrerer Areale untersucht werden, welche Mechanismen für die Synchronisation der Aktivität zwischen verschiedenen Arealen zuständig sein könnten.

## Literaturverzeichnis

- [GK02] Gerstner, W., Kistler, W.: Mathematical formulations of Hebbian learning, *Biological Cybernetics* 87, 404-415, 2002
- [GK05] Gerstner, W., Kistler, W.: *Spiking Neuron Models – Single Neurons, Populations, Plasticity*. Cambridge University Press, Cambridge, 2005
- [GR03] Grammont, F., Riehle, A.: Spike synchronization and firing rate in a population of motor cortical neurons in relation to movement direction and reaction time, *Biological Cybernetics* 88, 360-373, 2003
- [HE05] Hasselmo, M. E., Eichenbaum, H.: Hippocampal mechanisms for the context-dependent retrieval of episodes. *Neural Networks* 18(9), 1172-1190, 2005
- [He06] Hebb, D.E.: *The Organization of Behavior*. Appleton & Lange, East Norwalk, Third Edition, 2006
- [HH52] Hodgkin, A. L., Huxley, A. F.: A quantitative description of membrane current and its application to conduction and excitation in nerve, *J. Physiol.* 117, 500-544, 1952
- [Ra67] Rall, A.: Distinguishing theoretical synaptic potentials computed for different somadendritic distributions of synaptic input. *Journal of Neurophysiology* 30, 1138-1168, 1967
- [Sc98] Schmidt, R.F.: *Neuro- und Sinnesphysiologie*. Springer Verlag, Berlin, 3. Auflage, 1998
- [Si98] Singer, W.: Synchronization of Neuronal Responses as a Putative Binding Mechanism, In: Arbib, M. A. (eds.): *The Handbook of Brain Theory and Neural Networks*, MIT Press, 960-964, 1998
- [Wi99] Wilson, H.R.: *Spikes, decisions, and actions – The dynamical foundation of neuroscience*, Oxford University Press, Oxford, 3. Auflage, 1999

# Absolute Measurement Systems Based on Industrial Robots

Katayon Radkhah and Prof. Oskar von Stryk  
Technische Universität Darmstadt  
Department of Computer Science  
Simulation, Systems Optimization and Robotics Group  
{radkhah, stryk}@sim.tu-darmstadt.de

**Abstract:** Robotic measurement has become a hotly contested market. Due to manufacturing competitiveness and economical criteria, it plays an increasingly important role in both automobile and general industry. Particularly in the area of mounting and spot-welding of body-in-white, absolutely accurate robots are installed as measurement instruments replacing expensive coordinate measuring machines. Prior to such deployment, however, it needs to be ensured that the internal robot model fits the real model. Existing approaches to calibration fail in capturing both geometric and elastic effects appearing in the robot forward kinematics. Furthermore, they require expensive metrology equipments.

In my diploma thesis, titled “Model-Based Approach to Calibration of Industrial Robots Considering Geometric and Elastic Effects” and supervised by Prof. von Stryk, a feasibility study on achieving absolute positioning accuracy through static calibration of industrial robots with revolute joints has been conducted. The attained results demonstrate that the proposed model and procedure for the retrieval of the real model parameters enable the effective and significant improvement of the robot’s accuracy without the necessity of special and expensive metrology equipment.

## 1 Background and Industrial Motivation

Automated manufacturing involves more and more complex systems integration problems. In order to achieve the highly automotive and precise accomplishment of tasks, it is often necessary to redesign the entire work-cell and analyze the assembly process itself. The increasing task complexity of robots raised particularly the interest in off-line programming. Off-line programming, however, cannot have the promising effects and advantages as long as the internal robot model does not fit the real model. Hand in hand with the increasing interest in off-line programming goes therefore the interest in enhancing robot accuracy. As the automobile industry represents one of the most important industrial sectors, absolutely accurate industrial robots are crucial. In the area of mounting and spot-welding of bodywork, accurate manipulators can be installed as measurement tools. The capturing of welded points by such control robots represents a possible area of application. Incorrectly welded points and their late notice lead to problems and delay times of the production. It is crucial to notice such problems early in the production process. Reliable and fast warning systems, though, require an absolutely accurate robot. In order to improve the accuracy

of an industrial robot, robot calibration needs to be applied. It is a very practical problem facing those involved in the implementation of advanced automation.

## 2 Research Objectives and Contributions

The goal of my diploma thesis is to establish the theoretical and experimental basis for the deployment of industrial robots with revolute joints as measurement instruments, which could take over the role of fast and reliable warning systems. Particularly the automobile industry is highly interested in such warning systems for noticing incorrectly welded points on body-in-white early in the production process. Such measurement tools are also objects of interest for various other applications in automobile and general industry. Prior to such a deployment, however, it needs to be ensured that the real robot model matches the internal model. Robot calibration can offer a significant opportunity to improve the positioning accuracy and to cut costs of the production.

Existing approaches often make use of external measurement systems for the calibration of a given robot. The usage of an external measurement system, however, leads the idea of deploying a robot as a measurement tool to a contradiction. We propose a completely novel approach to both modeling of the forward kinematics and retrieving the real parameter values.

A parameterized extended forward kinematic model covering all physical properties and effects that influence the static and time-invariant positioning accuracy of a given manipulator is developed and implemented. In addition, the corresponding extended Jacobian matrices are developed and implemented in order to enhance the accurate retrieval of the real parameter values. Besides, a complete optimal procedure for the calibration of the parameters of the extended forward kinematic model is presented. We suggest a novel problem formulation of the occurring positioning deviations avoiding two disadvantages inherent to the existing approaches:

1. a difficult, time-consuming, and cost-intensive determination of the robot base, and
2. the usage of the internal robot error model.

The novel problem formulation facilitates an easier and a more cost-effective experimental setup. The new objective function represents the second important component of the proposed procedure for the calibration of industrial robots. Instead of absolutely accurate positioning data, only relatively accurate positioning data are required. Besides, in order to validate the novel problem formulation, the commonly used problem formulation is adapted to the proposed experimental setup.

The experimental setup consists of an industrial robot with revolute joints, a camera that is attached to the robot flange, and several appropriate calibration objects. The actual position of the robot flange is retrieved only indirectly by the attached camera system. It is therefore also not necessary that the flange is visible.

By means of simulatively generated data the nonlinear least squares regression equations are applied to a typical industrial robot. Both problem formulations are extensively

tested and the achieved accuracy ranges are compared. In these tests, also the identifiability of each parameter is investigated. Highly accurate results are achievable with both approaches, particularly with the novel approach using only a simple camera system. The experimental setup is simulated using a visualization tool. The most important objective of the simulations is to confirm that the numerical algorithm for the identification converges to the correct values.

To the best of our knowledge, this thesis is the first to investigate explicitly the identifiability of both the geometric and nongeometric parameters of an absolutely accurate model of a given robot arm with revolute joints. The proposed model-based approach considers a parameter set, such that the calculated robot end effector position matches the real robot end effector position. After compensation the robot end effector can be found at the desired position. Furthermore, this thesis presents a completely novel problem formulation for the identification of the desired parameters. To conclude, the work related to this thesis can be considered as a “building-block” contributing to the research on model-based calibration of industrial robots with revolute joints. This thesis proves by means of simulative data that a cost-minimal and effective realization of an absolutely accurate robot as a measurement tool is feasible.

### 3 Thesis Outline

My diploma thesis consists of seven chapters. **Chapter 1** gives an overview of the background, industrial motivation, and the contributions. In **Chapter 2** we clarify terms and definitions that are essential for a proper understanding of the thesis. Mathematical **basics, tools, and notations** used to develop the robot kinematic model are introduced. **Chapter 3** reviews the necessary techniques and methodologies used in calibration procedures. It presents a survey of the different kinematic modeling approaches and measurement methodologies. The existing calibration and compensation methods are summarized. The extended robot kinematic model is developed in **Chapter 4**. The standard **kinematic model** is extended to consider two important error sources, geometric and nongeometric issues. **Parameter identification** of the extended kinematic model and the subsequent **compensation** of the positioning deviations are the focus of **Chapter 5**. Different problem formulations and calibration methodologies are discussed. Numerical aspects and optimization issues, such as optimal starting values, are also subject of this chapter. With the knowledge of the desired reference positions and the real robot end effector positions, predicted by the developed model, the errors are corrected in the compensation step as a function of the joint configuration. **Chapter 6** discusses the replicated **experimental setup**. The developed model and procedure are applied to a 6 - DOF manipulator with revolute joints. This chapter also outlines the simulations and presents the obtained **numerical results**. **Chapter 7** concludes the presented work with a summary of the **contributions**. Furthermore, **suggestions** for possible future work are given.

## Literatur

- [BA93] R. Bernhardt und S. Albright. *Robot Calibration*. Chapman & Hall, London, 1993.
- [Bey04] L. Beyer. Genauigkeitssteigerung von Industrierobotern, October 2004.
- [BW04] L. Beyer und J. Wulfsberg. Practical Robot Calibration with ROSY. *Robotica*, 22(5):505–512, September 2004.
- [DH55] J. Denavit und R. S. Hartenberg. A Kinematic Notation for a Lower-Pair Mechanisms Based on Matrices. *ASME Journal of Applied Mechanics*, 22:215–221, 1955.
- [DSP93] M. R. Driels, W. Schwayze und S. Potter. Full-Pose Calibration of a Robot Manipulator Using a Coordinate-Measuring Machine. *Int. J. Adv. Manuf. Technol.*, 8:34–41, 1993.
- [Grä99] R.-G. Gräser. Ein Verfahren zur Kompensation temperaturinduzierter Verformungen an Industrierobotern, July 1999.
- [HM85] S. A. Hayati und M. Mirmirani. Improving the Absolute Positioning Accuracy of Robot Manipulators. *Journal of Robotic Systems*, 2(4):397–413, 1985.
- [HW96] J. M. Hollerbach und C. W. Wampler. The Calibration Index and Taxonomy for Robot Kinematic Calibration Methods. *Int. J. of Robotics Research*, 15(6):573–591, 1996.
- [MRD91] B. W. Mooring, Z. S. Roth und M. R. Driels. *Fundamentals of Manipulator Calibration*. Wiley Interscience, New York, 1991.
- [RKC07] K. Radkhah, D. Kulic und E. Croft. Dynamic Parameter Identification for the CRS A460 Robot Arm. In *Proc. of the Conf. on Intelligent Robots and Systems (IROS'07)*, San Diego, October 2007.
- [RMR87] Z. S. Roth, B. W. Mooring und B. Ravani. An Overview of Robot Calibration. *IEEE Journal of Robotics and Automation*, 3(5):377–385, October 1987.
- [SAG97] K. Schröer, S. L. Albright und M. Grethlein. Complete, Minimal and Model-Continuous Kinematic Models for Robot Calibration. *Robotics and Computer-Integrated Manufacturing*, 13(1):73–85, 1997.
- [Sch93] K. Schröer. *Identifikation von Kalibrationsparametern kinematischer Ketten*. Number 126 in Produktionstechnik-Berlin. Hanser Verlag, München and Wien, 1993.
- [WA92] C. Wampler und T. Arai. Calibration of Robots having Kinematic Closed Loops using Non-Linear Least-Squares Estimation. In H. J. Zimmermann, Hrsg., *Proc. IFTOMM Symposium*, Seiten 153–158, Nagoya, Japan, Sept. 24-26 1992.
- [Wie01] U. Wiest. *Kinematische Kalibrierung von Industrierobotern*. Shaker Verlag, Aachen, 2001.

# Context-Dependent and Self-Responsible Migration of Software Agents in Heterogeneous Environments

Dirk Bade

University of Hamburg, Department of Informatics  
Distributed Systems and Information Systems Group

bade@informatik.uni-hamburg.de

**Abstract:** Software agents are often employed in distributed environments to cope with the various dynamical dimensions of such systems. Furthermore, the paradigm of agent-oriented software engineering is of special interest in the course of ubiquitous computing and the emerging mobility of users. One reason for this are the constituting characteristics of agents, that perfectly suit the accompanying demands of these trends. Hence, this paper introduces a new kind of adaptive, mobile application supporting the "anywhere and anytime"-ability of mobile computing in general as well as the mobility of users and their adaptation to new computing environments. Therefore, a generic and extensible environment model, facing the heterogeneity of the infrastructure and entities within the environment, is presented in this paper. Further on, a mobility model for safe and efficient migration of agents is introduced. Using these models, agents may adaptively choose appropriate migration strategies at runtime by taking their current context into account.

## 1 Introduction

Since the early eighties agent-oriented software engineering gains more and more interest. Due to the offered higher level abstractions compared to e.g. the object-orientation, this paradigm is especially suited for the development of complex and distributed systems [LMSW05]. Another reason for this are the constituting characteristics of software agents like their ability to perceive the environment, to autonomously and proactively act in this environment and thereby to dynamically adapt to changes [WJ95, RN03]. Additionally, the notion of mobility is included in some definitions of agents, allowing them to migrate between different execution platforms and perform tasks in a suitable environment [FG96].

The ability to sense the environment combined with the ability to change the place of execution allows for dynamically choosing appropriate execution platforms, characterized e.g. by offered local resources. This feature is especially useful in the context of mobile computing, where agents are executed on resource constrained devices and hence may wish to temporarily migrate onto other platforms in order to process their tasks. This also allows agents to autonomously accompany their user by monitoring her behavior and migrating on a (mobile) device close to the user. This way, the user has access to the agent's data and services anytime and anywhere. Think of a mail-agent for example, retrieving

mails once a network resource is available and following it's user so that she can read and respond to mails no matter whether she is at her desktop computer at home or at work or with her mobile device on the way.

But having a personal agent following it's user to offer data and services is just one aspect. Since only a single instance of an application agent is needed, multiple installations, configurations and subsequent updates as well as manually synchronizing two or more instances can be avoided. This results in a higher convenience for the user as she can work with her application in a familiar setting and does not have to get used to new application settings on each device she is working with. Other examples include self-organizing service networks, where context-aware, mobile service agents are responsible for load-balancing and failure-recovery or applications for mobile ad-hoc networks (MANETs), in which unreliable and slow connections demand safe and efficient migration protocols that take the agent's current context into account.

Therefore, in this paper an environment model is presented in section 2, allowing agents to exchange information about entities in their surrounding. Furthermore, section 3 introduces a mobility model, which is used to specify the details of a migration process like e.g. the kinds of environmental events triggering the migration, the kind of migration (weak or strong), safety provisions (e.g. agent backups, error handling) and so on.

## **2 Environmental Awareness**

Sensing the environment plays a key role in an agent's internal reasoning process to choose an appropriate action towards reaching one of it's goals. Such an action is often directed to some kind of entity within the environment, e.g. another agent to coordinate with, a database to access or a service to engage. In order to perceive entities, an agent either needs some kind of sensor or it must exchange knowledge with other agents. The received percepts then need to be integrated into an internal model of the environment for further application-dependent processing.

This work therefore proposes a generic and flexible environment model, which is on the one hand responsible for gathering information about locally and remotely accessible entities and on the other hand for offering this information to the user and other agents in an easy, intuitive and yet expressive way. Furthermore, the model should not only allow for entity look-up or discovery, but should also support awareness [McG00], i.e. the continuous integration of state changes triggered by any environmental events. Due to the heterogeneity of the entities in the environment, it must allow for integrating information about any kind of entity (e.g. hardware entities, remote platforms, available services and other agents) and must be extensible to incorporate application-dependent information (e.g. documents, business objects, etc.). Finally, the model should make use of existing standards for exchanging and representing knowledge, but also provide simple proprietary mechanisms as a least common denominator.

In order to meet these requirements, the proposed model defines a set of generic interfaces for a) gathering information about local entities, b) exchanging information with other agents about remote entities and c) representing information. Local sensors are responsi-

ble for gathering information about e.g. the hardware, the operating system and execution environment as well as locally offered services and other local agents. Discovery protocols (e.g. Jini, Bluetooth SDP, JXTA, etc.) may be used to distribute information within the network and hence to collect information about remote entities. Different representation- and query languages as well as optional transformation services allow for expressing knowledge, ranging from simple string-based to expressive logic-based languages (e.g. RDF, OWL). Additional support for ontologies not only yields to a common conceptualization of entities, but also eases the transformation of representations into one another and hence supports the interoperability of different protocols and mechanisms.

In the context of this work a prototype for the *Jadex agent system* [BPL05] has been developed. This prototype has been implemented as a dedicated service agent (called *resource facilitator*, RF) responsible for managing and updating the environment model. Once started, this component may use different mechanisms to initially find other remote RFs and to subscribe for specific news channels (e.g. news about offered services, known contacts, resource load, etc.). In the following, information is exchanged using one or more standard discovery protocols, which are adaptively chosen, depending on the current context. This way, a virtual (multi-hop) network is spanned, in which environmental information is actively distributed among the RFs. Being aware of the environment is furthermore the basis for context-dependent, event-driven migration, as described in the following section.

### 3 Context-Dependent and Self-Responsible Migration

Considering the process of agent migration, there are three questions to answer: when, where and how to migrate an agent? In general, one might say, an agent needs to migrate once it's current platform lacks a required resource, which is available elsewhere. This answers the first two questions. But how to migrate strongly depends on the current conditions, because there are different kinds of migration techniques. One has to decide whether to migrate *weakly* (only transfer the object-state) or *strongly* (transfer the agent's object-state along with it's call stack and it's instruction pointer) and whether to push all or some of it's classes to the next destination or to pull them on demand. This decision is called a migration strategy [Bra03] and has a high impact on the efficiency of the migration process. For example, in a mobile ad-hoc network it might be better to push an agent, it's whole state and all of it's code, because it cannot be assured that the unreliable connection is still functional later on to reload some classes. On the other hand, with an expensive UMTS connection, one might want to push only the agent's object state and pull single classes once they are needed.

Answers to the questions raised above can often only be determined at runtime. And since an agent should be allowed to decide and act autonomously, user intervention should be avoided for convenience (surely not in all cases). Using the environment model introduced in the previous section an agent is able to perceive changes in the environment and may therefore decide at runtime when, where and how to migrate. But the process of migration is error-prone and hence reasonable precautions have to be taken, especially in case the user did not initiate the migration but the agent. As a consequence, the agent has to act



self-responsible, which means that the agent has to make sure, that the user's access to the agent is always granted. In unreliable networks (e.g. MANETs) the agent may for example create a backup of itself on the user's device, demand a transactional but less efficient migration and choose appropriate migration strategies in order not to get lost. In addition, sophisticated exception-handling for every single stage of the migration process needs to be applied in order to react to failures and to restart an agent once the migration failed. In order to meet these requirements, the proposed mobility model extends the *Kalong* model [Bra03], which focuses on the efficiency, in a way to also include safety aspects in a migration process. This way agents may autonomously find a trade-off between efficiency and safety of a migration process, depending on their current context.

## 4 Conclusion

This paper proposes a generic and flexible environment model as well as an adaptive mobility model. Although each of these models is useful in itself, the combination of these enables a new kind of adaptive mobile application. On the one hand, such an application may autonomously delegate subtasks to be executed on several nodes in a distributed system, thereby considering efficiency and safety aspects so that complete execution can be guaranteed. And on the other hand, this allows for an application to follow its user (or other mobile entities) in order to offer data and services anywhere and anytime <sup>1</sup>.

## References

- [Bad07] Dirk Bade. Kontextabhängige und eigenverantwortliche Migration von Software-Agenten in heterogenen Umgebungen. Master's thesis, Uni Hamburg - Dept. Informatik - AB VSIS, May 2007.
- [BPL05] Lars Braubach, Alexander Pokahr, and Winfried Lamersdorf. *Jadex: A BDI-Agent System Combining Middleware and Reasoning*, pages 143–167. Whitestein Series in Software Agent Technologies, Birkhäuser Verlag, 2005.
- [Bra03] Peter Braun. *The Migration Process of Mobile Agents-Implementation, Classification, and Optimization*. PhD thesis, Uni Jena, CS Dept., May 2003.
- [FG96] S. Franklin and A. Graesser. Is it an Agent, or just a Program? In *Intelligent Agents III. ATAL'96*, volume 1193. Springer-Verlag, 1996.
- [LMSW05] M. Luck, P. McBurney, O. Shehory, and S. Willmott. *Agent Technology: Computing as Interaction (A Roadmap for Agent Based Computing)*. AgentLink, 2005.
- [McG00] R.E. McGrath. Discovery and Its Discontents: Discovery Protocols for Ubiquitous Computing. Technical report, Uni. of Illinois, CS Dept., 2000.
- [RN03] Stuart Russell and Peter Norvig. *Artificial Intelligence: A Modern Approach*. Prentice-Hall, Englewood Cliffs, NJ, 2nd edition, 2003.
- [WJ95] Michael Wooldridge and Nicholas R. Jennings. Intelligent Agents: Theory and Practice. *Knowledge Engineering Review*, 10(2):115–152, 1995.

---

<sup>1</sup>Further information can be found in [Bad07]

# Clustering Techniques on Transformation Systems

Peer Bartels

De Montfort University  
Computing Sciences and Engineering  
peer.bartels@yahoo.de

Art der Arbeit: Abstract

Betreuer/in der Arbeit: Prof. Hussein Zedan, Prof. Karl Hayo Siemsen

**Abstract:** Legacy system migration is focused on the development of a target system which retains the functionality and, importantly, data of the original legacy system [1]. However, Legacy systems can be maintained and adapted easily to meet future business requirements. The idea is not new, as some approaches already do exist which allow the migration of such systems to new platforms and architectures which also provide tools and scientific methods to support such a strategy.

One approach is the "FermaT Transformation System" developed by Martin Ward at the University of Oxford, which has demonstrated that legacy systems can be transformed securely to a high level programming language such as C or Cobol [2]. The invented environment is based on a mathematical proven code transformations and is called "FermaT Transformation Engine". It uses an intermediate language known as "Wide Spectrum Language" or "WSL" [2] for the transforming steps. Although the invented "Wide Spectrum Language" has one lack, it does not contain cluster specific techniques and therefore it is not possible to parallelize sequential transformation tasks or to perform complex and time consuming transformations.

The objective aim of the proposed research is the development of well defined cluster technique which could remedy the situation along with a development of a systolic computing [3] infrastructure to perform parallel transformation sequences within a clustering system [4][5].

# 1 FermaT Cluster Architecture

To be able to support the clustering techniques within the FermaT transformation system, we have established a clustering architecture, which has similarities to the classical Beowulf architecture [6]. As in the Beowulf architecture, it also consists out of a headnode and the attached computing nodes. At the beginning of a parallel transformation task/process, the headnode has to know the status of the underlying architecture, to be able to successfully submit the job to the attached nodes. We call this first step “pre-processing”, by letting the headnode gather all the needed information about the underlying architecture such as a cluster or grid.

Within this approach, the underlying architecture can be a cluster or grid but the important information is, that it is a systolic-computing one. Within the processor architecture, a systolic computing architecture can be described as a specific arrangement of the involved data processing units within such a system. During a computation process, the processing units exchange data by communicating between each other and therefore the data or information is being triggered between the different nodes.

Communication and data exchange is being needed during a parallel transformation process since many transformations need to be applied during such a transformation process [7]. Therefore the data flows across the nodes and between their neighbours and can also flow in different directions. The name derives from analogy with the regular pumping of blood by the heart. The headnode is therefore in charge of distributing the processing data to the compute nodes of the systolic computation.

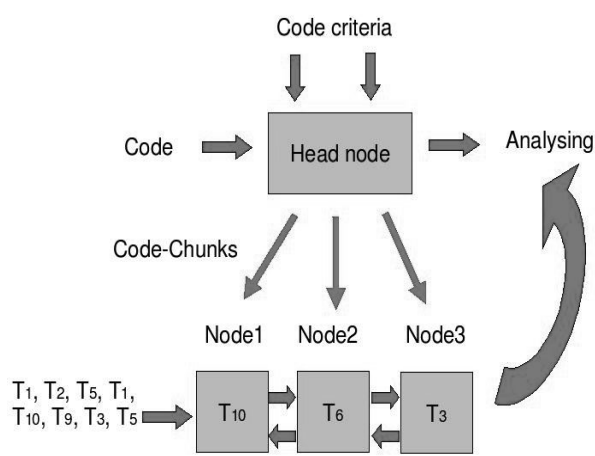


Abbildung 1: FermaT Cluster Architecture

## 2 Fermat Cluster Software Architecture

To be able to achieve such a special computing architecture we have developed a tool to support this parallelism for the Fermat transformation engine. The tool is known as the Fermat Cluster Environment (FCE) and is implemented the the platform independent language Java. As mentioned in the previous chapter, the tool does support the standard Beowulf cluster architecture and therefore the system can easily be expanded by adding more nodes in an ad-hoc manner to the system, including the gathering of information about the different processing nodes.

For such a system we have also established different models such as a one “Communication Model”, one “Termination Model”, one “Timing Model” and a “Monitor” which is monitoring the whole transformation process.

To be able to support other cluster or high performance architectures the software architecture is also based on the open source Torque and Maui job submission programming interfaces using the Intel C++ compiler suite.

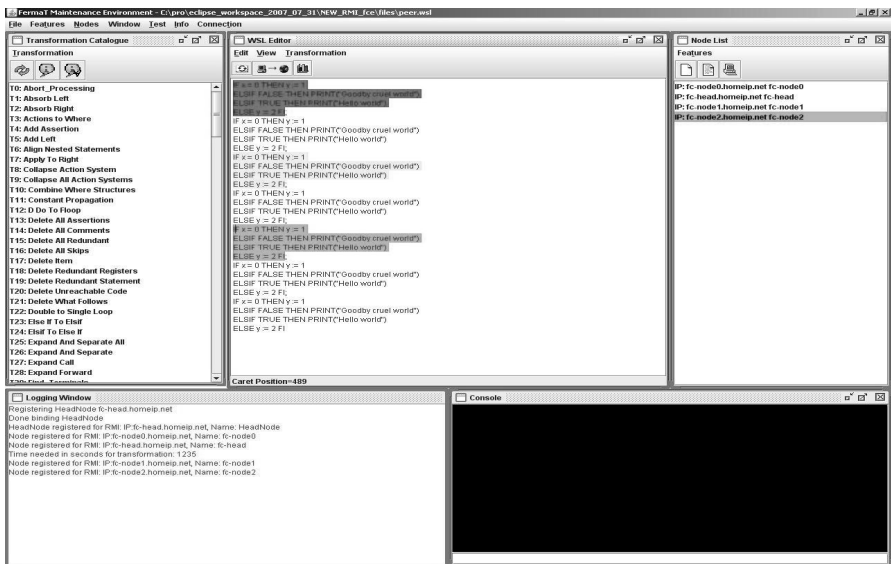


Abbildung 2: Fermat Cluster Environment

## Literaturverzeichnis

- [1] Martin P. Ward, Hussein Zedan, and T. Hardcastle. Legacy assembler reengineering and migration. In ICSM, pages 157–166. IEEE Computer Society, 2004.
- [2] Assembler to C migration using the fermaT transformation system. August 30 1999.
- [3] H.T. Kung. Why systolic architectures. IEEE Computer, pages 37–46, January 1982
- [4] Jeffrey Dean and Sanjay Ghemawat. Mapreduce: Simplified data processing on large clusters. In OSDI, pages 137–150, 2004.
- [5] R.J. Allan M. Antonioletti, E. Breitmoser. Software engineering and code development for hpc applications, 2000.
- [6] Beowulf. Beowulf, <http://www.beowulf.org/> 2007.
- [7] M. Ward, H. Zedan, M. Ladkau und S. Natelberg. Conditioned Semantic Slicing for Abstraction; Industrial Experiment. Software - Practice and Experience, Accepted for publication, 2008.

# Entwicklung einer OSEK/VDX-kompatiblen Systemschnittstelle für Linux

Johannes Bauer

JohannesBauer@gmx.de

**Abstract:** Dieses Paper fasst die Studienarbeit von Johannes Bauer zusammen, die sich mit der Entwicklung einer Systemschnittstelle für das im KFZ-Bereich häufig anzutreffende Echtzeitbetriebssystem OSEK/VDX beschäftigt.

## 1 Einführung

Für eingebettete Systeme, die Echtzeitbedingungen erfüllen müssen, eignen sich Betriebssysteme mit deterministischem, statischen Schedulingverhalten ausgezeichnet. Ein solcher Betriebssystemstandard ist OSEK/VDX [OSE05]. Da dieses vornehmlich auf tief eingebetteten Systemen zum Einsatz kommt, sind die Möglichkeiten des Testens funktionaler Komponenten und des Debuggings eher beschränkt, wenn überhaupt vorhanden. Eine Simulation des Laufzeitverhaltens auf einer mächtigeren Plattform ist hier also wünschenswert. Dies ist genau das Ziel der Studienarbeit. Durch die darin vorgestellten Werkzeuge wird es möglich, Applikationen, die gegen das OSEK-Betriebssysteminterface programmiert sind, auf einem herkömmlichen x86-32 oder x86-64 auszuführen, sofern auf diesem ein UNIX-artiges Betriebssystem läuft. Es gibt bereits zwei ähnliche Projekte, *openOSEK* [ope07] und *Trampoline* [BBFT06]. *openOSEK* befindet sich momentan noch im Entwicklungsstadium und ist nicht lauffähig, *Trampoline* hingegen hat bereits eine solide Codebasis. *Trampoline* verfolgt den Ansatz, durch Virtualisierung der Zielhardware möglichst portablen OSEK-Code zu erzeugen. In der vorliegenden Arbeit ist der Schwerpunkt ein anderer: Hier wird versucht, eine möglichst hardwarenahe Abbildung auf das darunterliegende UNIX-artige Betriebssystem zu erhalten. Dies geschieht zum Einen, um es zu ermöglichen, das OSEK-Betriebssystem auch direkt auf eingebettete Hardware zu übersetzen, zum Anderen um möglichst „echt“ zu Lehrzwecken zu zeigen, wie ein OSEK-Betriebssystem in der Praxis gebaut ist.

## 2 Zielstellung und Aspekte der UNIX-Implementierung

Die Studienarbeit beschäftigt sich mit der Erstellung eines Codegenerators, genannt *jo-sek*, welcher für eine gegebene OSEK-Betriebssystemkonfiguration C-Code erzeugt. In

der Konfigurationsdatei, dem sogenannten *OIL*<sup>1</sup> [OSE05, OSE04] wird die Konstitution des zu erzeugenden Betriebssystems beschrieben. Dies schließt also unter anderem die Anzahl der Tasks und Ressourcen sowie deren Beziehungen miteinander ein. *josek* parst die OIL-Datei und erzeugt dann mit Hilfe verschiedener situationsabhängiger Templates dynamisch den Betriebssystemcode. Hier kommt der Vorteil eines Codegenerators im Gegensatz zu vorgefertigtem Code zum Tragen: Spezielle Optimierungen können vom Codegenerator berücksichtigt werden und dadurch effizienter Code erzeugt werden. Dies ist zum Beispiel der Fall, wenn zur Generierungszeit ermittelt werden kann, dass zwei Tasks zur Laufzeit nie gleichzeitig ausgeführt werden. Zwei unterschiedliche Stackbereiche für die beiden Tasks sind dann nicht mehr nötig; sie können zu einem Stackbereich verschmolzen werden und so RAM sparen. Auch sind diverse Möglichkeiten vorgesehen, den Code durch Debuggingfeatures zu ergänzen. So kann, sofern das OSEK-OS auf einem Linux ausgeführt wird, automatisch Code zum Speicherschutz der Stackbereiche erzeugt werden. Sollte ein Task zur Laufzeit seinen Stackbereich überschreiten, führt dies zu einem synchronen Programmabbruch, der leicht zu debuggen ist. Neben den mannigfaltigen Optimierungs- und Spezialisierungsmöglichkeiten stellt sich noch ein dritter Vorteil von generiertem Code automatisch mit ein: das Endprodukt ist gut lesbar und kann so zu Lehrzwecken eingesetzt werden. Der generierte Code kann dann übersetzt und gegen die eigentliche Applikation gelinkt werden. Obwohl auf den Plattformen, auf denen UNIX-artige Betriebssysteme lauffähig sind, Ressourcen wie Plattenspeicher und RAM-Größe im Überfluss vorhanden sind, verfolgt *josek* dennoch eine duale Zielstellung: Praktische Anwendungen, auf der OSEK-konforme Betriebssysteme zum Einsatz kommen, sind in ihren Ressourcen möglicherweise stark limitiert. Deshalb wurde *josek* derartig konzipiert, dass es praktisch möglich ist, auch eine Abbildung auf direkte Hardware durchzuführen – hierbei wurde die 8-Bit Mikrocontrollerfamilie AVR anvisiert.

### 3 Implementierung

Soll nun aber – wie es der Hauptzweck der Arbeit war – eine Betriebssysteminstanz auf einem UNIX-artigen System simuliert werden, sind einige Abbildungen nötig, da direkter Hardwarezugriff weder einfach möglich noch wünschenswert ist. Deshalb wird eine komplette Betriebssystem-Instanz auf einen UNIX-Prozess abgebildet. Der Stack, den die verschiedenen Tasks benutzen, wird in diesem UNIX-Prozess in den Heap gelegt. Dies wird dadurch möglich, dass zur Generierungszeit des Betriebssystems die Größen für die einzelnen Stacks bereits feststehen müssen statt dynamisch zur Laufzeit zu wachsen. Letztendlich werden die Tasks selbst auf User-Level Threads abgebildet.

Um einem UNIX-Prozess eine asynchrone Programmunterbrechung zu senden, kann *josek* bei Bedarf eine ISR mit einem POSIX-Signal verknüpfen. Ein Signalhandler wird automatisch installiert und springt die ISR-Routine an, sobald der Prozess das Signal empfängt. Hierdurch wird die Interruptsemantik erfüllt und nachgebildet, allerdings mit einem Haken: Die ISRs werden zeitverzögert zugestellt, weil Signalhandler in Linux nur dann angesprungen werden, wenn ein Rescheduling-Point erreicht wird. Das heißt der ISR wird

---

<sup>1</sup>OSEK Implementation Language

erst dann ausgeführt, wenn der OSEK-OS-Prozess durch Ablaufen seiner Zeitscheibe verdrängt wird oder selbst einen Systemaufruf durchführt. Dies stellt allerdings kein Problem dar, da die OSEK-Spezifikation bei der Definition der Interrupts nicht wählerisch ist. Insbesondere wird kein maximaler Jitter gefordert.

Die Interaktion eines UNIX-Prozesses mit der Außenwelt, also beispielsweise einer Device-Node die einen CAN-Bus-Controller repräsentiert, ist nicht ganz einfach: Sollte der Prozess nämlich einen blockierenden `read` oder `write` Systemaufruf auf diese Device-Node ausführen, blockiert nicht nur der momentan laufende Task, sondern das ganze Betriebssystem. Dies kann natürlich durch nicht-blockierende Systemaufrufe und Polling unterbunden werden, ist aber unschön. Auf realer Hardware sähe die Realisierung schließlich so aus: die Peripherie signalisiert der CPU durch einen Interrupt, dass auf dem Gerät Daten anliegen oder dass das Gerät zum Schreiben bereit ist. *josek* unterstützt auch diese Art von IO-Interrupts: Hierzu kann der laufende Task eine Device-Node öffnen und *josek* durch eine Systemschnittstellenerweiterung den erhaltenen Filedeskriptor mitteilen. Dies geschieht mit der zusätzlichen Anforderung, dass Interrupts ausgelöst werden sollen, sobald der übergebene Deskriptor zum Lesen oder Schreiben bereit ist. Erhält das OSEK-Betriebssystem eine solche Anforderung, wird ein zweiter asynchroner Aktivitätsträger durch Zuhilfenahme der *pthread*s-Bibliothek erzeugt. Dieser neu entstandene Thread führt auf den überwachten Filedeskriptoren den POSIX-Systemaufruf `poll` aus, der solange blockiert, bis eine der Wartebedingungen erfüllt ist. Sobald dies geschieht, sendet der `poll`-Thread dem gesamten Prozess das Signal `SIGIO`. Im Hauptkontrollfluss ist vorher ein Signalhandler installiert worden, der dann die entsprechenden IO-Interrupts delegiert.

Werden von der Applikation, die auf dem OSEK-OS laufen soll, Alarms benutzt, so werden diese alle auf den Systemtimer, der jedem UNIX-Prozess zur Verfügung steht, abgebildet. Mittels `setitimer` wird hier also ein Systemalarm definiert, der dann wiederum in der Callback-Funktion – dem Signalhandler für `SIGALRM` – an die OSEK-Alarms entsprechend dispatcht.

## 4 Zusammenfassung

Insgesamt stellt der in der Arbeit entwickelte Codegenerator *josek* eine nützliche Möglichkeit dar, Programme, die gegen die OSEK-Schnittstelle programmiert sind, auf einem UNIX-System auszuführen und funktionale Komponenten so zu testen. Dies ist insbesondere deshalb sinnvoll, weil höhere Betriebssystemschichten, wie beispielsweise das an der Universität Erlangen entwickelte KESO [WSSP07], die relativ schlanke OSEK-Schnittstelle als Hardwareabstaktionsschicht voraussetzen. Zusätzlich zeigt die Arbeit in ihrem dualen Charakter, dass es möglich ist einen Codegenerator derart zu entwickeln, dass mit geringen Anpassungen auch Programmcode erzeugt werden kann, der direkt auf Hardware aufsetzt, welche stark ressourcenbeschränkt ist.



## Literatur

- [BBFT06] Jean-Luc Béchenne, Mikaël Briday, Sébastien Faucou und Yvon Trinquet. Trampoline – An OpenSource Implementation of the OSEK/VDX RTOS Specification. 2006. <http://trampoline.rts-software.org/IMG/pdf/trampoline.pdf>.
- [ope07] openOSEK.org. *The most common questions answered*. March 2007. <http://www.openosek.org/tikiwiki/tiki-index.php>.
- [OSE04] The OSEK Group. *OSEK/VDK Binding Specification 1.4.2*. July 2004. <http://portal.osek-vdx.org/files/pdf/specs/binding142.pdf>.
- [OSE05] The OSEK Group. *OSEK/VDK Operating System Specification 2.2.3*. February 2005. <http://portal.osek-vdx.org/files/pdf/specs/os223.pdf>.
- [WSSP07] Christian Wawersich, Michael Stalkerich und Wolfgang Schröder-Preikschat. An OSEK/VDX-based Multi-JVM for Automotive Appliances. In *Embedded System Design: Topics, Techniques and Trends*, IFIP International Federation for Information Processing, Seiten 85–96, Boston, 2007.

# Praktikum: Entwurf und Implementierung eines Softwareprojekts in einer funktionalen Sprache

Stephan Friedrichs, Henning Günther,  
Oliver Mielentz, Martin Wegner

**Abstract:** Im folgenden werden der Entwurf und die Implementierung einer komplexen Software zur Realisierung eines Chat-Protokolls in Ad-hoc-Netzwerken in der funktionalen Programmiersprache Haskell beschrieben. Es wird gezeigt, wie klassische objektorientierte Entwurfsmethoden auf eine rein funktionale Sprache angewendet werden können. Es werden zusätzliche, besonders hilfreiche Eigenschaften (wie z. B. Seiteneffektfreiheit und strenge Typisierung) dieser Sprache betrachtet.

## 1 Problemstellung

Das im Folgenden beschriebene Projekt ist im Rahmen des Software-Entwicklungs-Praktikums entstanden, dessen Aufgabe es ist, in kleinen Teams einen objektorientierten Entwurf einer komplexen Software nach dem Wasserfallmodell durchzuführen und anschließend zu implementieren.

Unsere konkrete Aufgabenstellung war ein Ad-hoc Chatsystem für drahtlose Netze (z. B. WLAN im Ad-hoc Modus über UDP). Die *Knoten* in einem solchen Netz sind mobile Geräte (Notebook, PDA, ...), an denen mindestens ein Nutzer chattet. Es gibt keinen zentralen Server, die einzelnen Knoten müssen selbstständig kooperieren, um die *Infrastruktur* des aktuellen Netzes (die sich laufend ändert) zu erkennen und zu nutzen. Beispielsweise muss eine flexible Routingstrategie entworfen werden, Teilnetze können hinzukommen (*Netmerge*) oder abgetrennt werden (*Netsplit*); dabei müssen z. B. veränderte Routen und unzustellbare Nachrichten als solche erkannt und speziell behandelt werden. Das Protokoll war in [SRLW07] vorgegeben, es gab aber keine Referenzimplementierung.

## 2 Entwurf

Die ersten beiden Entwurfsschritte, Pflichtenheft [FGMWc] und Grobentwurf [FGMWb] wurden unabhängig von der Programmiersprache entworfen. Erst mit dem Feinentwurf [FGMWa] spielten Aspekte der funktionalen Programmierung eine Rolle.

Der Grobentwurf [FGMWb] sieht im Wesentlichen folgende Komponenten vor:

**Parser:** Parst das XML-basierte Protokoll in interne Repräsentation

**Data:** Enthält dedizierte Datenstrukturen für die Infrastruktur des Netzes (Routinginformationen), Kanal- (ähnlich IRC) und Nutzerinformationen, Zertifikate der Nutzer für verschlüsselte Kanäle und verschiedene für das Protokoll benötigte Zwischenspeicher

**Distributor:** Realisiert die Protokolllogik, d. h. es werden Informationen aus eingehenden Nachrichten bewertet, ggf. abgespeichert und Nachrichten versendet bzw. weitergeleitet

**GUI:** Nutzeroberfläche, kommuniziert über das *ServerInterface*

**ServerInterface:** Kommunikation zwischen *GUI* und *Distributor*

Dabei ist hervorzuheben, dass die *GUI*-Komponente vom Rest der Software getrennt wurde und ausschließlich über die wohldefinierte Schnittstelle im *ServerInterface* mit dem *Distributor* verbunden ist. Das heißt, dass sie keine Protokolllogik implementiert, sondern ein reines Anzeigeelement ist. Darüber hinaus ist sie durch eine alternative Implementierung (z. B. über ein lokales Bussystem oder auf Konsole statt per GTK) austauschbar.

Es wurde außerdem beachtet, dass in keiner Komponente (das betrifft vor allem die *Data*-Komponente) Informationen redundant gespeichert werden, häufige Zugriffe aber trotzdem effizient erfolgen.

Ab dem Feinentwurf [FGMWa] müssen die Eigenschaften der geplanten Sprache (in unserem Fall die rein funktionale Sprache Haskell) berücksichtigt werden. Im Fall einer funktionalen Sprache ist diese Unterscheidung wichtiger, da nicht alle objektorientierten Mechanismen zur Verfügung stehen, dafür jedoch andere Möglichkeiten hinzukommen. Die Unterschiede betreffen vor allem das Typsystem, aber auch den Umgang mit Seiteneffekten: Haskell ist komplett frei von Seiteneffekten.

Seiteneffektfreiheit bedeutet, dass eine Funktion auf nichts anderes als ihren Rückgabewert Einfluss hat. So wäre beispielsweise das funktionale Äquivalent zu der C++-Funktion `void Set::insert(T t)` (Seiteneffekt: Das Set wird verändert) eine Funktion, die ein neues Set zurückgibt: `insert :: t → Set → Set1` (kein Seiteneffekt, das neue Set wird zurückgegeben).

Funktionen ohne Seiteneffekte sehen auf den ersten Blick umständlich aus, haben jedoch einen klaren Vorteil: Da sämtliche Änderungen im Rückgabewert enthalten sind, gibt es keine schwer zu überblickenden Wechselwirkungen zwischen Komponenten; alles ist isoliert testbar, Fehler sind reproduzierbar und können leicht eingegrenzt werden.

Vielen Komponenten konnten elegant ohne Seiteneffekte entworfen werden, dazu gehören *Parser*, *Data* und *ServerInterface*.

Die Funktionen des *Distributors*, die z. B. eingehende Nachrichten bearbeiten, ändern fast alle seinen Zustand (Instanzen der Datenstrukturen aus *Data*, die Informationen über das Netzwerk und dessen Nutzer enthalten). Dies kann zwar auf herkömmliche Weise ohne Seiteneffekte dargestellt werden, ist jedoch aufgrund der hohen Komplexität schwer zu handhaben. Deswegen wurde ein sog. State-Monad (vgl. [GHC]) verwendet. Monaden<sup>2</sup>

---

<sup>1</sup>angelehnt an Haskell-Syntax

<sup>2</sup>in Haskell wird damit beispielsweise I/O realisiert, vgl. IO Monaden [GHC]

sind algebraische Strukturen, die in der Lage sind, die Seiteneffekte von Operationen auf den gekapselten Zustand des Monads zu beschränken.

Einige Operationen sind jedoch nicht seiteneffektfrei darstellbar. Dazu gehören vor allem I/O (das betrifft in diesem Projekt vor allem die Interaktion mit dem Netzwerk), explizites Multithreading und die Nutzeroberfläche (in unserem Fall GTK). Dies ist ein generelles Problem funktionaler Sprachen. Haskell erlaubt eine strikte Trennung des seiteneffekt-freien Teils (der *funktionale Kern* der Anwendung) von dem Teil, der die anfallende I/O letztlich ausführt. Wir konnten den I/O-Teil sehr klein halten, so dass die Vorteile rein funktionaler Programmierung für den Großteil des Projekts erhalten blieben.

### 3 Fazit

Wenn man ab dem Feinentwurf die Sprache beachtet (was auch auf objektorientierte Sprachen zutrifft), ist es gut möglich, einen objektorientierten Entwurf auf eine rein funktionale Sprache anzuwenden. Dabei gehen die Vorteile der funktionalen Programmierung nicht verloren.

Es müssen im Wesentlichen zwei Dinge beachtet werden: Die Trennung von seiteneffekt-freien und -behafteten Operationen und die Unterschiede im Typsystem (bestimmte Vererbungshierarchien sind nicht direkt umsetzbar, andere jedoch sehr gut für algebraische Typen geeignet).

Von Nachteil ist, dass die aus der Objektorientierung gewohnten Denkmuster i. d. R. nicht direkt angewendet werden können: Vieles erfordert eine andere Denkweise (Was ist ohne Seiteneffekte zu bewältigen? Wo sollten diese durch ein Monad gekapselt werden?).

Die Vorteile überwiegen dies jedoch deutlich: Seiteneffektfreie Komponenten verhalten sich im Gesamtsystem genau so wie bei isolierten Tests. Die Implementierung einer solchen Komponente wirkt ausschließlich über ihre Schnittstelle auf das Gesamtsystem. Somit ist (ggf. unbeabsichtigte) „versteckte“ Kommunikation (z. B. über globale Variablen oder Singletons) kategorisch ausgeschlossen. Das erleichtert Tests und Debuggen un-ge- mein, beides kann viel strukturierter angegangen werden.

Weiterhin wird durch Haskell's Typsystem garantiert, dass zur Laufzeit keine Typfehler auftreten. Dadurch wird eine typische Fehlerquelle von vornherein ausgeschlossen.

Durch die hohe Abstraktion und Ausdruckskraft von Haskell reduziert sich die Menge an Quellcode signifikant (das Projekt wurde parallel von drei anderen Gruppen in Java und C++ durchgeführt, die Haskell-Implementierung hat etwa 70% weniger Code bei größerem Funktionsumfang, s. u.). Dies erhöht die Lesbarkeit und Wartbarkeit des Codes enorm.

Abschließend bleibt zu erwähnen, dass unsere Implementierung das Protokoll erfolgreich umsetzt: Die Netzstruktur wird korrekt erkannt, die Software kann mit Netsplit- und Net-merge-Vorgängen umgehen und Nachrichten werden korrekt geroutet; das Chatten in öffentlichen, anonymen und verschlüsselten Kanälen ist möglich.

Über die Mindestanforderungen hinaus können mehrere Benutzer gleichzeitig auf einem Knoten chatten, was das Protokoll zwar erlaubt, die Programmlogik jedoch deutlich kompliziert. Die grafische Nutzeroberfläche verbirgt die komplizierten Mechanismen des Ad-hoc Netzes durch die saubere Trennung der Komponente *GUI* vom Rest der Anwendung. Durch klares, intuitives Design und bewusst gewählter Ähnlichkeit zu bekannten Chatclients ist sie leicht zu bedienen.

## Literatur

- [FGMWa] Stephan Friedrichs, Henning Günther, Oliver Mielentz und Martin Wegner. <http://sep07.mroot.net/documents/Feinentwurf-2.pdf>. Feinentwurf.
- [FGMWb] Stephan Friedrichs, Henning Günther, Oliver Mielentz und Martin Wegner. <http://sep07.mroot.net/documents/Grobentwurf-5.pdf>. Grobentwurf.
- [FGMWc] Stephan Friedrichs, Henning Günther, Oliver Mielentz und Martin Wegner. <http://sep07.mroot.net/documents/Pflichtenheft-5.pdf>. Pflichtenheft.
- [GHC] <http://www.haskell.org/ghc/docs/latest/html/libraries/index.html>. Dokumentation der GHC Module.
- [SRLW07] F. Strauss, S. Ransom, S. Lahde und O. Wellnitz. P2P Chat - A Peer-to-Peer Chat Protocol, <http://sep07.mroot.net/documentation/draft-strauss-p2p-chat-09.txt>. Bericht 9, TU Braunschweig, April 2007.

# **Werkzeuggestützte Anforderungsanalyse für die benutzerzentrierte Entwicklung mobiler Unternehmenssoftware**

Roland Ellerweg

FH Aachen  
Elektrotechnik und Informationstechnik  
Ellerweg@FH-Aachen.de

Art der Arbeit: Diplomarbeit

Betreuer/in der Arbeit: Prof. Dr. –Ing Thomas Ritz

**Abstract:** Für die Entwicklung von mobiler Unternehmenssoftware eignet sich ein benutzerzentriertes Vorgehen. Gerade in der Anforderungsanalyse müssen hierbei ausführliche Informationen über den User sowie den Anwendungskontext erhoben werden. Der vorliegende Artikel stellt eine Erweiterung der UML vor, die eine detaillierte Beschreibung von Anwendungskontext und User erlaubt.

## **1 Einleitung**

Außendienstmitarbeiter profitieren von mobiler Unternehmenssoftware in zweierlei Weise. Zum einen ermöglicht die Mobile Lösung den ständigen Informationsaustausch mit den betrieblichen Informationsstrukturen, zum anderen wird die Kundenbetreuung verbessert. Die Entwicklung solcher unterstützenden Systeme gestaltet sich jedoch oft schwierig. Ursachen hierfür finden sich bereits in der Anforderungsanalyse, wo kritische Daten über Infrastrukturmerkmale oder Anwender kaum erfasst werden.

Dieser Artikel führt zunächst in ein benutzerzentriertes Entwicklungsmodell ein. Anschließend wird eine im Rahmen der Arbeit entwickelte Erweiterung der UML vorgestellt, die das Modell in der Anforderungsanalyse unterstützt.

## **2 Die benutzerzentrierte Entwicklung vom mobiler Unternehmenssoftware**

Um nachfolgende Sachverhalte zu verstehen ist es wichtig, die Definition von mobiler Unternehmenssoftware zu betrachten [Ri07]:

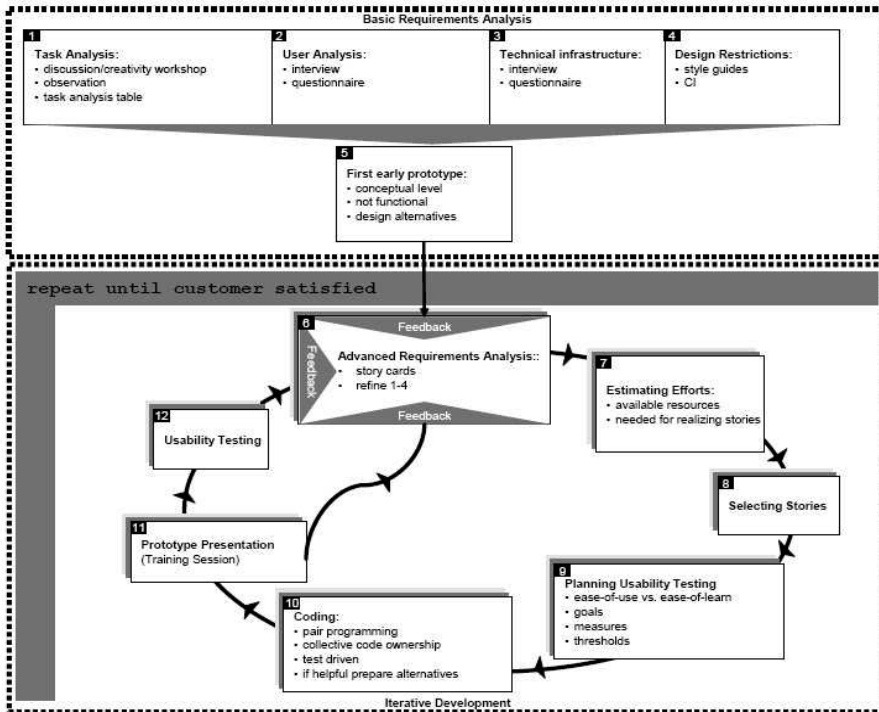
Mobile Unternehmenssoftware ist Software im mobilen Einsatz auf:

- Adäquaten mobilen Endgeräten: Das mobile Gerät muss für den Prozess geeignet sein.
- Mit angepasster Funktionalität: Es werden nur tatsächlich benötigte Funktionen

bereitgestellt.

- Basierend auf Daten adäquater Aktualität: Netzwerkverfügbarkeiten müssen berücksichtigt werden um Rückschlüsse auf eventuelle Synchronisationsmechanismen zu ziehen.

Basierend aus der Definition und den daraus resultierenden Anforderungen an mobile Unternehmenssoftware, entstand das folgende Entwicklungsmodell (siehe Abbildung 1).



Eine Nutzung von UML innerhalb der Basic Requirements Analysis ist nur bedingt möglich. Mit Anwendungsfalldiagramm und Aktivitätsdiagramm können zwar Prozesse abgebildet werden, Verrichtungsorte (Lokation), Anwendercharakteristika sowie Informationen zu der technischen Infrastruktur bleiben jedoch unberücksichtigt.

### 3 Integrierte Anforderungsanalyse für mobile Systeme

Vorangegangene Ausführungen haben grob in die Basic Requirements Analysis eingeführt. Im Folgenden werden die im Rahmen der Arbeit vorgenommenen Erweiterungen der UML vorgestellt.

Das Aktivitätsdiagramm unterteilt einen Anwendungsfall in mehrere Prozessschritte. Um an einem Prozessschritt Informationen zu dem Verrichtungsort zu speichern, wurde zunächst M-UML[Bh00] (M: Mobil) im Rahmen der Arbeit integriert. Dieses erlaubt eine differenzierte Betrachtung durch 3 Stereotype:

- `<<fix>>` (Fixe Lokation): Nur einmal auf der Welt existierend (etwa Längen, Breitengrade)
- `<<mobile>>` (Mobiles Objekt): Wird durch einen Symbolischen Ort (etwa „Kunde“) definiert. Dieser kann mehrfach an unterschiedlichen Weltkoordinaten auftreten
- `<<mobile2>>` (Mobiles Objekt an mobiler Lokation): Mobiles Objekt an einem wechselnden Ort (etwa Auto)

Um weitere Informationen zu der Lokation zu speichern, wurde ein neues Modellelement basierend auf dem von Kuhn vorgeschlagenen Mobile Process Landscapes [KG01] in das Aktivitätsdiagramm integriert. An jedem Landscape werden Umgebungseinflüsse (etwa Licht, Lautstärke) und die gewünschte Hardwareausstattung hinterlegt. Zur Beurteilung der Umgebungseinflüsse können verschiedene DIN Normen zugezogen werden (etwa DIN EN 12464-1, 12464-2, 33404-3).

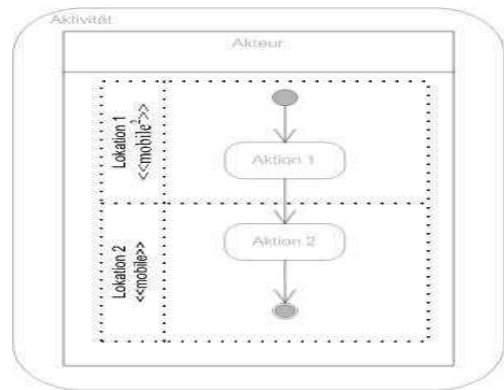


Abbildung 2: Lokation

Zur Aufstellung einer Anwendercharakteristik wurde das Akteurelement aus dem Use Case Diagram um Methoden des „Cross-Context Reasoning“ von Berkovsky [Be06] erweitert. Dieses setzt die Leistungsfähigkeit des Akteurs und die Komplexität des Anwendungsfalls in Relation. Es entstehen zwei Perspektiven:

1. **Akteur Perspektive:** Die Akteur Perspektive beschreibt den Akteur und seine Fähigkeiten. Bewertet werden Kognitive, Biographische und Körperliche Eignung sowie Verantwortlichkeiten des Akteurs. Zur Ermittlung der Leistungsfähigkeit eignet sich das Genfer Schema [BR00] oder auch die Arbeitbewertung nach REFA [Re91].
2. **Anwendungsfall Perspektive:** Die Anwendungsfall Perspektive beschreibt den Anwendungsfall und seine Anforderungen. Zusätzlich werden die von Berkovsky [Be06] vorgeschlagenen Arbeitsbedingungen, etwa Lichtverhältnisse, Witterungseinflüsse oder Lautstärke gespeichert, um auf weitere Störfaktoren einzugehen.

## 4 Fazit / Ausblick

Das beschriebene Vorgehen scheint ein viel versprechender Ansatz bei der Anforderungsanalyse von mobiler Unternehmenssoftware zu sein. Eine Prototypische Softwareunterstützung wurde im Rahmen der Arbeit bereits implementiert. Derzeit wird die Speicherung von Zugriffshäufigkeiten auf Audio/Video geprüft.



## Literaturverzeichnis

- [Be06] Berkovsky, S. et.al.: Predicting User Experiences through Cross-Context Reasoning. In (Eckert, C.; Polle, T.; Klein, H.J., Hrsg.): Hildesheimer Informatik Berichte. Universität Hildesheim, 2006; S. 27-31
- [BR00] Vom Brocke, J.; Robker, B.: Betriebswirtschaftliche Formeln, Schemata und Tabellen. Vahlen Verlag, Deutschland, Januar 2000.
- [De02] DEGAS: Design Environments for Global Applications. A FET European research project, 2002. <http://www.omnys.it/degas>
- [IS00] Baumeister, Hubert: AGILE: Architectures for Mobility. Information Society, LMU, München, 29. Apr. 2005. <http://www.pst.ifi.lmu.de/projekte/agile/>
- [Ja01] Jameson, A.: Modelling both the Context and the User. In: Personal and Ubiquitous Computing, Vol.5, No.1. Springer, London, Februar 2001. S. 29-33
- [KG01] Köhler, A.; Gruhn, V. (2004): Mobile Process Landscaping am Beispiel von Vertriebsprozessen in der Assekuranz. Key Pousttchi, Klaus Turowski (Ed.): Mobile Economy - Transaktionen, Prozesse, Anwendungen und Dienste. Proceedings zum 4. Workshop Mobile Commerce, Augsburg, Deutschland, 03.02.2004 - 03.02.2004, S. 12-24. Buchreihe LNI, GI-Lecture Notes in Informatics.
- [Ma99] Mayhew, D. J.: The Usability Engineering Lifecycle: A Practitioners's Handbook for User Interface Design. Academic Press, USA. 1999.
- [Ne02] Neumann, H.A.: Analyse und Entwurf von Softwaresystemen mit der UML. Hanser Verlag, München, 2002.
- [Ni93] Nielsen, J.: Usability Engineering. Academic Press, Boston, 1993.
- [Re91] REFA: Anforderungsermittlung (Arbeitsbewertung). Carl Hanser Verlag, München, 1991.
- [Ri03] Ritz, T.: Mobile CRM Systeme. In: ZWF-Zeitschrift für wirtschaftlichen Fabrikbetrieb. Hanser Verlag, 12/2003. S. 699-702
- [Ri07] Ritz, T.: Die benutzerzentrierte Entwicklung mobiler Unternehmenssoftware. In (König-Ries, B. et.al., Hrsg.): GI-Edition – Lecture Notes in Informatics (LNI). Bonner Köllen Verlag, Aachen, 2007. S. 23-35
- [SP05] Shneiderman, B.; Plaisant, C.: Designing the User Interface: Strategies for Effective Human-Computer Interaction. Pearson, Boston, 2005.
- [SR03] Stender, M.; Ritz, T.: Überbetriebliche Ad-hoc-Anwendungsintegration im Mobile Business. In: (Uhr, W.; Esswein, W.; Schopp, E., Hrsg.): Wirtschaftsinformatik 2003/Band 1. Tagungsband zur 6. Internationalen Tagung Wirtschaftsinformatik, Dresden, 2003.
- [We01] Wells, D.: Extreme Programming: A Gentle Introduction. 2006. <http://www.extremeprogramming.org>

# A Revival for Legacy Systems - Successful Reengineering of Software Systems

Matthias Ladkau

De Montfort University  
Computing Sciences and Engineering  
matthias@ladkau.de

Art der Arbeit: Abstract

Betreuer/in der Arbeit: Prof. Hussein Zedan, Prof. Karl Hayo Siemsen

**Abstract:** This paper gives a brief introduction into the FermaT Transformation Theory as well as current research on a Wide Spectrum Type Theory which is used within the FermaT technology to formulate strong statements about the reliability of program translations. The FermaT Transformation Approach uses formal proven program transformations to reduce the complexity of a given program while preserving or refining its semantics. These transformations form a transformation system and are applied on program code of legacy systems, currently with emphasis on assembler systems, to restructure and simplify as well as to extract higher-level representations (specifications). The discussed research is now also able to enhance legacy assembler code with a type system which is able to address and solve several critical translation problems. The FermaT technology together with the Wide Spectrum Type Theory makes it possible to prove that a reengineered legacy system has exact the same functionality and output than the original system provided that it gets the same input of data.

## 1 FermaT Transformation Technology

The roots for the FermaT Transformation Theory go back to 1989 [War89]. In the early 1990's a prototype transformation system called "Maintainer's Assistant" [WCM89] was developed at the University of Durham (UK). Although it already included a large number of transformations, it was very much an "academic prototype" whose aim was to test the ideas rather than to be a practical tool. For the latest version (called FermaT) the underlying runtime language was changed from LISP to Scheme. All remaining parts of the system were reimplemented in METAWSL and a METAWSL to Scheme translator was introduced (also in METAWSL). To enhance the usability of the FermaT tool a new graphical interface called FermaT Maintenance Environment (FME) has been developed in 2006. The FME is entirely written in the object-oriented language Java. It is used as an interface to access the facilities of the transformation system and the FermaT

Transformation Theory. All ongoing research in the area of the FermaT technology, such as the Wide Spectrum Type Theory, is utilising the FME as foundation.

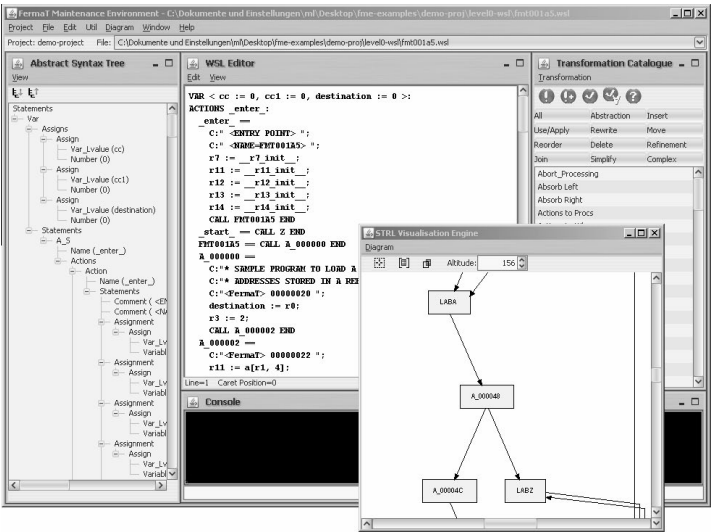


Abbildung 1: The FermaT Maintenance Environment

the core of the FermaT Transformation System is the Wide Spectrum Language (WSL) which utilises, among other things, Morgan’s specification statement [Mor88] and Dijkstra’s guarded commands [Dij76]. The intention is to form a language which acts as an intermediate language when processing a legacy system. WSL was designed for reengineering tasks and covers:

- Simple, clear and unambiguous syntax with formally specified semantics;
- A wide range of transformations with simple, mechanically checkable correctness conditions;
- The ability to express low-level programs and high-level abstract specifications.

The heart of WSL is a very small and mathematically tractable kernel language which includes already all necessary operations needed for a programming and specification language although it is not very expressive. In the context of this tiny kernel language it is relatively easy to prove the correctness of a transformation. The semantics of WSL are defined by denotational semantics where a program is defined as a function from states to sets of states. The kernel language is extended into an expressive programming language by composing new constructs out of primitive kernel statements (see [War04] for details). The “Wide Spectrum” in “Wide Spectrum Language” refers to the range of operations, from “low-level” constructs such as assignments, IF statements and gotos to “high-level” operations such as specification statements. The language is also “wide” because it suits all levels of reverse engineering [YW03]. The FermaT software migration approach to understanding and migrating

assembler code involves currently four stages:

1. Translate the assembler to WSL;
2. Translate and restructure data declarations;
3. Apply generic semantics-preserving WSL to WSL transformations;
4. Apply task-specific operations as follows:
  - For migration: translate the high-level WSL to the target language;
  - For analysis: apply slicing or abstraction operations to the WSL to raise the abstraction level.

Unlike simple line-by-line language migration technologies, the unique FermaT semantic preserving code transformations enable the original application to be automatically cleaned-up, simplified and restructured to its optimum state for migration to the chosen new language. The FermaT technology is currently used and distributed by Software Migrations Ltd. (SML)<sup>1</sup> to migrate large-scale commercial IBM 360 assembler systems into higher-level languages like C or COBOL. In a recent migration project a full commercial assembler system consisting of 2,296 modules with an average of 6,149 lines of code per modules with 11,959,084 lines in total was condensed down into 1,120,449 lines of C code with an average of 576 lines of code per module. The complete migration from assembler to C took 7 hours 38 minutes CPU time, averaging just over 14 seconds per module, running on a 2.6GHz PC. The FermaT Transformation System applied a total of 4,167,286 transformations, averaging 2,143 transformations per module and 152 transformations per second (for further details see [WZLN08]).

## 2 Wide Spectrum Type Theory

The translation process from or to WSL normally involves many assumptions as the semantic meaning of certain language statements differ slightly from programming language to programming language. The absence of an adequate, generalised and practical solution for the translation problem motivated the invention of a special layered type system for the FermaT Transformation System which is called Wide Spectrum Type System. The main aim of the research about the Wide Spectrum Type System is to create a theory of a layered type system for the WSL language and to research a transformation based approach to introduce type safety to an untyped legacy system. The type checker together with the precisely defined type system is able to identify dubious code [Car04, CW85]. With the type system it is possible to store and specify extracted data structures from an untyped assembler systems directly in WSL code. The result is a type safe WSL program which can be checked and afterwards translated to languages with a very strong type system [PS94, Pie02].

## Literaturverzeichnis

- [Car04] L. Cardelli. Type Systems. The Computer Science and Engineering Handbook, 2004.
- [CW85] L. Cardelli und P. Wegner. On Understanding Types, Data Abstraction, and Polymorphism. ACM Computing Surveys, 17(4):471–522, Dezember 1985.
- [Dij76] E. W. Dijkstra. A Discipline of Programming. Prentice-Hall, Englewood Cliffs, NJ, 1976.
- [Mor88] C. Morgan. The Specification Statement. ACM Transactions on Programming Languages and Systems, 10(3):403–419, 1988.
- [Pie02] B. C. Pierce. Types and Programming Languages. The MIT Press, Cambridge, Massachusetts, 2002.
- [PS94] J. Palsberg und M. I. Schwartzbach. Object-Oriented Type Systems. Wiley Professional Computing, 1994.
- [War89] M. Ward. Proving Program Refinements and Transformations. Dissertation, University of Oxford, 1989.
- [War04] M. Ward. Pigs from Sausages? Reengineering from Assembler to C via FermaT Transformations. Science of Computer Programming, Special Issue on Program Transformation, 52:213–255, 2004.
- [WCM89] M. Ward, F. W. Calliss und M. Munro. The Maintainer’s Assistant. In Proceedings of the International Conference on Software Maintenance, Page 307. IEEE, IEEE Computer Society Press, 1989.
- [WZLN08] M. Ward, H. Zedan, M. Ladkau und S. Natelberg. Conditioned Semantic Slicing for Abstraction; Industrial Experiment. Software - Practice and Experience, Accepted for publication, 2008.
- [YW03] H. Yang und M. Ward. Successful Evolution of Software Systems. Artech House, Inc., 2003.

# Beschreibung von agentenbasierten Anwendungen

Dominik Masur, Daniel Sevrin

*{Dominik.Masur,Daniel.Sevrin}@rwth-aachen.de*

**Abstract:** Die Bereitstellung von agentenbasierten Anwendungen ist auf Grund der sehr heterogenen Agentensysteme und fehlender Standards ein komplexer und zeitaufwendiger Prozess. Ein Teilprozess ist die Beschreibung der Agenten und deren Eigenschaften sowie der Agentengesellschaften und ihrer Zusammensetzung. Diese Beschreibungen werden von einem Werkzeug verarbeitet und dadurch, das Starten, Stoppen und die Überwachung von agentenbasierten Anwendungen automatisiert. Hier wird ein existierendes Metamodell für die Beschreibung der Agenten und Agentengesellschaften und das dafür entwickelte Werkzeug, den Agent-Society Configuration Manager and Lancher (ASCML), beschrieben. Für die Beschreibung von Agenten und Agentengesellschaften wird ein neu entwickeltes wissenbasiertes Werkzeug vorgestellt, welches Beschreibungen erzeugt, die dem Metamodell genügen und vom ASCML verwendet werden können.

## 1 Einführung

Ein Agent ist ein autonomes Programm, welches mit seiner Umwelt interagiert. Agenten arbeiten unabhängig voneinander und kommunizieren mit Hilfe von Kommunikationssprachen und Interaktionsprotokollen mit anderen Agenten. Intelligente Agenten können aus Erfahrungen lernen und ihr Verhalten an neue Gegebenheiten anpassen. Die Stärke der Agententechnologie liegt in den verfügbaren Modellierungsmethoden und der schon integrierten Unterstützung von Agentensystemen für dynamische und verteilte Einsatzgebiete. Ein Schwachpunkt beim Einsatz agentenbasierter Anwendungen war die fehlende Unterstützung für die Beschreibung und Orchestrierung von Agentengesellschaften, die die Funktionalität einer Anwendung erbringen. Das für diesen Zweck entwickelte Werkzeug ASCML unterstützt die Erstellung von Beschreibungen für Agenten und Agentengesellschaften, bietet aber nach unseren Erfahrungen keine benutzerfreundliche Bedienung. Hierdurch wird die Verwendung so erschwert, dass keine wesentliche Verbesserung erreicht wurde. Um dem entgegenzuwirken wurde für die Erstellung von Agenten und Agentengesellschaften ein neues, benutzerfreundliches Werkzeug entwickelt. Dies beruht auf dem Wissensmodellierungswerkzeug Protégé und stellt die Beschreibungen in einem Format zur Verfügung, die dem existierenden Metamodell genügen.

## 2 Theoretischer Ansatz

Das Beschreiben von Agenten und Agentengesellschaften wird mit Hilfe eines Metamodells realisiert. Um die Verwendung von Agenten in mehreren Agentengesellschaften zu ermöglichen werden die Beschreibungen für Agenten und Agentengesellschaften getrennt betrachtet und in XML-Dateien abgelegt. Die Agentenbeschreibung enthält eine informelle Beschreibung des Agenten, dessen eindeutigen Namen, den Pfad zu der Datei, in welcher der Agent gespeichert ist und eine Beschreibung der Parameter des Agenten. Für jeden Parameter wird ebenfalls eine informelle Beschreibung, ein Vorgabewert sowie ein durch den Benutzer angegebener Wert hinterlegt. Die Beschreibung einer Agentengesellschaft enthält Verweise auf die Beschreibungen von Agenten oder anderen Agentengesellschaften, die Teil der beschriebenen Agentengesellschaft sind. Des Weiteren werden Abhängigkeiten der Agenten oder Agentengesellschaften von anderen Agenten, Agentengesellschaften, Diensten, externen Ereignissen sowie logischen Ausdrücken darüber hinterlegt. Für eine Agentengesellschaft können mehrere Konfiguration definiert und eindeutig bezeichnet werden. In Abbildung ?? wird die Erzeugung der XML-Dateien für die

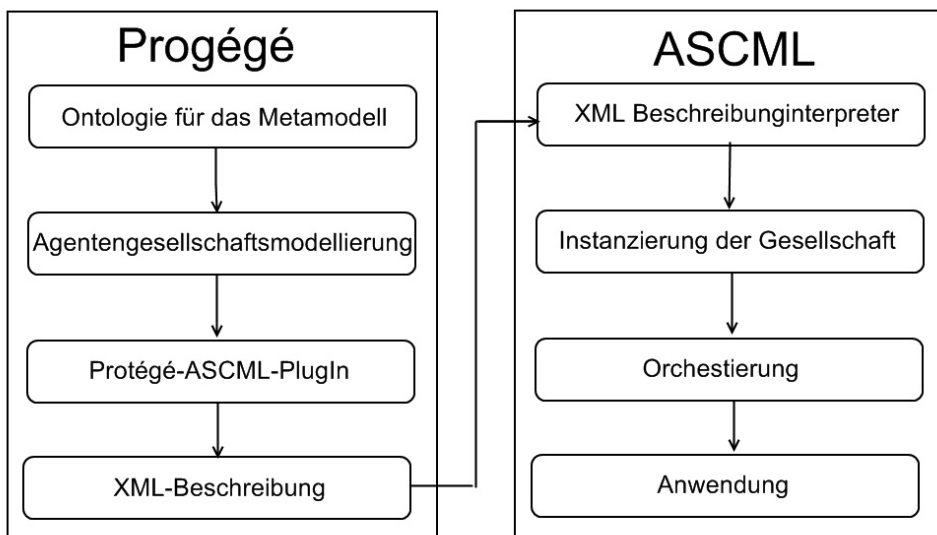


Abbildung 1: Protege in Verbindung mit ASCML

Beschreibung der Agenten und Agentengesellschaften dargestellt. In dem Wissensmodellierungswerkzeug Protégé kann mit Hilfe einer Ontologie die Beschreibung der Agenten und Agentengesellschaften benutzerfreundlich vorgenommen werden. Die dadurch modellierten Beschreibungen werden mit Hilfe des neu entwickelten ASCML PlugIns für Protégé in Form von XML-Dateien exportiert und können später von dem ASCML für die Orchestrierung von Agenten und Agentengesellschaften verwendet werden. Dabei werden

die XML-Dateien interpretiert, die darin beschriebenen Agenten und Agentengesellschaften instanziiert und anschliessend orchestriert, so dass der Funktionsumfang der agentenbasierten Anwendung erbracht werden kann.

### 3 Praktische Umsetzung

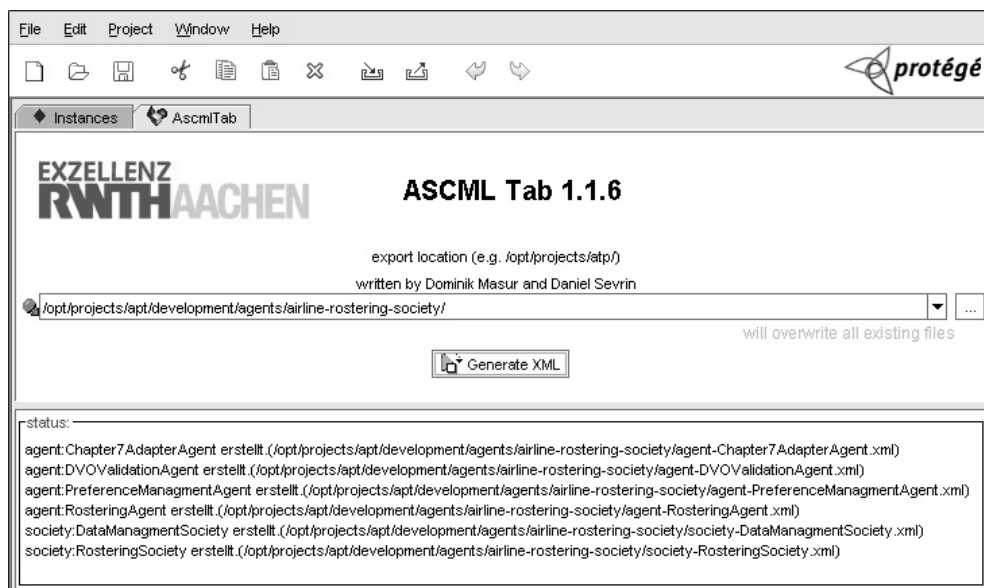


Abbildung 2: Das Protege-PlugIn ASCMLTab

Basierend auf dem Objektmodell für die Wissenspräsentation von Protégé wurde von den ontologischen Konzepten zur Beschreibung des Metamodells für die Agenten und Agentengesellschaften die Strukturen der XML-Dateien abgeleitet. Diese konnten dabei auf einfache Art und Weise durch einen Pre-Order-Durchlauf auf einer Baumstruktur generiert werden, wobei in den einzelnen Knoten des Baues die semantische Dekoration der entstehenden XML-Strukturen vorgenommen wurde. In Abbildung ?? ist die Ausgabe des Protégé-ASCML-PlugIns dargestellt und in Abbildung ?? dessen benutzerfreundliche Oberfläche.

### 4 Zusammenfassung und Ausblick

Mit Hilfe des neuen Protégé-ASCML-PlugIns wurde eine benutzerfreundliche Erstellung von Beschreibungen von Agenten und Agentengesellschaften ermöglicht. Hierdurch wird



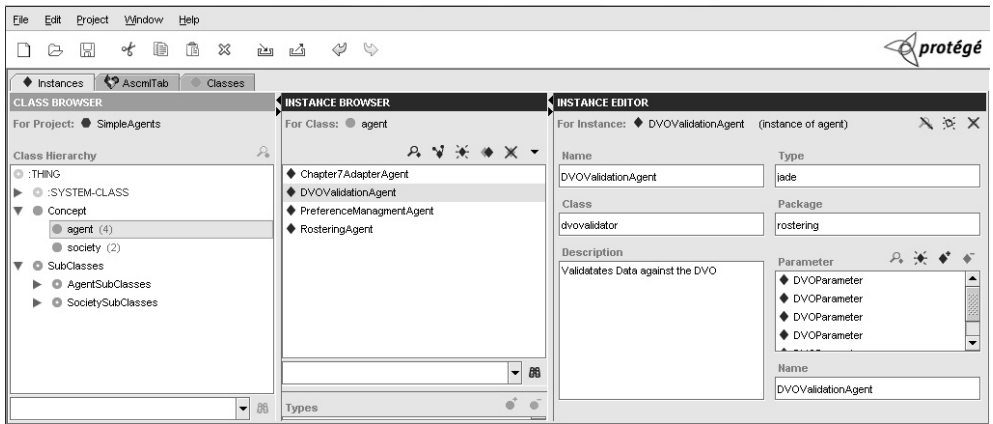


Abbildung 3: Das Protege-PlugIn ASCMLTab

Entwicklern von agentenbasierten Anwendungen die Bereitstellung dieser Anwendungen vereinfacht. Das neu entwickelte Protégé-ASCML-PlugIn wird unter der offenen OpenSource-Lizenz Apache Licence 2.0 veröffentlicht und über die Webseiten von Protégé und des ASCML angeboten.

## Literatur

- [BADE 2007] *Developing Multi-Agent Systems with JADE*; Agent-Society Configuration Manager and Launcher; 2007; Dirk Bade, Sven Lilienthal, Karl-Heinz Krempels und Ananda Sumedha Markus Widyadharma
- [BRAUB 2004] *Fifth International Workshop on Engineering Societies in the Agents World*; Deployment of Distributed Multi-Agent Systems; 2004; Lars Braubach and Alexander Pokahr and Karl-Heinz Krempels and Winfried Lamersdorf
- [LILEN 2006] *Informatiktage*; Agent Society Configuration Manager and Launcher; 2006; Sven Lilienthal und Sumedha Widyadharma
- [INFOR 2006] *Informatiktage*; Informatiktage 2006: Fachwissenschaftlicher Informatik-Kongress, 31. März und 1. April 2006, B-IT Bonn-Aachen International Center for Information Technology Bonn; 2006; GI



# Qualitätssicherung im Business Process Modeling durch automatische Ableitung visueller Kontrakte aus heterogenen UML-Modellen

Michael Mlynarski

Universität Paderborn  
Institut für Informatik  
funky@uni-paderborn.de

**Abstract:** Sowohl in der Fach- als auch der Forschungsliteratur werden Anwender mit Stichwörtern wie SOA, BPMS, WfMS, u.a. überschüttet. Vielzahl der innovativen Konzepte verfolgt den modellbasierten Ansatz. Ziel ist es aus den Modellen möglichst automatisch Code zu generieren. Dieses Vorgehen soll serviceorientierte Architekturen schnell und kostengünstig generieren. Die Schwachstelle solcher Generierungsprozesse liegt in der transparenten Qualitätssicherung. Um die Konsistenz zwischen der abstrakten Modellierungsebene (Geschäftsprozesse) und der Implementierung zu stärken, wende ich in meiner Arbeit die sog. visuellen Kontrakte um das Problem der Qualitätssicherung anzugehen.

## 1 Das Problem

Der Geschäftserfolg heutiger Organisationen steht und fällt mit einer mehr oder weniger gut entwickelten IT Infrastruktur. Eine innovative Softwarearchitektur zeichnet sich aus durch Eigenschaften wie Skalierbarkeit, Standardisierung oder Interoperabilität. Veränderungen im Organisationsumfeld bedeuten gleichzeitig eine effizientere Reaktion auf diese Erfordernisse durch eine geeignete IT Struktur. Wünschenswert ist also eine Abbildung der Geschäftsprozesse innerhalb der Softwarelandschaft. Ein gutes Konzept stellen hier die Serviceorientierten Architekturen (SOA) dar. Speziell die geschäftsprozessorientierte Variante (Enterprise SOA) basiert auf der Abbildung der Business Processes innerhalb der Softwarearchitektur [NL04]. Eine solche Darstellung (und speziell die Realisierung der Geschäftsprozessänderungen) kann viele Probleme in Hinsicht auf die Qualität bewirken. Allein die Implementierung eines Services muss Anforderungen eines Geschäftsprozesses (oder dessen Teile) erfüllen. Folgendes Problem tritt auf: Die Qualitätsprüfung bezüglich der Anforderungen muss bereits in der Modellierung der Geschäftsprozesse erfolgen und transparent bis hin zur Implementierung gewährleistet sein.

Das Problem wird deutlicher, wenn man sich die fatalen Folgen einer unzulänglichen Implementierung der Anforderungen in der Softwarearchitektur eines großen Unternehmens vorstellt. Die Kosten jeder Neuentwicklung und Änderung im laufenden System sind enorm. In einem kritischen System kann es sogar einen temporären Ausfall des Be-

etriebes bedeuten. Somit spielt die Qualität der Softwarearchitektur eine sehr bedeutende Rolle. Es wird nach Konzepten geforscht, die eine Qualitätskontrolle komplexer Systeme gewährleisten oder mindestens verbessern helfen.

## 2 Ein Lösungsansatz

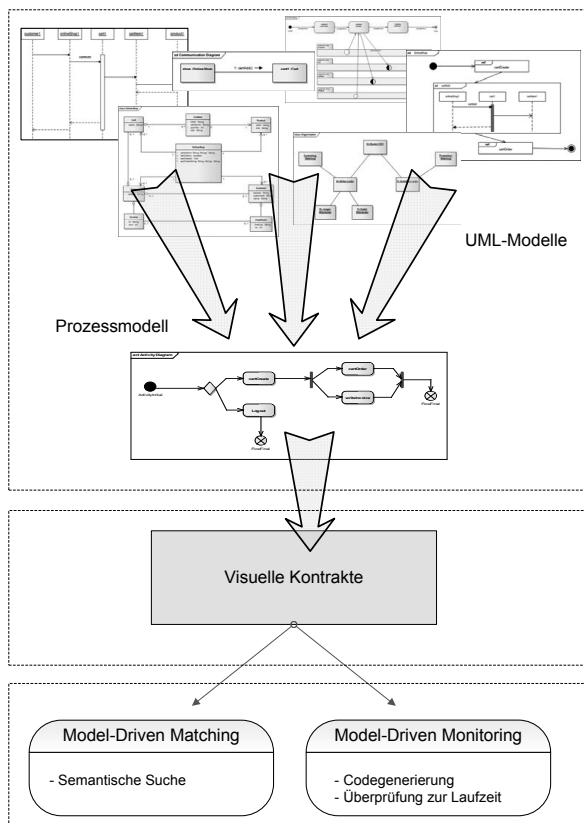


Abbildung 1: Grafische Darstellung des Generierungsprozesses. Für jede Aktion eines Geschäftsprozesses werden visuelle Kontrakte in Form von Objektdiagrammen generiert und aus diesen Definitionen Java Skeletons erstellt

Einen bereits sehr bekannten Standard in der Modellierung der Softwarearchitekturen und Systeme stellt die Unified Modeling Language (UML) vor [BME<sup>+</sup>07]. Mit dieser Sprache lassen sich Geschäftsprozesse in Form von UML Aktivitätsdiagrammen beschreiben. Die Verifizierung dieser Modelle kann mit Hilfe von Dynamic Meta Modeling (DMM) durchgeführt werden, wie von Soltenborn in [Sol06] vorgestellt. Ein Gesamtsystem kann mit verschiedenen Diagrammartentypen modelliert werden (bspw. Klassendiagramme, Inter-

aktionsdiagramme, etc.). Auf diese Weise können Architekturen wie die Enterprise Service Oriented Architecture fast vollständig definiert werden (wie in [EP00], [Mar00] oder [EHFT05] beschrieben). In einem modellbasierten Ansatz wird aus den Modellen Quellcode für die spätere Implementierung generiert. Diese soll natürlich die gewünschten Anforderungen erfüllen. Ein Konzept für die Definition der Anforderungen in Form von UML Objektdiagrammen bieten die sog. visuellen Kontrakte. Sie beschreiben die Semantik einer Operation durch Vor- und Nachbedingungen (gemäß Design by Contract [Mey92]). Die Kontrakte lassen sich automatisch (mit Hilfe der Java Modeling Language) in Java Skeletons überführen [Loh06]. Es gibt also zwei Konzepte (DMM und visuelle Kontrakte), mit deren Hilfe das Problem der Qualitätssicherung angegangen werden könnte.

Der Generierungsprozess wird in Abbildung 1 skizziert. Im Laufe eines UML-orientierten Softwareentwicklungsprojektes wird für das zu erstellende System eine Vielzahl von UML Diagrammen erstellt. Angefangen bei der abstrakten Prozessdefinition (Aktivitätsdiagramm), durch die Modellierung des Geschäftsfeldes (Klassendiagramm) bis hin zu exakten Definition des Verhaltens (Sequenzdiagramm, u.a.). Alle diese Modelle können relevante Informationen beinhalten, die für die automatische Generierung der visuellen Kontrakte eingesetzt werden können. Ein spezieller Algorithmus analysiert die Modelle und extrahiert für jede Aktion des Prozesses die gewonnenen Informationen. Die somit generierten visuellen Kontrakte werden in einer eindeutigen Form gespeichert. Anhand eines Tools Namens Visual Contract Workbench können Java-Skeletons erzeugt werden, welche die entsprechenden Anforderungen (Vor- und Nachbedingung) beinhalten. Wir sprechen dann von dem sog. Model-Driven Monitoring. Eine semantische Suche in einem serviceorientierten Umfeld anhand der visuellen Kontrakte ist mit Hilfe des sog. Model-Driven Matchings möglich. Somit wird eine vereinfachte Enterprise Service Oriented Architecture im Top-Down Ansatz erzeugt.

### 3 Zusammenfassung und Ausblick

Durch die Nutzung des oben beschriebenen Vorgehens kann die Konsistenz zwischen dem Business Process Modeling und der restlichen Systemspezifikation verbessert werden. Der Forschungsstand erlaubte es nicht, die Verbindung zwischen der Modellierungsebene und den visuellen Kontrakten zu knüpfen. Das Ziel der Masterarbeit war, aus heterogenen UML-Modellen (basierend auf der Definition der Geschäftsprozesse in Form von UML Aktivitätsdiagrammen) eine automatische Generierung von visuellen Kontrakten zu ermöglichen. Dazu wurde ein Algorithmus definiert und eine Fallstudie durchgeführt, die das Gesamtkonzept veranschaulichen soll.

Das Verfahren kann sowohl im Bereich des Model-Driven Monitorings (Generierung der Java-Skeletons) als auch des Model-Driven Matchings (semantische Suche in serviceorientierten Architekturen) angewendet werden. Die generierten Kontrakte können auf anderen Gebieten (wie bspw. Model-Based Testing) eingesetzt werden. Weiterhin können mehrere Modelle für die Informationsextraktion angewendet werden. Abschließend ist es wichtig zu erwähnen, dass die visuellen Kontrakte nicht nur im Gebiet der Serviceorientierten Architekturen eingesetzt werden, sondern auch das korrekte Verhalten jedes beliebigen

gen Softwaresystems in unterschiedlichen Softwarearchitekturen beschreiben können. Das präsentierte Spektrum stellt ohne Frage ein interessantes Feld der aktuellen und zukünftigen Forschung dar.

## Literatur

- [BKR03] J. Becker, M. Kugeler und M. Rosemann. *Process Management*. Springer Verlag, 2003.
- [BME<sup>+</sup>07] G. Booch, R. A. Maksimchuk, M. W. Engle, B. J. Young, J. Conallen und K. A. Houston. *Object-oriented analysis and design with applications. Third Edition*. Addison-Wesley, 2007.
- [BRJ99] G. Booch, J. Rumbaugh und I. Jacobson. *The Unified Modeling Language User Guide*. Addison-Wesley, 1999.
- [EGSW07] G. Engels, B. Güldali, C. Soltenborn und H. Wehrheim. Quality Assurance for Business Process Models using Visual Contracts. 2007.
- [EHFT05] G. Engels, R. Heckel, A. Förster und S. Thöne. *Process-Aware Information Systems. Process Modeling using UML*. Wiley Publishing, New York, 2005.
- [EP00] H.-E. Eriksson und M. Penker. *Business Modeling with UML: Business Patterns at Work*. Wiley & Sons, 2000.
- [Gad03] A. Gadatsch. *Grundkurs Geschäftsprozess-Management*. Vieweg Verlag, 2003.
- [Hel03] R. Helbig. *Prozessorientierte Unternehmensführung*. Physica-Verlag, 2003.
- [Loh06] M. Lohmann. *Kontraktbasierte Modellierung, Implementierung und Suche von Komponenten in serviceorientierten Architekturen*. Dissertation, Universität Paderborn, 2006.
- [Mar00] C. Marshall. *Enterprise Modeling with UML. Designing successful software through business analysis*. Addison-Wesley, 2000.
- [Mey92] B. Meyer. Applying Design by Contract. *IEEE Computer*, 25:40–51, 1992.
- [NL04] E. Newcomer und G. Lomow. *Understanding SOA with Web Services*. Pearson Education, 2004.
- [Sol06] C. Soltenborn. Analysis of UML Workflow diagrams with Dynamic Meta Modeling techniques. Diplomarbeit, Universität Paderborn, 2006.

# Constraint Based Transformation Theory

Stefan Natelberg

De Montfort University  
Computing Sciences and Engineering  
stefan.dmu@web.de

Art der Arbeit: Abstract

Betreuer/in der Arbeit: Prof. Hussein Zedan, Prof. Karl Hayo Siemsen

**Abstract:** The migration of a legacy system is one of the most challenging tasks in the area of software evolution. There exist several different approaches to migrate a legacy system with different advantages and disadvantages. One solution is to translate the software into an intermediate language which spans from the low abstraction level to the very high abstraction level [1]. Once the software has been translated to intermediate language, the abstraction level can be changed via mathematical proven transformation rules. Afterwards, a translator can easily translate the intermediate language into the target language [2].

To achieve an even better and more effective overall program transformation tactic, the results of several code transformation sequences for a specific project have to be compared and evaluated. The huge number of transformation rules and the search-space of possible transformation sequences make exhaustive search infeasible [3]. Given a transformation system with 20 possible code transformations and a sequence length of 20, there are 2020 possible combinations in the search-space.

To find suitable transformation sequences for specific requirements we have developed a formal language which is able to express transformation sequences. Powerful enough to describe complex transformation sequences, simple enough to be useful even for inexperienced maintainers. Also, a maintainer is able to influence the transformation process by defining constraints which have to be satisfied in certain states of the process [4]. The constraints support a priority system to avoid conflicts between them, and it is possible to group them to define abstract requirements.



# 1 Code Transformation Control via Regular Expressions

Dies To find suitable transformation sequences for specific requirements on the transformation result, we have developed a formal language which expresses transformation sequences. This language should be powerful enough to describe complex transformation sequences but also be simple enough to be useful even for people who are not very familiar with program transformation theory. Therefore, the maintainer should be able to describe a program transformation in a very abstract way and also to control the program transformation by defining the constraints which have to be satisfied. This constraints support a priority system to avoid conflicts between them, and it is possible to group constraints to define abstract requirements. In that way the user should be able to describe e.g. his definition of maintainability. The table 4.1 shows the Backus-Nauer-Form of the formal language which is very simple at this early state of the research.

<b>Sequence</b>	<code>&lt;sequence&gt; ::= &lt;alternation&gt; (“,” &lt;alternation&gt;)*</code>
<b>Alternation</b>	<code>&lt;alternation&gt; ::= &lt;repeat&gt; (“ ” &lt;repeat&gt;)*</code>
<b>Repeat</b>	<code>&lt;repeat&gt; ::= &lt;option&gt; (“*”)*</code>
<b>Option</b>	<code>&lt;option&gt; ::= &lt;transition name&gt; (“?”)</code>
<b>Transition Name</b>	<code>&lt;transition name&gt; ::= &lt;string&gt;   (“(” &lt;sequence&gt; “)”</code>

Abbildung 1: The Backus-Nauer-Form of the constraint based transformation language.

We choose a regular expression language for our approach because we need to convert it into a finite automaton. This automaton visualizes the transformation process. Therewith, the user is able to observe the entire transformation process which is a huge advantage for transformation comprehension.

# 2 Code Transformation Comprehension via Automata

A finite automaton consists of states and transitions. After we have parsed the regular language, every state of the automaton is a representation of a code version during the program transformation process. The initial state  $S_0$  is the unchanged code which the user puts into the system. The accepted or final state  $s_n$  is the program transformation result. The transitions between the states are the atomic code transformations of the FermaT Transformation System [2]. Because there are several ways to travel the automaton, we need additional user information to find the best way for the requirements of the user. This will be controlled by the constraints. If there are no constraints at all, the system will return several program transformation results and the user has to choose his favourite manually. The automaton and all details of the program transformation can

be shown in a graphical window. The user is able to obtain the entire transformation process and can take influence if necessary.

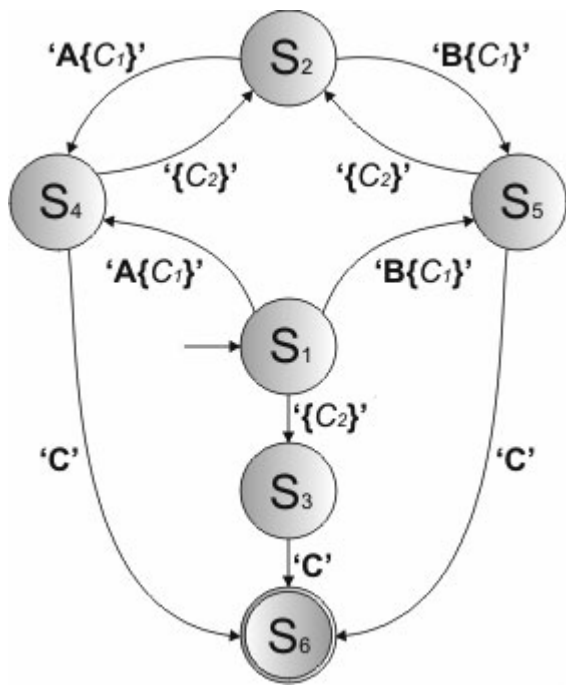


Abbildung 2: Example automaton.

The Constraint Based Transformation System is based on JAVA 6.0 and is an extension of the FermaT Maintenance Environment [5]. It consists of an editor and a parser for the formal language and an automaton generator to convert the language into an equivalent automaton [5]. Furthermore, the system provides several strategies to measure WSL code and to generate transformation sequences based on user defined constraints.

## Literaturverzeichnis

- [1] H. Yang and M. Ward. Successful Evolution of Software Systems. Artech House Publishers, January 2003.
- [2] M. Ward. Assembler to C migration using the FermaT Transformation System. International Conference on Software Maintenance, Oxford, August 1999.
- [3] D. Fatiregun, M. Harman, and R.M. Hierons. Evolving transformation sequences using genetic algorithms. In Department of Information Systems and Computing, Brunel University, 2004.
- [4] Martin Ward, Hussein Zedan, Matthias Ladkau, and Stefan Natelberg. Conditioned semantic slicing for abstraction; industrial experiment. Software - Practice and Experience, Accepted for publication, 2008.
- [5] Stefan Natelberg. Constraint Based Transformation Sequences. Transfer Report, De Montfort University, STRL, 2007.

# Towards non-exploitable processes: Extending the least privilege access principle support in OpenSolaris™

Johannes Nicolai  
johannes.nicolai@student.hpi.uni-potsdam.de  
Hasso-Plattner-Institut Potsdam, Potsdam, Germany

## Abstract

According to the least privilege access principle (LPAP)<sup>1</sup> which tries to minimize the potential damage in case of attacks, a process should restrict itself to be only allowed to perform actions it really needs to fulfill its job. This paper describes the development of the LPAP support from Solaris 1 to Solaris 10 and illustrates how LPAP support in OpenSolaris was extended by the author of this paper to allow processes running under the same user id to individually decide which file access and manipulation actions the operating system should grant to them.

## 1 Introduction

In almost every well-known operating system, processes started by a certain user run in the security context of this user. Two processes started by the same user have exactly the same power concerning the actions they are allowed to perform regardless of their task they should actually fulfill. Normally, the fact that a process is able to perform actions it does not need to perform in order to fulfill its task should not be a problem, because it does not perform them. This becomes a problem in the moment a process is attacked by an intruder that forces it to perform actions that were not intended by the programmer. In the worst case, the intruder is able to let the process execute arbitrary code in its security context.

This problem leads us to the least privilege access principle (LPAP): A process should only have the power to perform actions it also needs to fulfill its tasks.

Currently, it is not possible in OpenSolaris to restrict a process to only access world readable/writable/executable files, but this possibility is desired for processes that require access to global libraries and configuration files, but does not rely on the identity of the user that started the process. There are even situations a process does not need to access any file after starting up or does not need read/write/execute permissions at all.

---

<sup>1</sup>The least privilege access principle is also known under the term *principle of least privilege* [Nic05].

This paper will specify the necessary changes in OpenSolaris to exactly address these problems.

## 2 Solaris 1 Behaviour

In Solaris 1 either a process is running under the context of the super user (`euid == 0`) and may perform every operation other users might do plus some actions no other user is allowed to do<sup>2</sup> or it is running under the context of an ordinary user (`euid != 0`) and is only allowed to perform actions on objects that explicitly grant access rights for the user, its associated groups or for everybody.

It is not possible for a process to specify that it does not need to read the contents of files at all, even if it has the `euid`, `egid` or group memberships to do so, does not need entering owned directories but only globally searchable directories or does need reading every file but no write access at all.

To improve the LPAP support, the Solaris security model had to be constructed out of removable privileges, because if a process is able to remove certain privileges it does not need to fulfill its task, it cannot be exploited to use them. If privileges are arranged on multiple layers, the process has fine grained control over what privilege it really needs and what privileges it can drop without losing too much power.

## 3 Solaris 10 privileges

Up to Solaris 10, the Solaris operating system was based on the concepts of the Solaris 1 security model. With Solaris 10, a *privilege* based security model was introduced. You may imagine privileges as removable behaviour bricks. Whenever a privilege is defined to grant a certain action, the check is based on this privilege being set in the effective privilege set and not on the effective uid any longer. So if a process does not need a privilege, it may simply drop it from the effective privilege set. If it also drops the privilege from its permitted privilege set, the privilege is also removed from its effective set and there is no possibility for the process to get it back.

In contrast to the Solaris 1 security approach, privileges are now arranged in two layers whereas the bricks on super user level may be removed by the process or may not be present at process startup. Typically, a process started by the super user has associated all defined privileges with it, whereas a process started by an ordinary user may only use privileges on the ordinary user level.

The main difference to the Solaris 1 security approach is that now super user processes may decide to permanently drop their power to write every file, while they do not lose their power to read every file or vice versa. It is still not possible for a process to specify

---

<sup>2</sup>Actions like creating raw sockets, binding ports under 1024 or chrooting in restricted environments can only be performed by a process with `euid == 0`.

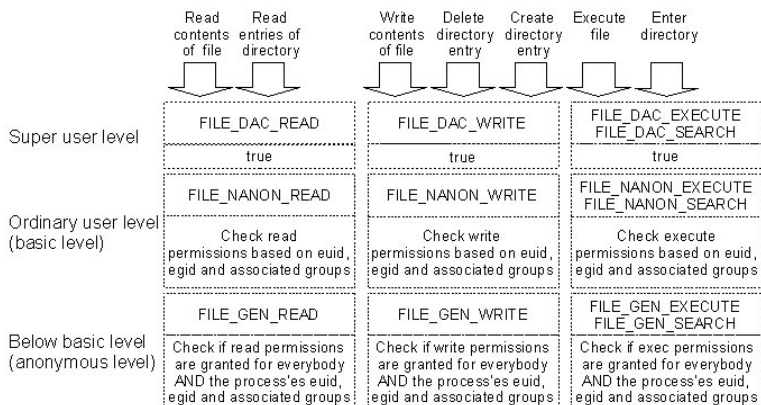


Abbildung 1: Proposed OpenSolaris File Access Behaviour

that it e. g. does not need to write the contents of files at all, even if it has the euid, egid or group memberships to do so or does not need to read the contents of owned files but only of globally readable files. This is because the privileges on the ordinary user level are not (yet) removable. To make these privileges removable is part of the next section.

The Solaris 10 privilege model also defined five *basic privileges* – removable behaviour bricks at ordinary user level – that are also present in the privileges sets for ordinary processes as well. So an ordinary process may remove some or all of these basic privileges from its effective or permitted privilege set and restrict its power below the normal level. The set of all defined basic privileges is called basic privilege set.

## 4 Extending the basic privilege set

Dropping the currently defined basic privileges in OpenSolaris only restricts some actions related with forking, signaling and observing processes in foreign process groups as well as hardlinking to non-owned files and executing binaries. There are no removable privileges in Solaris 10 to turn file access completely or partially off or to restrict the capability of a process to change meta information of its owned files.

In most cases, switching read/write/execute file access completely off is a too radical approach. The processes may still want to load globally readable libraries and configuration files and may need to change globally writable files or directories. This brought me to the idea of introducing a third layer of privileges – the *anonymous* level. Anonymous processes should be processes that choose to (partially) give up the user's identity, which means they cannot benefit from the fact that they are running with a special effective uid and are a member of certain groups<sup>3</sup>. In Figure 1 you see the results.

<sup>3</sup>Anonymous processes are not anonymous in a way that they are hidden in the process table, they only lose the power to benefits from the object access rights associated with their identity.

The semantics of the newly introduced basic privileges may explained with some examples (remaining privileges behave similarly). Without having set `FILE_NANON_READ`, a process is only able to perform read operations on globally readable files or directories because it can only act like an 'anonymous process' and cannot benefit from its `euid`, `egid` or group memberships. Additional rights gained through `FILE_DAC_READ` will not be restricted if `FILE_NANON_READ` is not set.

Without having set `FILE_GEN_WRITE`, file or directory operations that require write permissions will generally not be granted unless the `FILE_DAC_WRITE` or `FILE_NANON_WRITE` privilege is set. Additional rights gained through `FILE_DAC_WRITE` or `FILE_NANON_WRITE` will not be restricted if `FILE_GEN_WRITE` is not set.

As you may have noticed from the descriptions, the presence or absence of the new privileges is ignored if a privilege on a higher level is present in the same brick column. If for example, a privilege on super user level is present that grants read permissions for every file, it does not matter whether the process is anonymous or not. If the behaviour brick on basic level is present that allows writing the content of files the user associated with the effective `uid` of the process has write access to, it does not matter whether the privilege on anonymous level was removed to switch write access totally off.

The new removable privileges introduced up to this point are only associated with actions that operate on already existing files. Should an anonymous process be able to create files or change meta information of files owned by the user associated with the processes' effective `uid`? It probably does not because an anonymous process has given up its identity and therefore cannot create or modify meta information of files that are owned by this identity. To address this problem, I introduced another new basic privilege on basic level: `FILE_NANON_OWNER`. This privilege allows a process to perform operations that require ownership on a file, link or directory.

## 5 Acknowledgements

This work would not have been possible without the help of my Google Summer of Code mentor Darren J Moffat who encouraged me to expand the basic privilege set of OpenSolaris, provided the necessary technical information and is responsible for integrating my changes into the official kernel that can be found at [NM06]. I would also thank Google who paid me during the three months of implementation and Andreas Polze and Martin von Löwis who gave advice on how to structure this paper.

## Literatur

- [Nic05] Johannes Nicolai. Sichere Ausführung nichtvertrauenswürdiger Programme. Bericht, Hasso-Plattner-Institut Potsdam, Potsdam, 2005.
- [NM06] Johannes Nicolai und Darren J Moffat. *Open Solaris Security Project: Restricting file access by basic privileges*. <http://www.opensolaris.org/os/community/security/projects/gsoc2006/>, 2006.

# Ontology Composition using a Role Modeling Approach

Michael Pradel

michael@binaervarianz.de

**Abstract:** There is an increasing use of ontologies to represent domain knowledge, however, no notion of ontology components has so far been established. This lack hampers partial reuse of ontologies and makes a reuse-oriented engineering approach difficult. We propose to enable component-based ontology engineering by bringing together ontologies and role modeling, a well-known modeling paradigm from the object-oriented software community. Ontological role models provide an intuitive reuse unit, ontology components, and allow for a clearer and more natural way of modeling.

## 1 Introduction

An ontology is a model used to capture and represent knowledge by describing the concepts of a domain and relations between them. Ontologies are employed in artificial intelligence, software engineering, and biomedical informatics, as well as in the semantic web [BLHL01] to enhance web content with semantics. The distinguishing feature of an ontology compared to other knowledge representation mechanisms is its underlying formalism. Most ontology languages are based on Description Logics [BCM<sup>+</sup>03] that allow, for instance, to derive new knowledge from existing one and check the consistency of a given knowledge base.

A major question discussed in the research community is how to enable (partial) reuse of ontologies in order to build new ontologies from existing parts. One of today's most important ontology languages is the Web Ontology Language OWL, standardized and recommended by W3C [DS04]. Unfortunately, it provides very limited reuse mechanisms, namely *linking* and the *owl:imports* construct, that seem to be inflexible in the kind of reuse they provide, in particular regarding the granularity of components.

Our proposed solution makes use of *role modeling*, a well-known concept from object-oriented software engineering [Ree96, Ste00, Rie00]. The main idea is that objects can be abstracted in two ways. One abstraction regards their inherent properties and classifies them accordingly. This yields *natural types*, or classes, the major modeling concept in object-oriented software development as well as in ontologies. The other abstraction analyzes how objects collaborate in order to achieve a specific goal and leads to *role types* [Sow84]. *Person* for example is a natural type, while *Parent*, *Child*, or *Teacher* are role types. Related role types can be joined to a *role model*, a self-containing component describing one particular collaboration between objects.



This work tries to solve the problem of limited reuse in ontologies with the help of role modeling. We present *ontological roles* and *ontological role models* [PHA07], an extension of ontologies able to improve modeling and provide means for reuse. Our main contributions are:

- The conceptual idea of ontological roles illustrated by examples.
- Definition of ontological roles and role models on a conceptual and formal level.
- A methodology for role-based ontology engineering (omitted here).
- A syntactical extension of the Web Ontology Language OWL and translational semantics accompanied by a usable implementation.

## 2 Using Roles in Ontologies

The exclusive use of classes in ontological modeling to abstract from similar individuals has a number of deficiencies. First, classes are relatively coarse-grained units of abstraction. This not only leads to complex classes, but also to advanced complexity of their relations. Second, ontology classes often merge properties from different contexts, that is, multiple concerns are mixed into one conceptual unit. Finally, high complexity of class relations and missing separation of concerns hamper ontology reuse and composition. We argue that these problems can be overcome by introducing roles into ontologies as an additional modeling concept.

In object-oriented software development, a role describes the *behavior* of an object in a certain context. In ontologies, though, there is no dynamism or behavior. Therefore, an ontological role describes context-dependent *properties*.

### **Definition: Ontological roles and role types**

An ontological role describes the properties of an individual in a certain context. A set of roles with similar properties is abstracted by an ontological role type.

Based on this, we define what we consider a role model in an ontological setting.

### **Definition: Ontological role models**

An ontological role model describes a set of related ontological role types and as such encapsulates common relationships between ontological roles.

The ontology in Figure 1 includes three classes *Wine*, *Winery*, and *Food* representing natural types (rectangles with gray background), as well as five roles types (rectangles with rounded corners and white background). The roles types can be grouped into the two role models *Product* and *Meal*. An individual of class *Wine* may appear in different contexts, for instance being a product produced by a winery, or being a drink that accompanies a

meal. Depending on the context, individuals play completely unrelated roles, and thus, we express them as role types *Product* and *Drink*. This separation of concerns leads to a clearer and more natural way of modeling. Furthermore, reuse is facilitated: E.g. *Product* appears in a self-containing role model with two more role types (*Consumer* and *Producer*) providing an ontological component that may be applied in other ontologies as well.

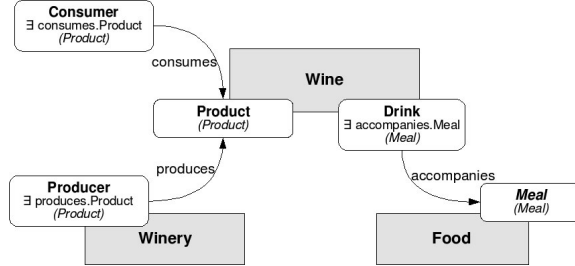


Figure 1: Different concerns of the *Wine* class are separated by the role types *Product* and *Drink*.

### 3 Semantics of Role-Based Ontologies

Description Logics (DLs) are the underlying formalism of most ontology languages. We therefore propose an extension of the  $\mathcal{ALC}$  DL. In order to be able to use existing tools such as reasoners, we provide a translation from our extended DL syntax into standard DL.

We introduce two axioms into the  $\mathcal{ALC}$  DL: The *role binding* axiom  $R \triangleright C$ , where  $R$  is a role type bound to the class  $C$ , and the *role assertion* axiom  $R(a)$  expressing that a specific individual  $a$  plays a role  $R$ .

The extended syntax may be translated to standard  $\mathcal{ALC}$  (and thus, into a standard ontology language) by the following algorithm:

1. Make all imported role type definitions available as classes in the ontology.
2. For each role type  $R$  used in the ontology:
  - (a) Let  $\{C_1, \dots, C_n\}$  be the set of classes to which  $R$  is bound ( $R \triangleright C_i$ ). Then add axiom  $R \sqsubseteq C_1 \sqcup \dots \sqcup C_n \sqcup \perp$  to the ontology.
  - (b) For each role assertion  $R(a)$ , make the same assertion available in the resulting ontology, now referring to the class-representative for the role type  $R$ .
3. Remove all role binding axioms.

Also, we propose an extension of OWL using the Manchester OWL syntax [HDG<sup>+</sup>06] and provide its translational semantics [PHA07]. This translation has been implemented based on the Reuseware composition framework, that allows for extending arbitrary languages for invasive composition [HAJZ07].

## 4 Conclusion

Our work shows how roles can be used in ontologies. Ontological roles expose context-dependent properties of ontology classes, leading to models which are more explicit and less complex. Even more important, ontological role models constitute intuitive ontology units that can be reused. This paves the way toward component-based ontology engineering where ontology components are developed independently, shared, and reassembled.

## References

- [BCM<sup>+</sup>03] Franz Baader, Diego Calvanese, Deborah L. McGuinness, Daniele Nardi, and Peter F. Patel-Schneider, editors. *The Description Logic Handbook: Theory, Implementation, and Applications*. Cambridge University Press, New York, NY, USA, 2003.
- [BLHL01] Tim Berners-Lee, James Hendler, and Ora Lassila. The Semantic Web. *Scientific American*, 2001.
- [DS04] Mike Dean and Guus Schreiber. OWL Web Ontology Language Reference. W3C recommendation, W3C, February 2004.
- [HAJZ07] Jakob Henriksson, Uwe Aßmann, Jendrik Johannes, and Steffen Zschaler. Reuseware - Adding Modularity to Your Language of Choice. In *Proceedings of Technology of Object-Oriented Languages and Systems Europe 2007, Zurich, Switzerland*, June 2007.
- [HDG<sup>+</sup>06] Matthew Horridge, Nick Drummond, John Goodwin, Alan Rector, Robert Stevens, and Hai Wang. The Manchester OWL Syntax. *OWL: Experiences and Directions (OWLED)*, November 2006.
- [PHA07] Michael Pradel, Jakob Henriksson, and Uwe Aßmann. A Good Role Model for Ontologies: Collaborations. In *International Workshop on Semantic-Based Software Development at OOPSLA*, October 2007.
- [Ree96] Wold-P. Lehne O. A. Reenskaug, T. *Working with Objects, The OOram Software Engineering Method*. Manning Publications Co, 1996.
- [Rie00] Dirk Riehle. *Framework Design: A Role Modeling Approach*. PhD thesis, 2000.
- [Sow84] J. F. Sowa. *Conceptual structures: information processing in mind and machine*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1984.
- [Ste00] Friedrich Steimann. On the representation of roles in object-oriented and conceptual modelling. *Data Knowledge Engineering*, 35(1):83–106, 2000.

# **Define a Web 2.0 Portal Solution Based on a New Lightweight Java Environment used to Bridge Web SOA and Enterprise SOA**

Oliver Rau

Fachhochschule Kaiserslautern, Standort Zweibrücken  
Informatik  
gi\_beitrag@olira.de

Art der Arbeit: Diplomarbeit

Betreuer/in der Arbeit: Prof. Adrian Müller (FHKL), Stephan Hesmer (IBM)

## **1 Berichtsvorschau und Beschreibung**

Durch das rasche Wachstum des Internets und der daran angeschlossenen Haushalte und Unternehmen wird das Bedürfnis zentraler Dienste und „Treffpunkte“ immer größer. Applikationen wie beispielsweise eMail oder Textverarbeitungsprogramme werden zunehmend zentral im Internet angeboten und benötigen nur einen Web-Browser um diese zu verwenden. In Verbindung mit solchen Rich Internet Applications (RIA) ist Web 2.0 [AL06] die bevorzugte Art und Weise zur Umsetzung. Die in diesem Rahmen verwendeten Techniken und Dienste werden meist auf der Client Seite ausgeführt, was für den Benutzer ein Programmverhalten wie bei einer Desktop Applikation ermöglicht.

Durch den Einsatz einer Portal Software nach Web 2.0 Prinzipien wird es Möglich ein System zu entwerfen, welches für Entwickler und Administratoren genauso einfach zu verwenden ist, wie von einem nicht ganz so erfahrenen Benutzer. Da jedoch hierfür die bisher im Rahmen von Web 1.0 angewendeten Architekturen meist nur eingeschränkt oder nicht verwendbar sind müssen hier sowohl auf Client- als auch auf Server-Seite die Konzepte überarbeitet werden.

Nach diesen Prinzipien ist diese Diplomarbeit und die daraus resultierende Software entwickelt worden. Bei der Serverseite handelt es sich um eine Model-View-Controller Architektur [WI07]. Da im Rahmen einer Web-2.0 Applikation jedoch viele Aktionen auf der Client Seite durchgeführt werden, mussten einige Probleme sowohl mit Server als auch mit Client Code realisiert werden.

Abbildung 1 zeigt ein solches Konzept am Beispiel einer JEE<sup>1</sup> Umgebung. Hierfür werden Teile der Logik der Serverseite auf der Client-Seite „dupliziert“ um das Look & Feel einer Desktop Applikation zu erhalten. (Client Side API).

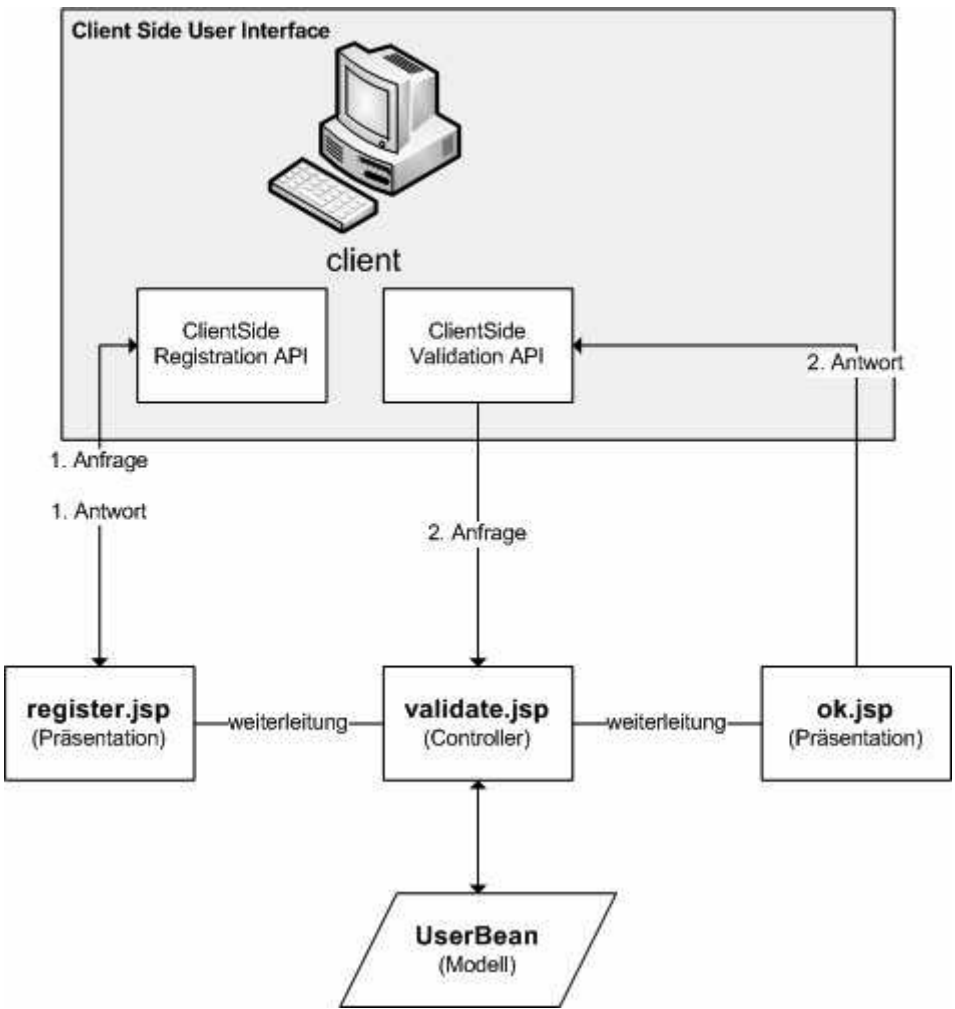


Abbildung 1: MVC in einem JEE environment mit Web 2.0 Aspekten

<sup>1</sup> JEE = Java Platform, Enterprise Edition; [SU99]

Um für Administratoren und Entwickler eine einfache Art der Verwaltung / Entwicklung zu ermöglichen wurde eine Dateisystem basierte Persistierung entworfen. Hierfür werden Verzeichnisse nach bestimmten Regeln erstellt und stellen dadurch die Portal Navigation dar. Die Inhalte des Portals werden durch die jeweiligen Dateien innerhalb der Verzeichnisse repräsentiert (z.B. HTML Quellen, Bild-Dateien, Stylesheets etc.).

Des Weiteren wurden zusätzliche Konzepte entworfen, mit denen es möglich ist, die Darstellung einer Portalseite komplexer zu gestalten. Dies wurde mit Hilfe von vordefinierten und standardisierten XML-Dateien realisiert, welche die Positionierung und das Aussehen einer Portal Seite beeinflussen. Solche XML-Dateien werden ebenfalls in den Verzeichnissen der Portal Seiten abgelegt.

Als Kommunikationsschicht zwischen dem Client- und Server-Code wurden REST-Services<sup>2</sup> verwendet, welche die Daten der Persistenzschicht in ein standardisiertes, ATOM<sup>3</sup>-basiertes Format umwandelt welches von der Client Seite konsumiert wird.

Dieses Format wurde größtenteils vom bereits existierenden IBM WebSphere Portal Web 2.0 Preview adaptiert um eine Produktkompatibilität zu gewährleisten.

Für diese REST-Services wurde mit Hilfe von verschiedenen Servlets und einem Frontcontroller Pattern eine dynamische und einfach erweiterbare Basis geschaffen.

Auf der Client Seite kommt ebenfalls aus dem Web 2.0 Preview stammende Client – Side-Aggregation (CSA) zum Einsatz. Diese wurde um einen Rich-Text-Editor erweitert, mit dessen Hilfe Portal Seiten und Inhalte nicht nur auf der Server Seite sondern auch auf der Client Seite erstellt und verändert werden können. Dadurch wird es Möglich Inhalte der Portal-Lösung ähnlich einem Wiki<sup>4</sup> zu erstellen und zu verwalten.

In der vollen Version meines Beitrags werde ich zeigen, wie die Server- und die Client-Seite aufgebaut sind und wie diese zusammenarbeiten. Hierfür werde ich zunächst auf die verschiedenen Persistierungsmöglichkeiten im Detail eingehen und wie diese mit Hilfe des Dateisystems realisiert wurde. Danach werde ich auf die Abstraktion der Schichten eingehen und erläutern wie durch diese die Daten zur weiteren Verarbeitung vorbereitet werden. Im Anschluß daran wird gezeigt wie mit Hilfe eines speziell dafür entworfenen und erweiterbaren SAX<sup>5</sup>-Frameworks das tatsächliche ATOM-Format erzeugt wird. Im Abschluss wird die Kommunikation mit der Client Seite an einer echten Anfrage gezeigt und wie diese Lösung produktisiert werden könnte bzw. bereits produktisiert wurde.

---

<sup>2</sup> REST = Representational State Transfer; [HA05]

<sup>3</sup> ATOM = Atom Syndication Format; [SN05]

<sup>4</sup> Wiki =Hawaiisch für „schnell“; [KL06]

<sup>5</sup> SAX = Simple API for XML; [MB02]

## Literaturverzeichnis

- [AL06] Alby, Tom: Web 2.0 Konzepte, Anwendungen, Technologie. Hanser-Verlag, 2006
- [HA05] Haas, Hugo: Reconciling Web Services and REST Services. W3 Tech Talk, 2005  
<http://www.w3.org/2005/Talks/1115-hh-k-ecows/>
- [KL06] Klobas, Jane: Wikis: Tools for Information Work and Collaboration. Chandos Publishing, Oxford, 2006
- [MB02] Means W., Bodie M.: The Book of SAX: The Simple API for XML. No Starch Press, 2002
- [SN05] Snell, James: An overview of the Atom 1.0 Syndication Format. IBM Developer Works, 2005  
<http://www.ibm.com/developerworks/xml/library/x-atom10.html>
- [SU99] Sun Microsystems, Inc.: Simplified Guide to the Java 2 Platform, Enterprise Edition. Sun Microsystems, Inc, 1999  
[http://java.sun.com/j2ee/reference/whitepapers/j2ee\\_guide.pdf](http://java.sun.com/j2ee/reference/whitepapers/j2ee_guide.pdf)
- [WI07] Wikipedia Foundation: Model View Controller, Wikipedia Foundation, 2007

# Technologiebasierte Performanceunterschiede in serviceorientierten Architekturen

Sascha Roth, B.Sc. cand.  
Roth.Sascha@gmx.net

Prof. Dr. Katja Lenz  
K.Lenz@fbi.h-da.de

Hochschule Darmstadt  
— Fachbereich Informatik —

**Abstract:** Dieser Artikel stellt die Ergebnisse einer Performanceanalyse unterschiedlicher SOA-Technologien vor. Es wird gezeigt, dass durch die Verwendung XML basierter Protokolle signifikante Performancenachteile entstehen. Zur besseren Visualisierung der Performance wird der *Business Process Performance Graph* eingeführt.

## 1 Einleitung

Eine serviceorientierte Architektur (SOA) ist als Software Architektur unabhängig von jeglicher Technologie (vgl. [BKM07, S. 4]). Allerdings ist durch Marketing Strategien der IT Hersteller der Eindruck entstanden, dass durch die Nutzung von Technologien wie Web Services und der Business Process Execution Language (BPEL) automatisch eine SOA mit all ihren Vorteilen entsteht.

Für eine Performanceanalyse wurden Prototypen mit den Technologien ActiveBPEL Enterprise [Rot07, S. 26ff], ActiveBPEL Open Source [Rot07, S. 31], Apache Axis2 [Rot07, S. 32ff] und Java RMI [Rot07, S. 34ff] nach Vorgabe eines Praxisbeispiels<sup>1</sup> entwickelt, um ein Benchmarking durchzuführen.

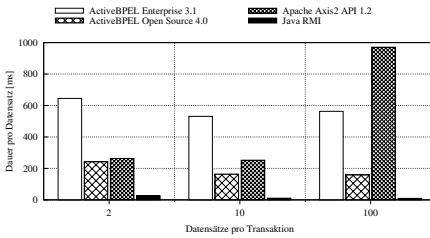
Um die Resultate der Performanceanalyse der vier Prototypen vorzustellen, sind einige kurze Definitionen notwendig: Der Begriff SOA wird im Rahmen dieses Artikels definiert als Software Architektur, welche unter Ausführung von black-box Services Geschäftsprozesse abbildet (vgl. [HBBK07, Kap. 2]). Geschäftsprozesse stehen in einer SOA wiederum als Service *höheren Levels* zur Verfügung. Unter *loser Kopplung* versteht man die Fähigkeit einen Service innerhalb einer SOA ändern zu können, ohne abhängige Prozesse zu beeinflussen.

Der übermäßige Einsatz des Prinzips von loser Kopplung beeinträchtigt die Performance negativ [HHV06, S. 9].

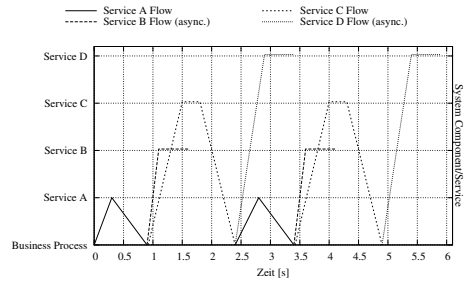
---

<sup>1</sup>Das Praxisbeispiel stammt von der Firma *Software AG*





(a) Zusammenfassende Messergebnisse



(b) Business Process Performance Graph

Abbildung 1: Grafische Auswertung des Geschäftsprozesses

## 2 Performanceanalyse einer serviceorientierten Architektur

Ein BPEL Geschäftsprozess kann durch jede Technologie, die eine Client/Server Architektur ermöglicht, mittels Kapselung ersetzt werden. Bei technischer Betrachtung stellt jeder Service eine Server Schnittstelle zur Verfügung – der Geschäftsprozess selbst stellt analog zu BPEL wiederum einen neuen Service dar. Die von der Firma ActiveEndpoints entwickelten ActiveBPEL Enterprise und ActiveBPEL Open Source Engines verwenden die Apache Axis API. Ein Nachbau des Geschäftsprozesses mit der *neueren und schnelleren* [Han07, S. 503],[Apa07a] Axis2 API sollte Performancevorteile bieten. Erhebliche Performancevorteile sollte auch die Java RMI Technologie bringen, da diese das binäre Java Remote Method Protocol (JRMP) verwendet. Services wurden in Form von Java Remote Objekten realisiert. Die Verwendung von Java Interfaces stellt eine lose Kopplung sicher<sup>2</sup>.

Um in den verteilten Systemen eine möglichst genaue Messung durchzuführen wurden Multicast UDP Pakete zur manuellen Zeit-Synchronisierung versendet. Mit Hilfe statistischer Methoden konnte ein max. Fehler von 3 ms bestimmt werden [Rot07, S. 42ff]. Für die Auswertung der Messergebnisse wurde der *Business Process Performance Graph* entwickelt. Eine detaillierte Beschreibung der Erstellung und des Nutzens dieses Graphen übersteigt den Umfang dieses Artikels. Jedoch erhält der Leser durch Abb.1(b) einen Eindruck des Nutzens mit zwei synchronen als auch asynchronen Services. Detaillierte Informationen sind in [Rot07, S. 49ff] zu finden.

Abb.1(a) zeigt eine Zusammenfassung der Messergebnisse. Die beiden BPEL Prototypen weisen ein relativ konstantes Verhalten auf. Die durchschnittlich höheren Laufzeiten der ActiveBPEL Enterprise Engine sind durch einen zusätzlichen Datenbank Server, welcher den Geschäftsprozess beinhaltet, zu erklären.

Bei dem Axis2 Prototypen wächst die Laufzeit eines einzelnen Datensatzes überproportional mit der Anzahl der Datensätze pro Transaktion. Diese Ergebnisse sind kritisch zu betrachten: Eine BPEL Engine benötigt i.d.R. nur begrenzte Informationen der XML Nutzdaten um den Geschäftsprozess fortsetzen zu können. Die Engine verwendet hierfür XPath

<sup>2</sup>Auf eine Versionierung und Protokollierung des Geschäftsprozess-Ablaufs wurde verzichtet.

Ausdrücke. Bei dem Axis2 Prototypen<sup>3</sup> wurden die vollständigen Nutzdaten zuerst auf Java Datentypen gemappt um anschließend in der Hochsprache weitere Anweisungen bzgl. des Geschäftsprozesses geben zu können. Auch bei Verwendung einer BPEL Engine innerhalb einer SOA findet häufig ein Mapping der XML Datentypen auf eine Hochsprache statt — üblicherweise in einem Endpoint-Service. Unerwartete, zusätzliche Flaschenhälse innerhalb einer SOA werden durch den XML Parsing Prozess daher begünstigt.

Der Java RMI Prototyp verhält sich sowohl konstant, als auch bei größeren Datenmengen um bis zu 100 fach schneller als die Web Service basierten Konkurrenten.

Eine abschließende Messung des Netzverkehrs konnte zeigen, dass die Laufzeiten der Web Service basierten Prototypen bei nur 2,3 bis 2,5 fachen Netzverkehr entstehen. Das lässt darauf schließen, dass durch den XML Parsing Prozess zusätzliche CPU Zeit und Hauptspeicher verbraucht werden.

Der hier gezeigte Effekt tritt in anderen Architekturen womöglich nicht in diesem Ausmaß auf — durch Dummy-Endpoint-Services wurde der Effekt zusätzlich begünstigt. Jedoch zeigen die Ergebnisse, dass durch die Verwendung von XML basierten Protokollen letztlich mit höheren Laufzeiten bei gleichzeitig höherem Verbrauch von Ressourcen zu rechnen ist.

### 3 Fazit und Ausblick

Bestandteil einer Migration zur SOA für ein Unternehmen ist die Auswahl einer geeigneten Technologie zur Designzeit. Ein Qualitätsattribut einer SOA ist ihre Performance (vgl. [OMB07]). Die Auswahl der einzusetzenden Technologie um eine SOA zu implementieren hängt von vielen weiteren Faktoren sowie politischen Entscheidungen ab. Nicht zuletzt da SOAs im Allgemeinen über Firmengrenzen hinweg skaliert werden [BKM07, p. 61].

Dieser Artikel hat gezeigt, dass zwischen 2,3 bis 2,5 fache Netzlast durch Einsatz XML basierter Protokolle entsteht. Dadurch fallen auch in heutigen Netzwerken Kosten an. Zusätzlich verursacht der XML Parsing Prozess einen überproportionalen Anstieg der Ausführungsdauer im Verhältnis zu den Nutzdaten.

Ein verwandter Artikel nennt zwischen 10 bis 20 fache Nachrichtengrößen und spricht die Performancenachteile des Einsatzes von XML Standards an [OMB07, S. 2]. Darüber hinaus haben Performanceanalysen gezeigt, dass Web Services im Vergleich mit CORBA und RAW Sockets ein Vielfaches an Zeit benötigen [OO02].

Die BPEL Standards sehen bisher keine Integration von bestehenden binären Protokollen vor. Insbesondere bei einer Intra-Business Kommunikation scheint dies aus Performance-sicht durchaus sinnvoll. Im Rahmen performanter Service Orchestrierung ist eine Erweiterung von BPEL zur Integration von *nicht* XML Standards, insbesondere binärer Protokolle wie CORBA(RMI-IIOP) und JRMP denkbar. Dadurch können ggf. Web Service Wrapper in Altsystemen entfallen, was zusätzliche Performancevorteile verspricht.

---

<sup>3</sup>In dieser Arbeit wurde POJO Modus (siehe [Apa07b]) der Axis2 API verwendet.

## Literatur

- [Apa07a] APACHE SOFTWARE FOUNDATION (Hrsg.): *Apache Axis2*. USA: Apache Software Foundation, 2007. <http://ws.apache.org/axis2/>, Abruf: 20.12.2007
- [Apa07b] APACHE SOFTWARE FOUNDATION (Hrsg.): *POJO Web Services using Apache Axis2*. USA: Apache Software Foundation, 2007. [http://ws.apache.org/axis2/1\\_2/pojoguide.html](http://ws.apache.org/axis2/1_2/pojoguide.html), Abruf: 20.12.2007
- [BKM07] BIANCO, Phil ; KOTERMANSKI, Rick ; MERSON, Paulo: Evaluating a Service-Oriented Architecture / Carnegie Mellon University (Software Engineering Institute). 2007 (15). – Forschungsbericht
- [Han07] HANSEN, Mark D.: *SOA Using Java<sup>TM</sup> Web Services*. Prentice Hall Professional, 2007
- [HBBK07] HURWITZ, Judith ; BLOOR, Robin ; BAROUDI, Carol ; KAUFMAN, Marcia ; LEVENSQUE, Paul (Hrsg.): *Service Oriented Architecture For Dummies<sup>®</sup>*. Indiana, USA : Wiley Publishing, Inc., 2007
- [HHV06] HESS, Andreas ; HUMM, Bernhard ; VOSS, Markus: Regeln für serviceorientierte Architekturen hoher Qualität. In: *Informatik Spektrum* 6 (2006), Dezember, S. 2–14
- [OMB07] O'BRIEN, Liam ; MERSON, Paulo ; BASS, Len: Quality Attributes for Service-Oriented Architectures. In: *SDSOA '07: Proceedings of the International Workshop on Systems Development in SOA Environments*. Washington, DC, USA : IEEE Computer Society, 2007. – ISBN 0-7695-2960-7, S. 3
- [OO02] OLSON, Mike ; OGBUJI, Uche ; IBM INC. (Hrsg.): *The Python Web service developer: Messaging technologies compared*. IBM Inc., Juli 2002. <http://www-128.ibm.com/developerworks/library/ws-pyth9/>, Abruf: 20.12.2007
- [Rot07] ROTH, Sascha: *Benchmarking of a Service-Oriented Architecture*, Hochschule Darmstadt (Fachbereich Informatik), Bachelorarbeit, 2007

# Erweitertes Hyperblock-Scheduling für VLIW-Prozessoren

– Kurzfassung der Diplomarbeit –

Dipl.-Inform. Dirk Tetzlaff

Institut für Softwaretechnik und Theoretische Informatik,

Fachgebiet Programmierung eingebetteter Systeme,

Technische Universität Berlin, FR 5-6, Franklinstr. 28/29, 10587 Berlin

<http://pes.cs.tu-berlin.de>    [tetzlaff@cs.tu-berlin.de](mailto:tetzlaff@cs.tu-berlin.de)

Betreuer:

Prof. Dr. Sabine Glesner

Dipl.-Inform. Lars Gesellensetter

## 1 Einleitung

Neu entstehende Prozessor-Architekturen werden stetig komplexer, um Programme effizienter abarbeiten zu können, und erfordern somit immer ausgefeiltere Strategien bei der Übersetzung von Programmen. Dies gilt insbesondere für VLIW-Prozessoren (*Very Long Instruction Word*) mit parallel arbeitenden Funktionseinheiten, die in letzter Zeit wieder verstärkt Anwendung finden. Bei diesen Prozessoren ist der Compiler für die Effizienz des erzeugten Maschinencodes verantwortlich, denn er entscheidet bei der Übersetzung, welche Befehle parallelisierbar sind. Dafür muss er vorhersagen, wie sich das Programm zur Laufzeit verhält. Obwohl diese Herausforderung seit langem bekannt ist, erzeugen auch aktuelle Compiler immer noch keinen effizienten Maschinencode. Experimente mit den bekannten SPEC CPU2006 Benchmarks auf Intels Itanium Prozessor zeigen, dass bestenfalls zwei Befehle pro Takt ausgeführt werden ([KHD<sup>+</sup>07]). Hauptursache ist die geringe Auswahl an Befehlen, die der Scheduler zur Verfügung hat, da die Grundblöcke der Zwischendarstellung ohne weitere Optimierung im Durchschnitt nur 9 Befehle enthalten (vgl. Tabelle 1). Bei kontrollflussdominierten Programmen können schwer vorhersagbare Verzweigungen auch zu vielen *Branch Mispredictions* führen, weshalb die Pipelines dieser Prozessoren die Arbeit dann unterbrechen müssen (engl.: *stall*).

In meiner Diplomarbeit habe ich ein Verfahren entwickelt, um eine höhere Auslastung der Funktionseinheiten von VLIW-Prozessoren zu erreichen und somit eine effizientere Abarbeitung der Befehle zu ermöglichen. Ich habe das Verfahren implementiert und konnte anhand von Experimenten mit den SPEC CPU2006 Benchmarks zeigen, dass durch mein Verfahren die Laufzeit der Programme deutlich verkürzt wird.

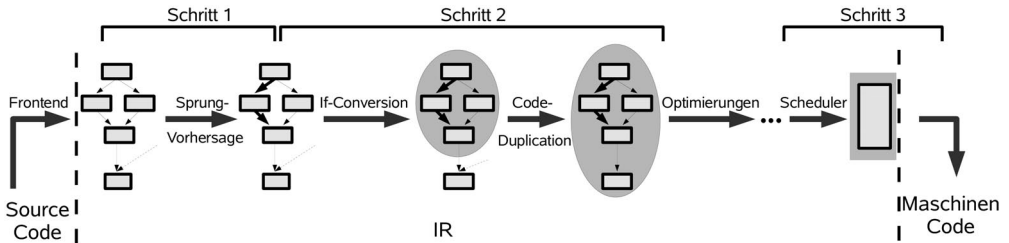


Abbildung 1: Erzeugung von erweiterten Hyperblöcken

## 2 Durchführung

In diesem Abschnitt erläutere ich mein Verfahren, dass aus folgenden drei Schritten besteht (siehe Abbildung 1):

1. Statische Sprungvorhersage
2. Vergrößerung der Grundblöcke
3. Scheduling innerhalb des vergrößerten Kontextes

Im ersten Schritt führe ich zunächst statisch eine Sprungvorhersage durch und berechne daraus die Ausführungs-Frequenzen der Grundblöcke und der Zweige des Kontrollfluss-Graphen basierend auf dem Ansatz von Wu und Larus [WL94]. Den Algorithmus habe ich verbessert, indem ich vorhandene Heuristiken weiterentwickelt und eine neue Heuristik entwickelt habe, wodurch genauere Sprungvorhersagen erzielbar sind.

Im zweiten Schritt werden zuerst durch iterative Anwendung von *If-Conversion*<sup>1</sup> Grundblöcke zu Hyperblöcken ([MLC<sup>+</sup>92]) vergrößert. [Mah96] zeigt, dass die Performanz von VLIW-Prozessoren mit dieser Technik besser ist als mit anderen Scheduling-Verfahren<sup>2</sup>. Choi et al. [CKGN01] betonen, dass *If-Conversion* gezielt angewendet werden muss, da es sonst negative Auswirkungen auf das Laufzeitverhalten hat. Daher begrenze ich die maximale Größe des Then- und des Else-Blockes und auch die Differenz zwischen beiden Größen. Außerdem wende ich *If-Conversion* nur bei schwer vorhersagbaren Verzweigungen an und begrenze die Anzahl an kopierten Befehlen, um steigende *Instruction-Cache-Stalls* zu verhindern. Die Relation zwischen den verwendeten Prädikaten in einem Hyperblock wird in einer von mir entworfenen, effizienten Baumstruktur festgehalten. Mit dieser ist eine optimale Register-Allokation durchführbar und die Entscheidung, ob zwei Prädikate disjunkt sind (d.h. sich gegenseitig ausschließen), kann mit linearem Aufwand getroffen werden. Anschließend werden Grundblöcke, deren Anzahl an Befehlen eine gesetzte Grenze unterschreitet, durch *Code Duplication* vergrößert.

Beim Scheduling im dritten Schritt meines Verfahrens wird zuerst ein Datenabhängigkeitsgraph unter Beachtung der Prädikat-Relation erzeugt (d.h. zwischen Befehlen disjunkter Prädikate wird keine Kante erstellt). Anschließend werden die Befehle mittels *List Scheduling* den Befehlsworten zugeordnet. Die Prioritätsfunktion zur Festlegung der Reihenfolge der Befehle sorgt u.a. dafür, dass Befehle mit höherer Frequenz früher eingeplant werden, so dass innerhalb derer Latenz Befehle mit geringerer Frequenz eingeplant werden können.

<sup>1</sup>If-Conversion überführt Kontrollabhängigkeit bei *If-Then-Else*-Konstrukten in Datenabhängigkeit mithilfe spezieller Prädikatregister.

<sup>2</sup>Neuere Arbeiten ([WWK<sup>+</sup>01], [QPG06], [KMSP06]) zeigen, dass *If-Conversion* auch die Performanz von Out-Of-Order Prozessoren steigern kann.

Bench	Befehle/Block			Stat. IPC		Bin (%)	Perf (%)	IPC		Sprünge/		Squash/		Stalls/Takt (%)					
	-O4	If-C	Dupl	-O4	Opt			o. Stalls		Bef. (%)		Bef. (%)		Total	I-Cach	Br	Mispr		
perlb.	5.35	6.01	8.32	1.63	1.67	8.3	0.8	1.69	1.68	17.5	14.6	0.7	3.7	33.1	34.1	6.3	7.0	6.3	6.0
bzip2	9.88	11.35	13.93	1.78	1.82	5.4	3.2	1.60	1.61	9.8	9.5	0.0	1.2	32.8	31.7	0.2	0.2	4.8	4.3
gcc	5.54	6.23	7.77	1.73	1.76	4.5	0.5	1.59	1.58	19.4	18.3	1.3	2.8	49.7	49.8	4.8	5.3	9.2	8.8
mcf	6.91	8.41	10.14	1.66	1.70	-1.4	0.5	1.32	1.39	14.2	13.3	0.1	2.0	88.5	89.2	0.0	0.0	0.9	0.8
gobmk	6.86	8.21	10.04	1.85	1.85	0.9	-0.2	1.50	1.51	13.0	11.8	0.8	2.1	42.0	43.1	7.1	8.4	9.7	9.4
hmmer	8.57	9.78	12.88	1.75	1.79	10.7	11.4	1.60	1.55	8.7	1.7	0.0	8.7	16.4	9.7	0.1	0.0	9.0	0.5
sjeng	7.48	9.01	11.31	1.64	1.68	4.5	0.8	1.53	1.51	13.8	12.7	0.2	2.3	28.7	28.6	5.2	5.9	13.8	12.7
h264ref	10.95	15.97	22.40	1.76	1.82	6.7	3.2	2.02	2.04	5.9	4.8	0.3	1.2	21.7	20.7	0.8	0.9	4.3	2.8
milc	11.82	13.35	17.19	2.12	2.16	7.1	0.7	1.91	1.92	2.0	2.0	0.0	0.1	79.6	79.4	0.0	0.0	0.1	0.0
lbm	21.38	25.74	35.18	2.67	2.87	13.5	1.8	1.95	1.99	0.9	0.5	0.2	0.7	69.2	69.3	0.1	0.1	0.1	0.1
sphinx	9.42	10.47	13.22	1.90	1.94	8.4	0.9	2.01	2.00	6.1	5.2	0.4	2.0	62.4	62.1	0.3	0.2	1.7	1.3
Ø	<b>9.47</b>	<b>11.32</b>	<b>14.76</b>	<b>1.86</b>	<b>1.92</b>	<b>6.2</b>	<b>2.1</b>	<b>1.70</b>	<b>1.71</b>	<b>10.1</b>	<b>8.6</b>	<b>0.4</b>	<b>2.4</b>	<b>47.6</b>	<b>47.1</b>	<b>2.3</b>	<b>2.5</b>	<b>5.4</b>	<b>4.2</b>

Tabelle 1: Statische Auswirkungen und Laufzeit-Ergebnisse ohne/mit Optimierung<sup>3</sup>

### 3 Ergebnisse

Ich habe mein Verfahren in einer am Fachgebiet vorhandenen Compiler-Infrastruktur<sup>4</sup> implementiert und anhand von Experimenten mit 11 SPEC CPU2006 Benchmarks die statischen und dynamischen Auswirkungen analysiert (siehe Tabelle 1). Als Referenz dienen die Werte für die **-O4**-Einstellung des CoSy Compilers, die über 50 Standard-Optimierungen umfasst. Die Laufzeit-Ergebnisse sind jeweils der Median von drei Durchläufen mit dem *train*-Datensatz. Durch If-Conversion wird die Block-Größe von durchschnittlich 9.47 Befehlen auf 11.32 erhöht. Nach dem Vergrößern durch Code Duplication werden 14.76 Befehle erreicht. Die moderate Erhöhung der statischen Anzahl von Befehlen pro Takt (*Instructions per Cycle, IPC*) im Maschinencode von 1.86 auf 1.92 ist durch den hohen Grad an Datenabhängigkeiten zwischen den Befehlen zu erklären. Durch die Integration einer präziseren Alias-Analyse erwarte ich eine deutlichere Steigerung der Parallelität. Verglichen mit der Vergrößerung der Grundblöcke um durchschnittlich 56% steigt die Code-Größe der Binärdateien nur gering (6.2%). Die Performanz der Benchmarks zur Laufzeit wird im Mittel durch mein Verfahren um 2.1% verbessert. Herausragend dabei ist die Laufzeit-Verkürzung von *hmmer* um über 10%. Die erreichte dynamische Parallelität (IPC ohne Stalls) konnte bei *mcf* am stärksten vergrößert werden, aber durch den hohen Anteil an *Total Stalls* steigt die Performanz nur marginal. Durch die Anwendung meines Verfahrens wird der Anteil an Sprungbefehlen von 10.1% auf 8.6% gesenkt und der Anteil an Stalls, die durch Branch Mispredictions verursacht werden, sinkt von 5.4% auf 4.2% (d.h. um 22%). Die geringen Auswirkungen auf die I-Cache-Stalls zeigen, dass die Kosten-Abschätzungen während der Vergrößerung der Grundblöcke gut gewählt wurden. Der mögliche negative Effekt von If-Conversion durch nicht ausgeführte Befehle (*squashed*), die aber Funktionseinheiten belegen, konnte begrenzt werden. Bei dem einzigen Benchmark mit einer Laufzeit-Verschlechterung, *gobmk*, stieg der Anteil an I-Cache-Stalls am stärksten und Branch-Misprediction-Stalls konnten nicht wesentlich verringert werden.

Abschließend lässt sich festhalten, dass durch das entwickelte Verfahren alle Benchmarks mit einer Ausnahme erheblich profitiert haben. Durch die Vergrößerung der Grundblöcke wird die Anzahl ausgeführter Sprünge und die Anzahl an Branch Mispredictions reduziert, ohne zu einer Zunahme von I-Cache-Stalls zu führen. Es hat sich gezeigt, dass die Heuristiken das dynamische Sprungverhalten präzise geschätzt haben. Mit Hilfe meines Verfahrens lässt sich somit das Potential von modernen VLIW-Prozessoren besser ausnutzen.

<sup>3</sup>Total Stalls beinhalten neben den gezeigten Werten auch Memory Stalls.

<sup>4</sup>Basierend auf dem CoSy<sup>®</sup> Compiler Development System von ACE [ACE07].

## Literatur

- [ACE07] Associated Compiler Experts bv., Amsterdam, The Netherlands. 2007. <http://www.ace.nl>.
- [CKGN01] Youngsoo Choi, Allan Knies, Luke Gerke und Tin-Fook Ngai. The Impact of If-Conversion and Branch Prediction on Program Execution on the Intel® Itanium™ Processor. In *MICRO 34: Proceedings of the 34<sup>th</sup> Annual International Symposium on Microarchitecture*, Seiten 182–191, Washington, DC, USA, 2001. IEEE Computer Society.
- [KHD<sup>+</sup>07] Arun Kejariwal, Gerolf F. Hölflöhner, Darshan Desai, Daniel M. Lavery, Alexandru Nicolau und Alexander V. Veidenbaum. Comparative Characterization of SPEC CPU2000 and CPU2006 on Itanium™ Architecture. In *SIGMETRICS '07: Proceedings of the 2007 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems*, Seiten 361–362, New York, NY, USA, 2007. ACM Press.
- [KMSP06] Hyesoon Kim, Onur Mutlu, Jared Stark und Yale N. Patt. Wish Branches: Enabling Adaptive and Aggressive Predicated Execution. *MICRO TOP PICKS: Micro's Top Picks from Microarchitecture Conferences, Special Issue*, 26(1):48–58, 2006.
- [Mah96] Scott A. Mahlke. *Exploiting Instruction-Level Parallelism in the Presence of Conditional Branches*. Doktorarbeit, Department of Electrical and Computer Engineering, University of Illinois, Urbana, IL, USA, 1996.
- [MLC<sup>+</sup>92] Scott A. Mahlke, David C. Lin, William Y. Chen, Richard E. Hank und Roger A. Bringmann. Effective Compiler Support for Predicated Execution Using the Hyperblock. In *MICRO 25: Proceedings of the 25<sup>th</sup> Annual International Symposium on Microarchitecture*, Seiten 45–54, Los Alamitos, CA, USA, 1992. IEEE Computer Society Press.
- [QPG06] Eduardo Quiñones, Joan-Manuel Parcerisa und Antonio Gonzalez. Selective Predicate Prediction for Out-Of-Order Processors. In *ICS '06: Proceedings of the 20<sup>th</sup> Annual International Conference on Supercomputing*, Seiten 46–54, New York, NY, USA, 2006. ACM Press.
- [WL94] Youfeng Wu und James R. Larus. Static Branch Frequency and Program Profile Analysis. In *MICRO 27: Proceedings of the 27<sup>th</sup> Annual International Symposium on Microarchitecture*, Seiten 1–11, New York, NY, USA, 1994. ACM Press.
- [WWK<sup>+</sup>01] Perry H. Wang, Hong Wang, Ralph M. Kling, Kalpana Ramakrishnan und John P. Shen. Register Renaming and Scheduling for Dynamic Execution of Predicated Code. In *HPCA '01: Proceedings of the 7<sup>th</sup> International Symposium on High-Performance Computer Architecture*, Seiten 15–26, Washington, DC, USA, 2001. IEEE Computer Society.

# Maschinelles Lernen der semantischen Auszeichnung medizinischer Textdokumente

Elisabeth Cremers

Rheinische Friedrich-Wilhelms-Universität Bonn  
Institut für Informatik  
cremerse@cs.uni-bonn.de

**Abstract:** Bei der Realisierung des Semantic Web geht es darum, die Inhalte des herkömmlichen Webs um Informationen auf Ebene der Bedeutung (Semantik) anzureichern, die für Maschinen lesbar sind. Das Ziel, dadurch autonome Systeme miteinander agieren lassen zu können, ist von hoher wirtschaftlicher Bedeutung. Das Gesundheitssystem befindet sich in einer Phase der Wandlung hin zu einer Struktur von Versorgungsnetzen aus Krankenhäusern und Arztpraxen und die Kommunikation von Informationssystemen innerhalb und zwischen diesen Institutionen ist ein relevantes Thema [HGE07]. Die semantische Heterogenität der einzelnen Systeme muss dabei überwunden werden, um Kommunikation und Qualität langfristig zu sichern und Kosten einsparen zu können.

## 1 Domäne eHealth

EHealth hat zum Ziel, die Akteure im Gesundheitswesen digital zu vernetzen und die Informations- und Kommunikationsmöglichkeiten für individuelle Gesundheitsfragen zu nutzen. Mit der Einführung von eHealth Lösungen soll die Herausforderung der Finanzierung des Gesundheitssystems angegangen werden, um langfristig das Gesundheitssystem auf hohem Niveau zu sichern. Ein aktuelles Thema aus dem Bereich eHealth ist die Einführung der elektronischen Gesundheitskarte in Kombination mit dem elektronischen Rezept (eRezept) und der Arzneimitteldokumentation. Mit dem eRezept erhofft man sich durch verbesserte automatische Abrechnungsprüfungen bei den Krankenkassen ohne Mehraufwand Einsparungen von über 70 Millionen Euro.

Im freiwilligen Teil der elektronischen Gesundheitskarte (wahlfrei für den Inhaber) soll der elektronische Arztbrief (eArztbrief) realisiert werden. Arztbriefe werden in Praxen und Krankenhäusern täglich in großen Mengen verfasst und versendet. Sie sind ein wichtiges Kommunikationsmedium zwischen Arzt und Patient, aber vor allem auch zwischen Ärzten untereinander. Auf Grundlage eines Arztbriefes werden häufig bei Mit- und/oder Weiterbehandlung eines Patienten durch einen anderen Arzt wichtige Entscheidungen getroffen. Für Ärzte liegt die Priorität grundsätzlich nicht in der Datenerfassung und -bearbeitung, sondern in der medizinischen Tätigkeit. Deshalb sollten Ärzte an dieser Stelle effektiv durch ein Informationssystem unterstützt werden.



Eine Unterstützung des Arztes ist nicht nur bei der Erstellung eines Arztbriefes, sondern ebenso bei der Auswertung der Informationen, die über den Arztbrief transportiert werden, sinnvoll. Eine automatische inhaltliche Erschließung der Dokumente kann dann von besonders großer Bedeutung sein, wenn schnell über die medizinische Sachlage eine Übersicht gewonnen werden muss. Ein Notfall oder die ambulante Behandlung insbesondere chronisch kranker Menschen sind Beispiele dafür. Die benötigte Unterstützung kann durch semantische Interoperabilität gewährleistet werden.

## 2 NER

Semantische Interoperabilität kann erreicht werden auf dem Wege einer semantischen Auszeichnung medizinischer Textdokumente. Die *Named Entity Recognition* (NER) ist eine Teilaufgabe der *Informationsextraktion* (IE). Bei der IE geht es darum, aus unstrukturierten Dokumenten automatisch Information zu gewinnen und sie strukturiert zu repräsentieren. Ein NER-System hat dabei die Aufgabe, in einem unstrukturierten, in natürlicher Sprache verfassten Textdokument gewisse für den Benutzer relevante Bestandteile zu erkennen und diese einer zuvor definierten Kategorie zuzuordnen. Diese Bestandteile, so genannte *Named Entities* (NE), können Eigennamen, Zeitangaben oder numerische Ausdrücke sein und kontextbezogen definiert werden.

Die NER lässt sich einteilen in die Aufgabenbereiche *Named Entity Detection* (NED) und *Named Entity Classification* (NEC). Bei der NED geht es darum, Named Entities in einem natürlich-sprachlichen Text als für den Benutzer interessante Bestandteile zu entdecken. Die Bestandteile, also Wörter oder Wortsequenzen, werden dann in der NEC einer geeigneten Kategorie zugeordnet und entsprechend maschinenlesbar annotiert.

## 3 Hybrides NER-Verfahren für medizinische Textdokumente

Ein hybrides Verfahren für die NER nutzt gezielt die Vorteile mehrerer einzelner Verfahren. Es soll hier beispielhaft die Auswahl von Verfahren in Bezug auf eine gewählte NE-Kategorie für medizinische Textdokumente stattfinden.

NEs numerischer Kategorien tauchen in medizinischen Textdokumenten häufig auf und sind für den Arzt ein wichtiger Anhaltspunkt. Der Einsatz eines listenbasierten Verfahrens macht an dieser Stelle wenig Sinn, da sich eine explizite Speicherung der möglichen Zahlenkombinationen verbietet, wenn die allermeisten davon nicht benutzt werden. Ein befriedigendes Ergebnis ist dagegen damit zu erreichen, einen regulären Ausdruck innerhalb eines *regelbasierten Verfahrens* zu formulieren. Da die Angabe z.B. einer NE der Kategorie DATUM zumeist in einer syntaktisch einheitlichen Form erscheint, kann ein Muster dafür spezifiziert werden, nach dessen Vorkommen dann automatisch in Texten gesucht wird.

Einen wichtigen inhaltlichen Beitrag eines informativen Arztbriefes liefern NE der Kategorie DIAGNOSE. Entscheidend ist bei der Auswahl eines NER-Verfahrens für diese Kategorie die Erkenntnis, dass die Medizin bereits über ein breites Spektrum an Begriffs- und Ordnungssystemen verfügt. Grundbestandteil eines *listenbasierten Verfahrens* ist eine Wortdatenbank, in der alle Wörter und Wortsequenzen aufgelistet sind, die als NEs erkannt und klassifiziert werden sollen. Die in dem Text vorkommenden Wörter werden dann mit denen in der Wortdatenbank gelisteten Worten verglichen. Bei Übereinstimmung eines Wortes aus dem Text wird dieses genau der Kategorie zugeteilt, die in der Datenbank dasselbe Wort klassifiziert. Dem Nachteil des Verfahrens, der bei der aufwendigen Erstellung einer umfangreichen Wortdatenbank liegt, kann für die Kategorie DIAGNOSE entgegengewirkt werden, indem beispielsweise auf Snomed CT [C08] als Referenzontologie zurückgegriffen wird. Es sei darauf hingewiesen, dass mit Ontologien durch deren hohe semantische Reichhaltigkeit bessere Ergebnisse erzielt werden als mit einfachen Listen. Die Grundstruktur einer Ontologie besteht schließlich nicht nur aus Begriffen und einem Lexikon, sondern zusätzlich aus semantischen Relationen und regelhaften Zusammenhängen [ES06].

Eine wichtige Rolle im Rahmen der Übermittlung und Bearbeitung von medizinischen Textdokumenten spielt die Medikation eines Patienten. Ein weiterbehandelnder Arzt kann eine Therapie erst dann erfolgreich weiterführen, wenn er über die dem Patienten verabreichten Medikamente Bescheid weiß. Auch für die Kategorie MEDIKAMENT ist ein listenbasiertes Verfahren gut anzuwenden. Als Referenz kann die Rote bzw. Gelbe Liste als Arzneimittelverzeichnis zum Einsatz kommen. Zusätzlich soll hier aber auf die Möglichkeit eingegangen werden, ein *Maschinelles Lernverfahren* für die NER der Kategorie MEDIKAMENT anzuwenden.

Das Verfahren der Subgruppenentdeckung sucht in einem Textdokument Gruppen, die sich in wenigen Attributen von der Gesamtpopulation unterscheiden. Bereits bekannte Medikamente bilden bei Anwendung dieses Verfahrens die Gesamtpopulation. Bei der Neueinführung eines Medikaments kommt es häufig vor, dass lediglich die Dosierung geändert wurde oder aber ein Zusatz hinzugefügt wurde, der namentlich einen neuen Prä- oder Postfix des Medikamentennamens zur Folge hat. Die Subgruppenentdeckung kann an dieser Stelle erkennen, dass der Term im Grunde genommen einer Gruppe von Termen (den bereits bekannten Medikamenten) ähnlich ist, er sich aber in der Ausprägung des Zielattributs von diesen unterscheidet. Als Zielattribut versteht sich hier die Zuordnung, ob es sich um ein Medikament handelt oder nicht. Somit wird erreicht, dass Medikamente, die in einer Liste von Medikamenten (noch) nicht enthalten sind, dennoch als solche klassifiziert werden können.

## 4 Anwendungsszenario und Ausblick

Mit der Erreichung der semantischen Interoperabilität von medizinischen Informationssystemen ist eine Vielzahl von nützlichen Anwendungen denkbar. Das hybride NER-Verfahren kann dabei in verschiedene Szenarios eingebettet werden, von denen eines nun vorgestellt werden soll:

Ein weiterbehandelnder Arzt möchte sich einen schnellen Überblick über den Krankenhausaufenthalt seines neuen Patienten machen. Er möchte möglichst schnell wissen, welche Diagnose sein Vorgänger im Krankenhaus gestellt hat, welche Ergebnisse Labor- und Blutuntersuchungen des Patienten ergeben haben und welche Medikamente dieser in den letzten Tagen bekommen hat. Durch die (semi-)automatische semantische Auszeichnung des Arztbriefes ist es möglich, dass der Arzt sein System auffordert, alle Medikamente aufzulisten, die der Patient im Laufe seines Krankenhausaufenthalts bekommen hat. Diese sind seit der semantischen Auszeichnung des Textdokumentes in der Kategorie MEDIKAMENT gespeichert. Dazu könnte er beispielsweise die Information anfordern, an welches Datum die Medikation gebunden ist. Indem der Arzt dann die Medikation der Diagnose gegenüberstellt, kann er Entscheidungen darüber treffen, wie eine Weiterbehandlung aussehen kann. Tatsächlich ist es aber auch vorstellbar, das hybride NER-Verfahren so zu erweitern, dass ein Vorschlag für die Weiterbehandlung automatisch vom System gemacht werden kann. Dafür müsste das System lediglich auf bestehende Therapiepläne zugreifen können. Angenommen zu jedem Therapieplan wären diejenigen Patientendaten hinterlegt, die zu diesem Vorgehen geführt haben, so könnte das System genau den Therapieplan vorschlagen, dessen zusätzliche Informationen den vorliegenden Patientendaten am ähnlichsten sind. Je detaillierter die semantische Auszeichnung des Arztbriefes ist, desto mehr Patientendaten und desto genauere Informationen über die Zusammenhänge der Daten kann das System speichern. Das ist wiederum ausschlaggebend dafür, wie gut eine Übereinstimmung verschiedener Daten gemessen werden kann. Liegt beispielsweise ein Therapieplan vor, der auf den Daten basiert, dass ein Patient nach allergischer Reaktion auf ein bestimmtes Medikament ein anderes Medikament verabreicht bekam, spielt der Zusammenhang zwischen Medikamentauswahl und Allergie eine besonders wichtige Rolle. Ist hingegen keine Kategorie für die NE „Allergie“ hinterlegt, so ist das System nicht in der Lage, den Zusammenhang zu vermitteln, dass mit dem ersten Medikament eine Unverträglichkeit verbunden ist. Bei der automatischen Therapieplanerstellung kann das unerwünschte Folgen haben. Es ist wichtig, dass Fachexperten an der Definition von Kategorien für NEs beteiligt sind.

## Literaturverzeichnis

- [C08] College of American Pathologists: Snomed CT. [www.cap.org](http://www.cap.org), Abruf am 02-02-2008.
- [Dö01] Dörre, J. et al.: Volltextsuche und Text Mining. In: Carstensen, K.-U. et al. (Hrsg.): Computerlinguistik und Sprachtechnologie- Eine Einführung. Spektrum Akademischer Verlag, Heidelberg, 2001, S. 425-441.
- [ES06] *Ehrig, Marc; Studer, Rudi*: Wissensvernetzung durch Ontologien. In: Blumauer, A.; Pellegrini, T.(Hrsg.): Semantic Web- Wege zur vernetzten Wissensgesellschaft. Springer-Verlag, Heidelberg, 2006, S. 469-484.
- [Ha05] Haas, Peter: Medizinische Informationssysteme und elektronische Krankenakten. Springer Verlag, Heidelberg 2005.
- [HGE07] Holzinger, A.; Geierhofer, R.; Errath, M.: Semantische Informationsextraktion in medizinischen Informationssystemen. In: Informatik-Spektrum 30(2), S. 69-78. Springer, München, 2007.
- [SH05] Stuckenschmidt, Heiner; van Harmelen, Frank: Information Sharing on the Semantic Web. Springer-Verlag, Heidelberg, 2005.

# Gestaltung eines papierbasierten Tangible Interface zur Unterstützung kunsthistorischer Arbeitstechniken

Tanja Döring

tanja.doering@informatik.uni-hamburg.de

Art der Arbeit: Diplomarbeit im Fach Informatik an der Universität Hamburg  
Betreuer: Prof. Dr.-Ing. Steffi Beckhaus und Prof. Dr. Horst Oberquelle

**Abstract:** Grafische User Interfaces bieten nicht für alle Aufgaben optimale Unterstützung. Die beschränkte Bildschirmgröße, die Konzeption als Einzelarbeitsplatz und die fehlende Integration papier- und computerbasierten Arbeitens sind einige der Gründe für die Entwicklung neuer Schnittstellen. Zu diesen zählen Tangible User Interfaces, welche greifbare Gegenstände mit digitalen Daten und Funktionen verknüpfen und somit neue Möglichkeiten der Interaktion mit dem Computer und seiner Integration in die Umgebung des Menschen liefern. Die in diesem Beitrag vorgestellte Diplomarbeit behandelt die Gestaltung eines Tangible User Interface für den Anwendungskontext der Kunstgeschichte. In dem nach einer Methode des Participatory Design realisierten Projekt wurde eine Fallstudie über kunsthistorische Arbeitstechniken durchgeführt und auf dieser Basis ein Konzept für ein System ermittelt, das als tischbasiertes Tangible User Interface umgesetzt und im Rahmen eines Benutzungstests erprobt wurde.

## 1 Motivation und Ziel

Bei kreativen und schwach strukturierten Tätigkeiten verwenden viele Menschen gerne Papier, Stift und greifbare Gegenstände. Sie arbeiten räumlich und nutzen große Flächen zum Sammeln, Ordnen und Strukturieren. Viele dieser Arbeitstechniken lassen sich nur sehr bedingt in Systemen mit Grafischen User Interfaces (GUIs) umsetzen. Erschwerend kommt hinzu, dass Benutzer oft einen großen Teil ihrer Aufmerksamkeit der Bedienung von Applikationen widmen müssen, statt sich ganz auf ihre kreative Aufgabe konzentrieren zu können. Diese Bedingungen führen vielfach zu umständlichen Medienbrüchen wie etwa der parallelen Verwendung von papierbasierten und digitalen Versionen eines Dokumentes. Im betrachteten Anwendungskontext der Kunstgeschichte zeigt sich dieser Konflikt in der Arbeit mit Abbildungen. Trotz vieler Ansätze zur Computerunterstützung [KK03] werden traditionelle Arbeitstechniken wie etwa der Zettelkasten mit Bildkarten weiterhin mit großer Beliebtheit eingesetzt [Bru99].

Tangible User Interfaces (TUIs) [Hor04] ermöglichen eine bessere Integration von realen Gegenständen und digitalen Daten, indem greifbare Objekte mit entsprechenden Daten und Funktionen verbunden werden. Sie können eingesetzt werden, um Einschränkungen der GUI-Interaktion zu überwinden, ohne dabei auf eine Einbettung in vorhandene Com-

puteranwendungen zu verzichten (siehe z.B. [KNF<sup>+</sup>01]). Eine Reihe von TUI-Prototypen sind tischbasierte Systeme mit greifbaren Interaktionselementen (siehe z.B. [WHZ<sup>+</sup>06]).

Ziel des Projektes war es, ein innovatives tischbasiertes Tangible Interface für Kunsthistoriker zu entwickeln, welches traditionelle Arbeitstechniken aufgreift und möglichst nahtlos in computerunterstützte Arbeitsweisen integriert. Der Nutzer sollte sich auf die Arbeit mit Bildkarten konzentrieren können, ohne seine Hauptaufmerksamkeit auf das System lenken zu müssen.<sup>1</sup> Das TUI sollte frei und flexibel verwendbar sein, indem es individuelle und kreative Arbeitsweisen unterstützt. Als spezieller Anwendungskontext wurde der Bereich der ikonografischen Arbeitsweise betrachtet, in dem Bildmotive genau untersucht, miteinander verglichen und klassifiziert werden.

## 2 Durchführung und Ergebnisse

Im Rahmen des Diplomprojektes wurde eine Fallstudie zu kunsthistorischen Arbeitstechniken durchgeführt und auf dieser Basis ein Konzept für ein Tangible Interface zur Erstellung kunsthistorischer Bildertafeln entwickelt. Dieses wurde in Form eines tischbasierten Prototyps umgesetzt und im Rahmen eines Benutzungstests erprobt. Um den Bedürfnissen der Nutzer aus dem Bereich der Kunstgeschichte gerecht werden zu können, wurde für die Durchführung des Projektes eine Methode des *Participatory Design* angewendet [BKS04]. Am Projekt nahm eine Bildhistorikerin als potenzielle Anwenderin und Testperson teil. Benutzerbeteiligung wurde durch Techniken wie *In Situ Interview*, *Dokumentenanalyse*, *Paperprototyping* und *experimentelles Prototyping* durch den gesamten Gestaltungsprozess hindurch sichergestellt.

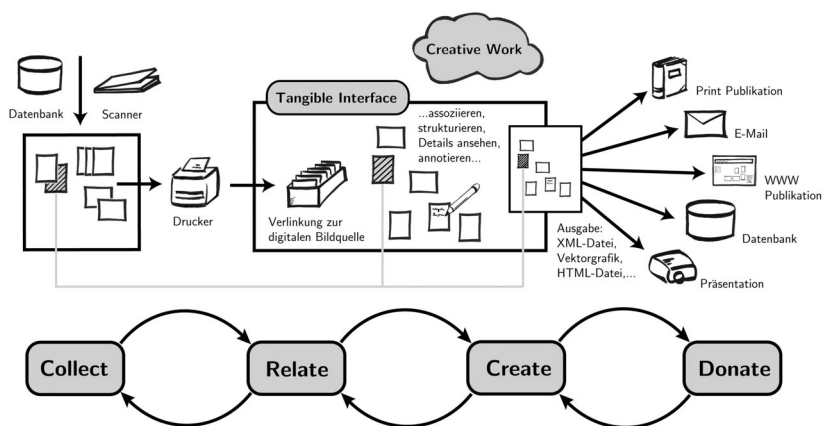


Abbildung 1: Integration des Tangible User Interface in den kreativen Arbeitsprozess

<sup>1</sup>Dies entspricht Mark Weisers Konzepten des *Calm* und *Ubiquitous Computing* [Wei91].

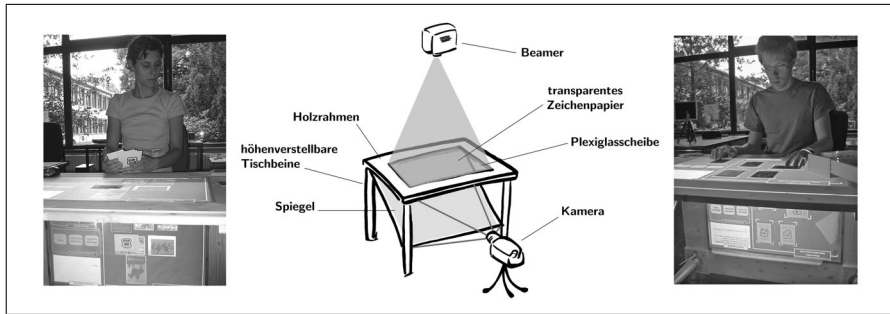


Abbildung 2: Der Aufbau des tischbasierten Tangible User Interface

Die Fallstudie über ikonografische Arbeitstechniken analysierte einzelne Arbeitsschritte. Dabei diente ein Modell über vier Stufen kreativer Arbeit (Collect-Relate-Create-Donate) [Shn02] zur Strukturierung. Die Studie ermittelte, dass die vorgelagerten Aktivitäten der Collect-Stufe (Sammlung und Anlage des Bildmaterials) gut durch vorhandene GUI-Applikationen unterstützt werden konnten. Die papierbasierten Bildkarten waren während der Relate- und Create-Aktivitäten (Strukturierung und Diskussion des Bildmaterials, Analyse und Herausarbeitung von Motivgruppen) das essentielle Arbeitsmittel für den kreativen Erkenntnisprozess. Für die Aktivitäten der Donate-Stufe (Erstellung von Publikationen und Vortragsfolien) waren wiederum vorhandene Computeranwendungen gut geeignet.

Zu den zentralen Potenzialen eines TUI für Kunsthistoriker zählen die Bewahrung und Erweiterung traditioneller, papierbasierter Arbeitsweisen, die Unterstützung von kreativer, freier Arbeit und die Integration in bestehende (GUI-)Anwendungen. Abb. 1 zeigt ein Schema der Integration des Tangible User Interface in den kreativen Arbeitsablauf. Hieraus wurde ein Konzept für ein System erstellt, welches auf der Relate- und Create-Stufe die freie Arbeitsweise mit Bildkarten aus Papier auf einer Tischfläche ermöglicht. Dazu zählt u.a. ihre Clusterung, Arrangierung und die Erstellung von Annotationen und Schlagworten zu den Bildmotiven. Das System beinhaltet Import- und Exportschnittstellen für die Integration in vor- und nachgelagerte Aktivitäten der Collect- und Donate-Stufe.

Das als funktionierender Prototyp umgesetzte System (siehe Abb. 2) konzentriert sich auf die Unterstützung der Relate-, Create- und Donate-Stufe. Die Interaktion erfolgt über eine semitransparente Tischfläche, auf der Bildkarten aus Papier platziert werden können. Sie sind rückseitig mit eindeutigen schwarz-weißen Mustern (Fiducials) versehen und können somit per Videokamera, die die Tischfläche von unten über einen Spiegel aufnimmt, identifiziert und lokalisiert werden. Als visuelle Trackingkomponente verwendet das System die ReacTIVision Vision Engine [KB07], die über das Protokoll MIDI<sup>2</sup> die Daten der erkannten Fiducials an die TUI-Anwendung weitergibt. Die grafische Ausgabe der in Macromedia Director<sup>3</sup> erstellten Applikation wird von oben auf den Tisch projiziert. Die Anwendung ordnet den Abbildungen auf Papierkarten zugehörige Bilddateien zu, so dass aus Zusammenstellungen von Bildkarten beispielsweise eine entsprechende Websei-

<sup>2</sup>Siehe „MIDI Manufacturers Association“ unter <http://www.midi.org> (Zugriff am 5.1.2008).

<sup>3</sup>Siehe <http://www.adobe.com/products/director/> (Zugriff am 5.1.2008).

te oder Vektorgrafik generiert werden kann. Eine erste formative Evaluation des Tangible Interface mit einer Bildwissenschaftlerin zeigte, dass sich ein papierbasiertes TUI für die wissenschaftliche Arbeit mit Bildkarten gut eignet. Die Nutzerin konnte frei mit ihren persönlichen Bildkarten arbeiten, ohne sich auf die Bedienung des Systems konzentrieren zu müssen. Sie bewertete die Kombination von traditioneller Arbeitstechnik und innovativer Anwendung als sehr nützlich und intuitiv bedienbar. Es zeigte sich, dass für den Einsatz des TUI seine nahtlose Integration in den kreativen Arbeitsablauf entscheidend ist.

### 3 Fazit

Dieser Beitrag hat die nach einem benutzerzentrierten Ansatz durchgeführte Konzeption und Entwicklung eines Tangible User Interface für Kunsthistoriker vorgestellt. Der erstellte Prototyp bietet durch innovative Interaktionsformen neue Möglichkeiten zur Unterstützung schwach strukturierter, kreativer Arbeitsprozesse. Die entstandene Tischkonstruktion und die TUI-Applikation eignen sich als Ausgangspunkt für die Entwicklung und Erprobung weiterer Anwendungen mit Tangible User Interfaces.

### Literatur

- [BKS04] Ken Bødker, Finn Kensing und Jesper Simonsen. *Participatory IT Design. Designing for Business and Workplace Realities*. MIT Press, Cambridge, USA, 2004.
- [Bru99] Matthias Bruhn. The Warburg Electronic Library in Hamburg: A Digital Index of Political Iconography. *Visual Resources*, XV:405–423, 1999.
- [Hor04] Eva Hornecker. *Tangible User Interfaces als kooperationsunterstützendes Medium*. Dissertation, Universität Bremen, 2004.
- [KB07] Martin Kaltenbrunner und Ross Bencina. reactIVision: A Computer-Vision Framework for Table-Based Tangible Interaction. In *Proceedings of TEI '07*, Seiten 69–74, New York, USA, 2007. ACM Press.
- [KK03] Hubertus Kohle und Katja Kwastek. *Computer, Kunst und Kunstgeschichte. Theorie und Praxis*. Deubner Verlag für Kunst, Köln, 2003.
- [KNF<sup>+</sup>01] Scott R. Klemmer, Mark W. Newman, Ryan Farrell, Mark Bilezikjian und James A. Landay. The Designers' Outpost: A Tangible Interface for Collaborative Web Site Design. In *Proceedings of UIST'01*, Seiten 1–10, New York, USA, 2001. ACM Press.
- [Shn02] Ben Shneiderman. *Leonardo's Laptop: Human Needs and the New Computing Technologies*. MIT Press, Cambridge, USA, 2002.
- [Wei91] Mark Weiser. The Computer for the 21st Century. *Scientific American*, 265(3):94–104, September 1991.
- [WHZ<sup>+</sup>06] Manuela Waldner, Jörg Hauber, Jürgen Zauner, Michael Haller und Mark Billinghurst. Tangible Tiles: Design and Evaluation of a Tangible User Interface in a Collaborative Tabletop Setup. In *Proceedings of OZCHI '06*, Seiten 151–158, New York, USA, 2006. ACM Press.

# VRLab – Interaktion in virtueller Realität

Tobias Eger, Simeon Eichel, Christoph Kluck,  
Pascal Liedtke, Bastian Rackow

Hochschule Reutlingen  
Medien- und Kommunikationsinformatik Master

Tobias\_Manuel.Eger@student.reutlingen-university.de  
Simeon.Eichel@student.reutlingen-university.de  
Christoph.Kluck@student.reutlingen-university.de  
Pascal.Liedtke@student.reutlingen-university.de  
Bastian.Rackow@student.reutlingen-university.de

Art der Arbeit: Masterprojekt im 1. und 2. Mastersemester

Betreuer/in der Arbeit: Prof. Dr. Gabriela Tullius, Prof. Dr. Uwe Kloos

Abstract: Virtual Reality, Interaktion und stereoskopische Projektion sind die Bereiche des VRLab der Hochschule Reutlingen. Diese werden in mehreren Teilprojekten bearbeitet, von denen hier die „Remote Control“, das „Low-cost Tracking System“ und das „3D-Modellhandbuch“ vorgestellt werden.

## 1 Das VRLab

Als Virtual Reality, kurz VR, bezeichnet man eine computergenerierte, echtzeitfähige und interaktive virtuelle Umgebung. VR wird dank steigender Rechnerkapazitäten immer interessanter für Wirtschaft und Forschung und schon in vielen Bereichen eingesetzt. So beispielsweise in der Medizin [AM07], Maschinenbau (CAD) und der Pilotenausbildung (Flugsimulator). Das Projekt VRLab im Studiengang „Medien- und Kommunikationsinformatik“ an der Hochschule Reutlingen beschäftigt sich mit Fragestellungen rund um dieses Themengebiet, mit den Schwerpunkten Darstellung, Mensch-Maschine-Interaktion, Tracking und Kommunikation.

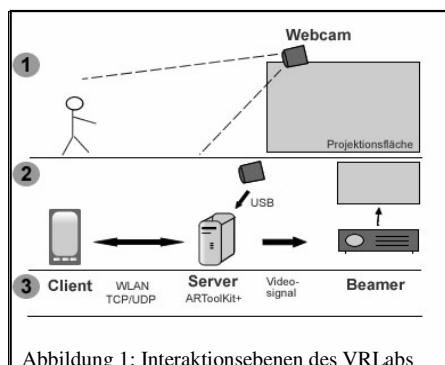


Abbildung 1: Interaktionsebenen des VRLabs

Das Labor beinhaltet ein fest installiertes stereoskopisches 3D-Projektionssystem (Powerwall), das mit Polarisation arbeitet und ein, auf Interferenz-Technik basierendes,

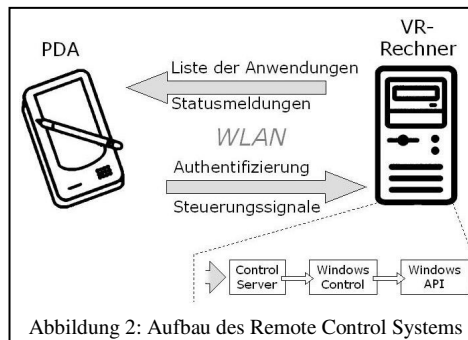


mobiles System [IT03]. Die einzelnen Projekte arbeiten auf drei Interaktions- und Kommunikationsebenen: Mensch-Maschine (1), Maschine-Maschine (2), und Protokoll-Protokoll (3), siehe Abb. 1. Ziel des VRLab-Projekts ist es den Funktionsumfang in Teilprojekten stetig auszubauen, die Darstellungsqualität zu optimieren und die Interaktionsmöglichkeiten zu erweitern.

## 2 Remote Control

Das Teilprojekt „Remote Control“ entstand als Arbeit im Bereich „Interaktion mit großen Anzeigen“. Es wurde eine Software entwickelt, die eine unhandliche, platzraubende und auch störanfällige Steuerung des Präsentationsrechners mittels drahtloser Tastatur und drahtloser Maus ersetzt. Als Lösung wird eine Steuerung per PDA vorgeschlagen.

Im aktuellen Stand ist es möglich, den Präsentationsrechner über WakeOnLan einzuschalten und fernzusteuern. Neben der Möglichkeit der Maus- und Tastatureingaben können Audioparameter bearbeitet, sowie vordefinierte Anwendungen gestartet werden. Vorteil ist, dass ein umständliches manuelles Starten und Beenden der Anwendungen entfällt. Ein Kernproblem, das bei der Interaktion mit großen Anzeigen auftritt, ist die Unübersichtlichkeit beim Navigieren. Dieses Problem wird somit gelöst.



Das System ist in C# implementiert und arbeitet nach dem Client- Server- Prinzip (siehe Abb. 2). Der Client ist als GUI auf dem PDA realisiert und empfängt per HTTP eine Liste der verfügbaren Anwendungen vom Server. Per TCP/IP sendet der Client Steuerbefehle für zeitunkritische Daten. Zeitkritische Daten, wie Mausbewegungen, werden per UDP an den Server übertragen. Dieser führt die Befehle über die Windows API aus.

Hierbei kommunizieren die beiden Geräte über ein privates, gesichertes WLAN. Eine Authentifizierung des Clients am Server findet statt, ist im momentanen Stand jedoch noch nicht verschlüsselt.

### 3 Low-cost Tracking System

Das Tracking von Bewegungen eines Benutzers ist eine häufig verwendete Technik, um die Interaktion mit virtuellen Welten zu ermöglichen. Professionelle Trackingsysteme sind jedoch sehr teuer. Hauptziel dieses Teilprojektes ist es eine low-cost Lösung zu entwickeln, die in der Lage ist, den Benutzer mit der virtuellen Welt interagieren zu lassen.

Da das System im VRLab meist im Stehen benutzt wird, ist es wünschenswert Interaktionstechniken anzubieten, die das Interagieren mit dem System im Stehen unterstützen. In einem ersten Schritt wird die Bildschirmmaus über Handbewegungen des Benutzers im Raum gesteuert. Dazu wird der Bereich in dem der Benutzer interagieren soll, mit einer handelsüblichen Webcam erfasst. Ein Marker in Form einer leuchtenden Infrarot-LED wird von dem System erkannt und seine Bewegungen in Mausbewegungen umgesetzt. Die Wahl der Bibliothek zur Verarbeitung der Bilddaten der Webcam fiel nach einer Evaluierung [NE07] auf das Framework Aforge.NET [AF07]. Aufgrund der beschränkten Auflösung der Webcam ist die Maussteuerung nicht pixelgenau.

Ein nächster Schritt ist die Entwicklung eines 3D-Trackingsystems, mit Hilfe zweier Webcams. Vor allem für Interaktionstechniken, die auf Head-Tracking z.B. zur Blicksteuerung oder auf Tracking von Eingabegeräten basieren, ist ein solches Trackingsystem notwendig. Beispiele für solche Techniken sind Aperture [FO96] und HOMER [BO97] zur Auswahl oder Manipulation von Objekten im virtuellen Raum. So kann zum Beispiel durch Drehen eines Eingabegerätes ein Objekt im Raum gedreht werden (siehe Abb. 3). Mit Hilfe der intrinsischen und extrinsischen Kamera-Parameter und den Koordinaten eines Punktes in zwei Kamerabildern, lässt sich die Position eines Punktes im Raum bestimmen [FA96], und somit 3D-Tracking realisieren.

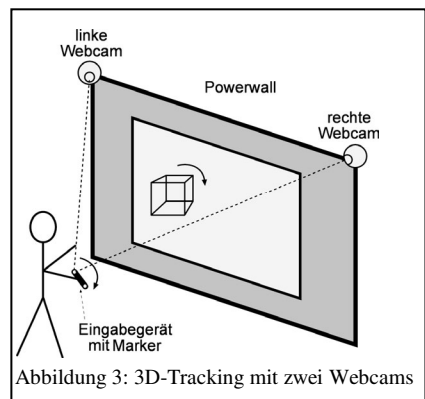


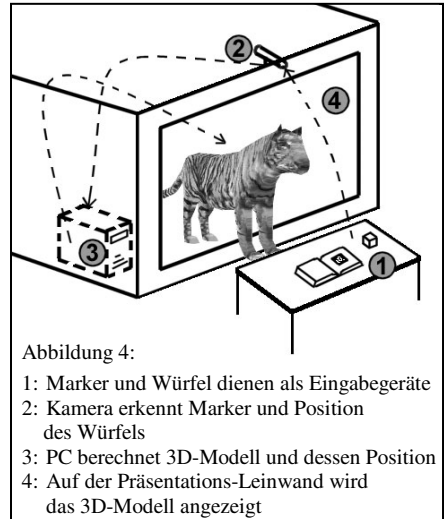
Abbildung 3: 3D-Tracking mit zwei Webcams

Um die Maussteuerung zu demonstrieren, wird das Spiel „Balloons“ entwickelt. Der Spielablauf ist einfach gehalten: Der Spieler muss aufsteigende Ballons mit einer virtuellen Nadel zerstechen und bekommt dafür Punkte. Er steuert die virtuelle Nadel durch die Bewegung eines Markers im Blickwinkel einer Webcam. Dafür steht entweder der oben erwähnte Infrarotmarker oder ein musterbasierter Marker zur Verfügung. Für die Erkennung des musterbasierten Markers auf dem Kamerabild wird eine Bibliothek für Augmented Reality („ARToolkitPlus“ [AR07]) verwendet. Die virtuelle Nadel kann in allen drei Dimensionen bewegt werden.

## 4 3D-Modellhandbuch

Eine weitere Anwendung des Trackingsystems ist das 3D-Modellhandbuch. Die Komponenten des Systems sind in Abb. 4 skizziert.

Mittels einer handelsüblichen Webcam erkennt das System schwarzweiße Marker im 3D-Modellhandbuch. Dieses Buch ähnelt einem herkömmlichen Fotoalbum. Anstelle von Fotos befinden sich darin musterbasierte Marker. Je nach Marker werden die entsprechenden Objekte auf der Präsentations-Leinwand in 3D angezeigt. Bisher lässt sich das 3D-Modell entweder durch den Marker selbst oder durch Tastaturbefehle um alle Achsen drehen. Die nächste Stufe der Entwicklung wird ein selbstkonstruierter, kabelloser Würfel sein, der die Aufgabe der Rotation übernimmt. Beispielsweise durch Farberkennung kann die momentane Position des Würfels bestimmt werden, um diese dann auf das 3D-Objekt übertragen zu können ([SE07] und [ST93]). Es soll erreicht werden, dass ein Benutzer den Würfel intuitiv als 3D-Eingabegerät benutzt. Die dafür entwickelte Software setzt auf der C++-Bibliothek „ARToolKitPlus“ [AR07] auf und basiert auf DirectX.



## 5 Ausblick

Der aktuelle Stand der Projekte wird stetig ausgebaut. Die Authentifizierung der Remote Control wird weiter entwickelt und komplexere Steuerbefehle werden implementiert werden. Auch eine Steuerung der Präsentationsprojektoren bzw. das Umschalten der verschiedenen Projektoreneingänge per PDA ist so möglich.

Das Trackingsystem soll genutzt werden, um verschiedene Interaktionstechniken, die ein solches System benötigen, verfügbar zu machen. Diese können dann getestet und verglichen werden. Neue innovative 3D-Interaktionstechniken sollen entwickelt werden.

Weitere Schritte des Projekts 3D-Modellhandbuch werden die Entwicklung eines Backends und die Anbindung der Software an eine Datenbank sein. Auf einfache Art und Weise können dann weitere 3D-Modelle dem Modellhandbuch hinzugefügt werden.

## Literaturverzeichnis

- [AF07] Aforge.NET Projektseite - <http://code.google.com/p/aforge/>, Stand: 09.12.2007
- [AR07] ARToolkitPlus - [http://studierstube.icg.tu-graz.ac.at/handheld\\_ar/artoolkitplus.php](http://studierstube.icg.tu-graz.ac.at/handheld_ar/artoolkitplus.php)  
Stand: 09.12.2007
- [AM07] AMIRA, Visualisierung von 3D-Rohdaten, <http://amira.zib.de>, Stand 08.01.2008
- [BO97] D. Bowman, L. Hodges: "An Evaluation of Techniques for Grabbing and Manipulating Remote Objects in Immersive Virtual Environments" in Proceedings of the 1997 symposium on Interactive 3D graphics SI3D, ACM Press, 1997
- [FA96] Faugeras, O.: Three-Dimensional Computer Vision – A Geometric Viewpoint, The MIT Press, 1996
- [FO96] A. Forsberg, K. Herndon, R. Zeleznik: "Aperture Based Selection for Immersive Virtual Environments" in Proceedings of the 9th annual ACM symposium on User interface software and technology UIST, ACM Press, 1996
- [IT03] INFITEC – A new stereoscopic visualisation tool by wavelength multiplex imaging, 2003, Helmut Jorke, Markus Fritz,  
[http://www.infitec.net/INFITEC\\_english.pdf](http://www.infitec.net/INFITEC_english.pdf), Stand 08.01.2008
- [NE07] Neumann T.: Auswahl einer Tracking-Bibliothek, Projektbericht des VRLabs, Hochschule Reutlingen, 2007
- [SE07] Seidel Alexander, Eger Tobias; 3D-Animationshandbuch, Hochschule Reutlingen, 2007
- [ST93] Steinbrecher Rainer, Bildverarbeitung in der Praxis [2. Auflage], München  
[u.a.], Oldenburg Verlag, 1993, ISBN 3-486-22372-0  
kostenloser Download: <http://www.rst-software.de/dbv/download.html>  
Stand 18.12.2007



# Interessenprofile in virtuellen Identitäten

Sebastian Kurt <http://skurt.de>  
Freie Universität Berlin, Deutschland  
[kurt@inf.fu-berlin.de](mailto:kurt@inf.fu-berlin.de)

**Abstract:** Mit dieser Arbeit wird das Auffinden, Extrahieren und Auswerten von Interessen in virtuellen Profilen betrachtet. Die Nutzer von Internetdiensten übergeben durch unterschiedliche Aktionen Daten an Dienste. Diese bilden ein virtuelles Abbild der Interessen einer Person. Der Vergleich zu den per Fragebogen ermittelten realen Interessenschwerpunkten zeigt inwieweit Übereinstimmung zwischen den Welten herrscht. Mit dem virtuellen und realen Interessenprofilabdruck besteht die Möglichkeit grob zu charakterisieren.

## 1 Einführung

Mit der Verlagerung vieler Bereiche des öffentlichen und privaten Lebens in das Internet bringt jeder zunehmend persönliche Informationen in die virtuelle Welt ein. Die sozialen Gemeinschaften (Bsp. Facebook, StudiVZ<sup>1</sup>) fordern ihre Teilnehmer dazu auf, die Profile für ihr Netzwerk zu füllen. Nur so sei gewährleistet, von allen Freunden gefunden zu werden und viele neue Bekanntschaften zu ergünden. All dies führt dazu, dass Menschen im World Wide Web immer gläserner werden. Werbetreibende wollen anhand von Profildaten zielgerichtete Botschaften überbringen können. Teilweise hat der Nutzer mit seinem Profil und dem Verhalten im Internet die Möglichkeit Geld zu verdienen. Inwieweit Offenheit mit eigenen Daten im virtuellen Raum die reale Persönlichkeit widerspiegelt, soll hier geklärt werden.

Das Anwendungsszenarien, welches diese Arbeit behandelt, dreht sich um den Vergleich von Interessen. Die untersuchte Person möchte erfahren (Bsp. vor Bewerbungsanfragen), inwieweit ihr virtuelles Abbild mit den realen Vorlieben übereinstimmt.

## 2 Problem

Die Offenheit mit privaten Daten im Netz steht im Widerspruch zum Datenschutz. Was aber die Anbieter wirklich anhand der Daten tun können und wie dabei für Werbende Vorteile entstehen, ist fraglich. In der Arbeit sollen diese Verwertungsmöglichkeiten betrachtet werden.

Dabei muss auf die Fragestellung, inwiefern die virtuellen Aktivitäten der Internetnutzer mit ihren realen Aktionen übereinstimmen, [Boy07] genauso eingegangen werden wie auf

---

<sup>1</sup><http://www.facebook.com>, <http://www.studivz.de> vom 11.10.2007.

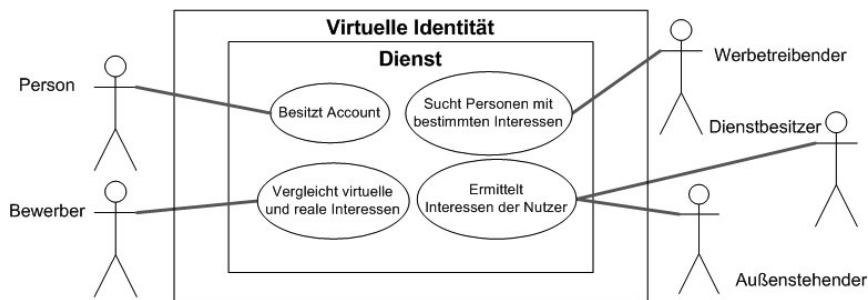


Abbildung 1: Anwendungsszenarien

die Inhalte, welche Nutzer online hinterlegen. Auch die Nutzung zu beruflichen Zwecken oder im Bildungsbereich fällt in den Betrachtungsraum. So lassen sich aus privatem Material (Bsp. Fotos vom Urlaub) leichter persönliche Interessen ermitteln als aus den Daten, welche am Arbeitsplatz entstehen. [Bro02]

Um die Nutzer in Gruppen einteilen zu können und ihre Interessen zu kategorisieren, werden verschiedene Ansätze diskutiert. Neben Klassifikationen aus der Werbebranche oder den Medienunternehmen, gelangen auch klassische Interessenmodelle in die Überlegung. So wird neben Adsense von Google, die Studie „@facts extra - Online-Nutzertypen 2007“ von SevenOne<sup>2</sup> und der Allgemeine Interessen-Struktur-Test zur Erfassung schulisch-beruflicher Interessen betrachtet.

### 3 Interessenprofil und Datenquellen

Da das Profil einer Person zeit-, umwelt- und kontextabhängigen Einflüssen ausgesetzt und damit nicht festgelegt ist, kann nicht vom „Interessenprofil“ als etwas klar abgegrenztem gesprochen werden. Es handelt sich viel mehr um temporäre Schwerpunkte in Interessengebieten. Neben diesen Hindernissen existiert die Hürde der vielfältigen Benennungen und Zuweisung zu Oberkategorien einzelner Punkte. Der Versuch, allgemeine Interessenschwerpunkte zu klassifizieren, ergibt die Kernbereiche Kultur, Entertainment, Wirtschaft, Gesellschaft, Sport, Technik, Wissenschaften. Die Haupt- und Unterkategorien der Lexika (Wikipedia und Freebase) sowie von Digg dienen dabei als Quelle<sup>3</sup>.

Die oben angesprochenen Segmentierungsmodelle unterteilen Personen aufgrund weiterer Faktoren (Bsp. soziale Herkunft, Wertorientierung, Nutzungsverhalten im WWW) in Gruppen und werden in der Diplomarbeit näher behandelt.

Die verwendeten Dienste Flickr, 43Things, LastFM, Technorati, Upcoming, Digg und Delicious<sup>4</sup> decken die Bereiche Foto, Lebensziel, Musik, Weblogs, Termin, Nachrichten und Lesezeichen ab und stellen damit auch die Grundlage des erhaltenen Interessenprofils dar.

<sup>2</sup>Werbevermarkter von ProSiebenSat.1.

<sup>3</sup><http://www.wikipedia.org/>, <http://www.freebase.com/>, <http://www.digg.com/> vom 10.11.2007.

<sup>4</sup><http://flickr.com/> <http://www.43things.com/> <http://technorati.com/> <http://upcoming.yahoo.com/> <http://del.icio.us> <http://last.fm/> <http://digg.com/> vom 10.11.2007.

## 4 Verfahren

Die Software *That'sMe* automatisiert das Auffinden, Extrahieren und Verarbeiten der Daten. Mit Hilfe der APIs der verwendeten Dienste werden die in den durch den Nutzer identifizierten Profilen gespeicherten Informationen eingesammelt. FUTEF-API sowie Cat-Graph<sup>5</sup> bestimmen die Einordnung der Daten in Wikipedia-Kategorien. Aus dieser berechnet das Verfahren die Interessenschwerpunkte im WWW. [AKD06][BL00][WW07]

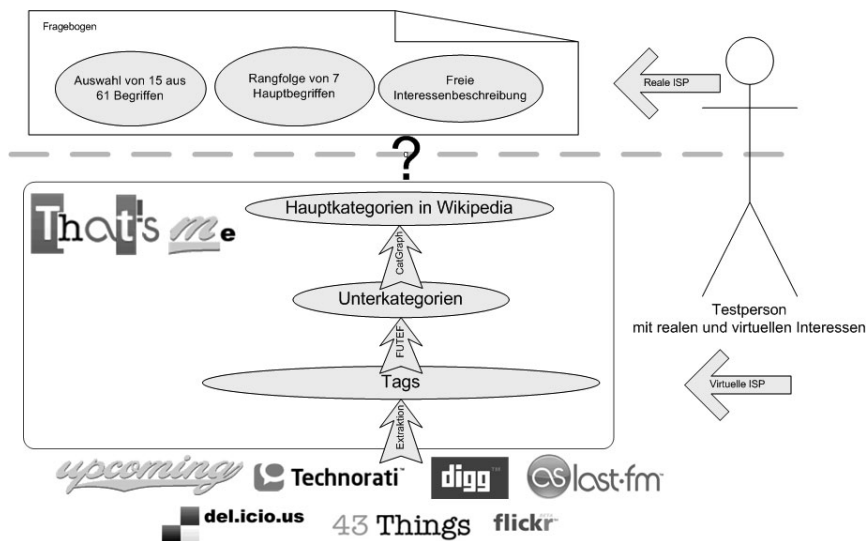


Abbildung 2: Schema des Verfahrens

Drei ausgesuchte Faktoren (Persönlichkeit, Spezialisierung, Tag-Häufigkeit) gewichten die Aussagekraft der Informationen aus den Quellen und fließen in die Bewertung ein. Dieses bildet den Ausgangspunkt für die Gewichtung von Tags, Kategorien und schließlich Interessenschwerpunkten.

Der Ansatz dieser Arbeit beschränkt sich im Verfahren auf die Verwendung von Tags. Jegliche Personendaten, Orte, Kontakte, Gruppen, Kategorien oder zeitliche Informationen fließen nicht ein. Die Betrachtung der Aufgaben von Tags fließt bei der Filterung in das Verfahren ein. [MNBD06][GH06]

Der Vergleich des virtuellen Interessenprofilabdrucks (VIPA) mit dem realen Abdruck (RIPA, aus dem Fragebogen) ist einer der drei Wege um die Übereinstimmung beider Welten zu überprüfen. Die Ergebnisse zeigen ein von Menge und Herkunft der Nutzerdaten abhängiges wirklichkeitsnahes Bild einer Person. Mit domänenspezifischen Kategoriesystemen oder alternativen Einordnungsmöglichkeiten wäre eine Optimierung möglich. Sprachliche Barrieren sowie die Lücke zwischen reinen durch Nutzer erzeugten Daten und computerverständlichen Informationen (Semantic Web) verhindern derzeit noch ein besseres Ergebnis. [AKD07][Kra92][LMD06]

<sup>5</sup><http://api.futef.com/> und <http://tools.wikimedia.de/~dape/categraph/> vom 10.11.2007.



Weitaus differenzierter wird das Thema in der Diplomarbeit besprochen.

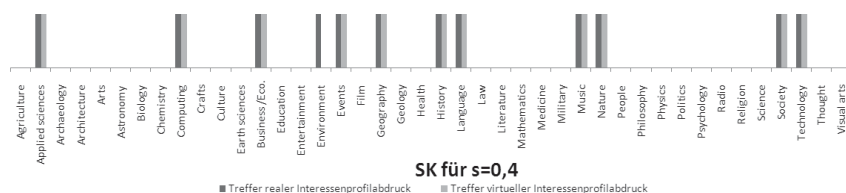


Abbildung 3: VIPA vs. RIPA einer Testperson mit Schwellwert s

## Literatur

- [AKD06] Hend S. Al-Khalifa und Hugh C. Davis. Folksonomies versus automatic keyword extraction: An empirical study. In *IADIS Web Applications and Research 2006 (WAR2006)*, 2006.
- [AKD07] Hend S. Al-Khalifa und Hugh C. Davis. Exploring The Value Of Folksonomies For Creating Semantic Metadata. *International Journal on Semantic Web and Information Systems (IJSWIS)*, 3(1):13–39, March 2007.
- [BL00] Geoffrey C. Bowker und Susan Leigh Star. *Sorting Things Out: Classification and Its Consequences (Inside Technology)*. The MIT Press, August 2000.
- [Boy07] Danah Boyd. *Why Youth (Heart) Social Network Sites: The Role of Networked Publics in Teenage Social Life*. MacArthur Foundation on Digital Learning, Identity Volume (ed. David Buckingham). MIT Press, 2007.
- [Bro02] Vanda Broughton. Facet analytical theory as a basis for a knowledge organization tool in a subject portal. In María J. López-Huertas und Francisco J. Muñoz-Fernández, Hrsg., *Challenges in knowledge representation and organization for the 21st century: Proceedings of the the 7th ISKO Conference 2002, Granada, Spain*, Seiten 135–142. Ergon Verlag, July 2002.
- [GH06] Scott Golder und Bernardo A. Huberman. Usage patterns of collaborative tagging systems. *Journal of Information Science*, 32(2):198–208, April 2006.
- [Kra92] Martin Kracker. *Vom Nutzen unscharfen Begriffswissens. Fuzzy Thesauri für die Formulierung von Anfragen in Faktendatenbanken*. In: *Kognitive Ansätze zum Ordnen und Darstellen von Wissen*. Number 2. Indeks Verlag, 1992.
- [LMD06] Hugo Liu, Pattie Maes und Glorianna Davenport. Unraveling the Taste Fabric of Social Networks. *Int. J. Semantic Web Inf. Syst.*, 2(1):42–71, 2006.
- [MNBD06] Cameron Marlow, Mor Naaman, Danah Boyd und Marc Davis. Position Paper, Tagging, Taxonomy, Flickr, Article, ToRead. In *Collaborative Web Tagging Workshop at WWW2006, Edinburgh, Scotland*, May 2006.
- [WW07] Fei Wu und Daniel S. Weld. Autonomously semantifying wikipedia. In *CIKM '07: Proceedings of the sixteenth ACM conference on Conference on information and knowledge management*, Seiten 41–50, New York, NY, USA, 2007. ACM.

# T-Prox: Vorstellung einer Lösung zur Proxy-gestützten Usabilityanalyse von webbasierten GUIs

Sven Lilienthal  
sven.lilienthal@rwth-aachen.de  
Lehrstuhl für Informatik 4  
RWTH Aachen

**Abstract:** In diesem Papier wird ein Ansatz vorgestellt, der eine kostengünstige Usability-Analyse von web-basierter Benutzerschnittstellen ermöglicht, indem automatisch das Nutzerverhalten protokolliert und für eine weitere Auswertung bereitgestellt wird. Dies wird benötigt, da viele Anbieter die eigentlich erforderlichen Tests scheuen, obwohl der Erfolg von EDV-Systemen nachweislich direkt von deren Gebrauchstauglichkeit abhängt.

## 1 Einführung

Die DIN 9241-11 [DIN98b] definiert Gebrauchstauglichkeit folgendermaßen:

Das Ausmaß, in dem ein Produkt durch bestimmte Benutzer in einem bestimmten Nutzungskontext genutzt werden kann, um bestimmte Ziele effektiv, effizient und zufriedenstellend zu erreichen.

Bei vielen Informationssystemen wird bei deren Entwicklung der Aspekt der Gebrauchstauglichkeit vernachlässigt. Da dies zu einer sinkenden Akzeptanz, einer der hauptsächlichen Erfolgsfaktoren [Ted99] führt, besteht ein großer Bedarf an automatischen bzw. unterstützenden Systemen, die eine problemlose Usability-Analyse ermöglichen, ohne das große Änderungen an den existierenden Informationssystemen oder den Rechnersystemen der Anwender vorgenommen werden müssen. Des Weiteren haben Studien gezeigt, dass eine schwierige Navigation sich negativ auf die Glaubwürdigkeit und das entgegengebrachte Vertrauen auswirkt [FML<sup>+</sup>01]. Viele Firmen lehnen jedoch Tests der Gebrauchstauglichkeit auf Grund der zur Zeit vermeintlich hohen Kosten ab. Krug [Kru05] und auch Nielsen [Nie97] zeigen Möglichkeiten für kostengünstige Usability-Tests auf, welche allerdings entweder einen hohen manuellen Aufwand erfordern oder auf der kostengünstigen Beschaffung einer Videokamera beruhen. Dies motiviert die Entwicklung eines eigenen Werkzeuges, welches automatisch die Aktionen des Benutzers mitprotokolliert und für eine einfache Auswertung bereitstellt.

Als existierende Lösungsansätze werden die beiden Proxy-basierten Lösungen *Web Event-logging Tool* (WET) von AT&T und das an der University of California in

Berkeley entwickelte *WebQuilt* hier berücksichtigt. WET entstand 1999 [EC99], um die hausinternen Webanwendungen von ATT zu testen, wurde jedoch nie veröffentlicht. Es fängt die auf Clientseite im Browser generierten Javascript-Events ab, um so eine detaillierte Übersicht über die Aktionen des Nutzers zu bekommen. WebQuilt [HHWL01] wurde in den Jahren 2001/02 in Berkeley entwickelt. Bemerkenswert ist hier vor allem die grafische Auswertung der gesammelten Daten. Dabei stellt WebQuilt den Pfad des Benutzers durch die Benutzeroberfläche als Graphen dar, die besuchten Webseiten als Knoten und die Links als Kanten. Verbindungen werden hier deutlich an Hand der Häufigkeit ihrer Nutzung farblich markiert.

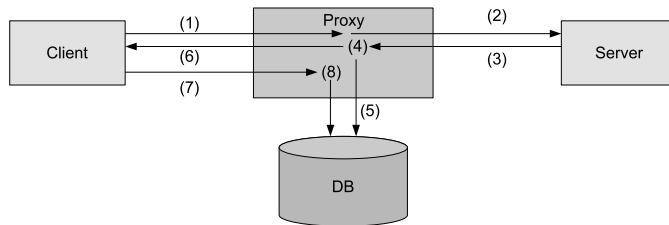
## 2 Anforderungen

Für die aussagekräftige Analyse der Gebrauchstauglichkeit sind insbesondere einige Metriken wichtig. Um die Seiten identifizieren zu können, die besonders Problematisch sind, sollten einige Daten automatisch erhoben werden. Dazu gehören die Effizienz, gemessen in der benötigten Zeit um das gewünschte Ziel zu erreichen, also der Zeitdifferenz von der ersten bis zur letzten aufgerufenen Seite, die durchschnittlich auf einer Seite verbrachte Zeit, die Anzahl und Art der getätigten Aktionen, wo runter auch die so genannten Blind-Klicks fallen, also Klicks auf nicht Maus-sensitive Elemente, sowie die durchschnittliche Anzahl von Aktionen pro Zeit. Die Anzahl der hierarchischen Rücksprünge bzw. die Anzahl der Rücksprünge auf die Anfangsseite sollten auch herausgestellt werden da sie andeuten, dass sich der Benutzer verirrt hat. Für die Auswertung dieser Daten können mit Hilfe maschineller Unterstützung Navigationsschleifen und Irrtümlichwerweise als aktiv wahrgenommenen Elemente der Benutzeroberfläche einfach identifiziert werden. Die beiden vorgestellten Werkzeuge bieten keine hinreichende Unterstützung für eine aussagekräftige Usability-Analyse der Benutzerschnittstellen von IT-Systemen.

## 3 Umsetzung

Das Programm besteht aus zwei Hauptkomponenten: einem Server, der die Aktionen protokolliert, und einem Werkzeug zur Auswertung der gespeicherten Daten. Der Server besteht aus dem eigentlichen Proxy, einem minimalen Webserver sowie einer Datenbank. Der typische Ablauf einer Anforderung ist in Abb.1 dokumentiert. Da die Sicherheitsbestimmungen bei JavaScript keine XMLHttpRequests [Web07] zu anderen Webseiten zulassen, müssen die Anfragen an die Protokollkomponente an eine zur ursprünglichen Webseite relative URL gerichtet sein. Dazu wird eine spezielle URL aufgerufen, die vom Proxy erkannt wird. Im Message-Body HttpRequest werden die Daten übermittelt und in der Datenbank gespeichert.

Das auf dem Client ausgeführte Javascript fügt dem Body-Element für alle benötigten Aktionen Verarbeitungsroutinen hinzu. Aufgrund des Event-Bubbling bzw.



1. Client fordert Seite an
2. Proxy holt Seite
3. Server liefert Seite
4. Proxy modifiziert die Seite
5. Proxy archiviert die Seite
6. Modifizierte Seite wird an den Client weiterleitet
7. Client schickt Benachrichtigungen über Aktionen an den Proxy
8. Proxy speichert die Aktionen

Abbildung 1: Ablauf eines Requests

Event-Capturing, also des weiterleitens aller Events entlang des Pfades im DOM-Tree, wird jedes erzeugte Event auch allen im Baum darüber liegenden Elementen mitgeteilt. Damit reicht eine Registrierung als Event-Handler beim Body-Element, um alle auf einer Seite erzeugten Events zu erhalten. Diese werden dann durch das Script protokolliert und per XMLHttpRequest an den Proxy zurückmeldet.

Das zur Auswertung entwickelte Werkzeug stellt die von einem einzelnen Nutzer besuchten Seiten als Graph dar. Dabei werden häufig besuchte Seiten hervorgehoben, diese scheinen besonders kritisch zu sein und bedürfen einer weiteren Untersuchung. Durch die Auswahl eines Knotens werden die auf dieser Seite getätigten Aktionen, sowie die auf dieser Seite verbrachte Zeit und die durchschnittliche Zeit pro Aufruf angezeigt. Des Weiteren werden allgemeine Informationen zu diesem Benutzer aufgelistet, darunter die vom ersten bis zum letzten Aufruf vergangene Zeit, die Gesamtzahl der getätigten Aktionen, sowie die durchschnittliche Zeit pro Aktion.

## 4 Zusammenfassung und Ausblick

Das hier vorgestellte Werkzeug erleichtert die Analyse der Gebrauchstauglichkeit von Webseiten in erheblichem Maße. Da es kostenlos zur Verfügung gestellt werden soll<sup>1</sup>, stellt es eine günstige Alternative zu den bestehenden, teuren Lösungen dar. Im Bereich der Auswertung ist noch Potential für Verbesserungen vorhanden, so könnten fehlerhaft angeklickte Elemente besser automatisch identifiziert und gekennzeichnet werden.

---

<sup>1</sup><http://www.t-prox.net>

## Literatur

- [DIN98a] DIN Deutsches Institut für Normung e.V., Hrsg. *Ergonomie der Mensch-System-Interaktion ; Teil 151: Leitlinien zur Gestaltung von Benutzungsschnittstellen für das World Wide Web ; DIN EN ISO 9241-151*. Beuth Verlag GmbH, Berlin, 1998.
- [DIN98b] DIN Deutsches Institut für Normung e.V., Hrsg. *Ergonomische Anforderungen für Bürotätigkeiten mit Bildschirmgeräten ; Teil 11: Anforderung an die Gebrauchstauglichkeit - Leitsätze ; DIN EN ISO 9241-11*. Beuth Verlag GmbH, Berlin, 1998.
- [EC99] Michael Etgen und Judy Cantor. What does getting WET (Web Event-logging Tool) mean for Web Usability? In *Proceedings of the 5th Conference on Human Factors & the Web*, June 1999. <http://zing.ncsl.nist.gov/hfweb/proceedings/proceedings.en.html>.
- [FML<sup>+</sup>01] B. J. Fogg, Jonathan Marshall, Othman Laraki, Alex Osipovich, Chris Varma, Nicholas Fang, Jyoti Paul, Akshay Rangnekar, John Shon, Preeti Swani und Marissa Treinen. What makes Web sites credible?: a report on a large quantitative study. In *CHI '01: Proceedings of the SIGCHI conference on Human factors in computing systems*, Seiten 61–68, New York, NY, USA, 2001. ACM.
- [HHWL01] Jason I. Hong, Jeffrey Heer, Sarah Waterson und James A. Landay. WebQuilt: A proxy-based approach to remote web usability testing. *ACM Trans. Inf. Syst.*, 19(3):263–285, 2001.
- [Kru05] Steve Krug. *Don't Make Me Think: A Common Sense Approach to the Web (2nd Edition)*. New Riders Publishing, Thousand Oaks, CA, USA, 2005.
- [Nie97] Jakob Nielsen. Discount Usability for the Web. Online, 1997. [http://www.useit.com/papers/web/\\_discount/\\_usability.html](http://www.useit.com/papers/web/_discount/_usability.html).
- [Ted99] Bob Tedeschi. Good Web Site Design Can Lead to Healthy Sales. *The New York Times*, August 1999. <http://www.nytimes.com/library/tech/99/08/cyber/commerce/30commerce.html>.
- [Web07] Web API Working Group. The XMLHttpRequest Object. Working draft, W3C, June 2007. <http://www.w3.org/TR/XMLHttpRequest/>.

# An Analysis Dashboard for Evaluating System Performance and User Behavior in Web Portals

Markus Pahs

Department of Computer Science, FSU Jena  
Ernst-Abbe-Platz 2, 07743 Jena  
markus.pahs@uni-jena.de

Andreas Nauerz

IBM Germany Research & Development  
Schönaicher Str. 220, 71032 Böblingen  
andreas.nauerz@de.ibm.com

Diploma Thesis

(Supervisors: Prof. Dr. Birgitta König-Ries, Prof. Dr. Martin Welsch)

**Abstract:** Web portals are Web sites that offer users a single point of access for their information needs. In order to enhance user satisfaction, it is essential that navigation and structure of the portal are intuitive to the users and meet their expectations. For Web pages in general, this problem is being addressed by applying data mining techniques to the usage logs and optimizing the Web page according to the results of this analysis. In this thesis, we investigate in how far the techniques developed for the Web are suitable for Web portals and present a prototypic framework, the Analysis Dashboard, for analyzing user behavior and system performance in Web portals.

## 1 Introduction

Web portals are Web sites that *provide site users with a single point of access to multiple types of information and applications. Regardless of where the information resides or what format it uses, a portal aggregates all of the information in a way that is pleasing and relevant to the user* [RCG<sup>+</sup>05]. This definition emphasizes the importance of designing Web portals in such a way that navigation and structure are intuitive to the user and meet their expectations. The main challenge is how to structure a portal to achieve this aim. Today, usually some central instance, most often the administrator, creates a portal structure that is optimal in his view and should satisfy the need of the majority; the check whether the goal is really achieved will be done, if ever, in the future. Since the impact of Web sites' behavior on business success and customer loyalty is known and "the right quantitative and qualitative information is essential" [MDH01], the adaptation process has to be more accurate and faster.

This problem, of course, is not restricted to Web portals, but exists for Web sites in general. One method to measure the satisfaction and the usage behavior of Web site visitors is performing Web analytics. Web analytics, in particular Web mining, is the process of discovering potential useful and previously unknown information from Web data. This often includes the same metrics, such as number of visits, the time of residence, the date of request, frequent itemsets, sequential rules, and so on. These metrics can be used to detect (past) trends on the huge amount of logged data, but the question is do they meet our requirements?

This paper will provide an overview about the state of the art of Web analytics and Web mining [FPSS96, CMS99, SCdT00], and will then continue to investigate the usability of these techniques for the optimization of business Web portals [RCG<sup>+</sup>05, Nau07]. As a result, we will present the analysis dashboard, our prototypical framework for analyzing user behavior in Web portals.

## 2 Web Usage Mining

The definition of Web mining given by [CSM97] contains both the search of information in Web resources, i.e., Web content mining, which is different from Knowledge Discovery in Databases due to the heterogeneous and semi-structured kind of accessible data, and the identification of user access patterns on Web servers, Web Usage Mining (WUM). The analysis of such data can help to identify the user's needs, to increase the usability and performance of the Web site, or show errors and bottlenecks on a Web resource. The Web usage mining process elucidated by [SCdT00] can be divided into three main steps: pre-processing, pattern discovery, and pattern analysis. Applying this process to Web portals needs to take a number of differences into account:

*Preprocessing* is the first step in the Web usage mining process and transforms the usage, content, and structure information into abstract data. In Web portals it is important to mention that there could be different resource types, such as portlets. Also possible is additional information within a log file that shall be mined by a Web analytics tool.

The second step in the WUM process covers the *discovery of useful patterns* in the pre-processed data. The methods and algorithms used in this step often are derived from other fields of research, such as statistics, data mining, and machine learning. The final component responsible for pattern discovery will be encapsulated and therewith the commonly known concepts can be used. The only points we have to respect are the differing kinds of input data. Sometimes it is also necessary to abstract or convert several steps of the methodology to make sense in the Web domain.

The last step, *pattern analysis*, is used to filter uninteresting patterns of the previous stage. The filtering will be done by the application that displays the data, because of the varying interests of the users for which Web mining is done. For a better understanding of the collected data, and maybe a decision which information is relevant, visualizations are used. These visualization techniques could be graphs, charts, or simply different colors. In contrast to most of the other tools, Web portals are able to use portlets for separation and customization of a user's result. Also some portal specific interpretations are necessary.

## 3 Architecture

Some of the tools successfully used today for Web site analytics can be used in Web portals [LU06]; however they do not exploit the portal's specifics. We did not find any Web statistics tool that is really integrated into a Web portal and consequently utilizes provided portal features. First of all, besides some privacy issues, the usage of a decentralized Web mining tool is not a real option due to the lacking attention on portal specifics. Addition-

ally, according to Mason [Mas06] we think the most required benefits of a good Web stats tool or framework should be usability, extendibility, scalability, and an easy to handle data storage like data warehousing.

Our Analysis Dashboard framework consists of the three primary components *preprocessor*, *core*, and *evaluator* (cp. Fig. 1) adapting tasks from the common WUM process and a fourth component called *data manager* responsible for data provision and persistent data storage. The function and structure of the data manager is inspired by the data warehouse concept.

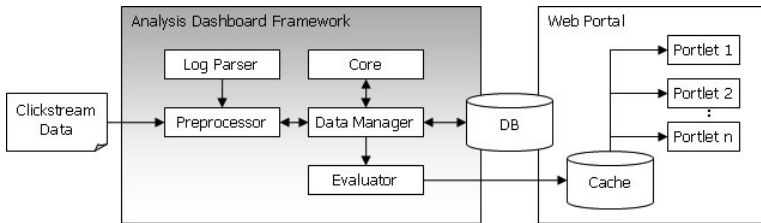


Figure 1: Architecture of the Analysis Dashboard framework

The *preprocessor* receives its data record via a defined interface. This enables the application of different clickstream data sources. Thus an event listener or log parser only has to extract the desired information out of the raw record and transmit them to the interface. The preprocessor component normalizes, aggregates, or splits the raw data into an internal data format. After the preprocessing all the aggregated records are stored by the *data manager*. The *core component* requests the needed data sets from the data manager depending on the data mining algorithm or the method used to discover patterns. The data mining algorithms are stored externally, so that each algorithm must be accessed via an interface, too. That way encapsulation and easy extensibility are ensured. The result of the core component is again stored by the data manager component. The last component, the *evaluator*, also takes the relevant data from the data manager to evaluate it with respect to the used metric. The result of the evaluator is transmitted to the requesting actor.

## 4 Conclusion

In this paper, we describe a portal specific framework in the rapidly growing area of Web Usage mining. The interest in Web usage mining becomes more and more commercial, since the correlation between user's satisfaction and Web site structure and layout is known. Therefore many tools have been introduced to build knowledge about usage behavior out of raw log data. However, in the Web portal domain some things are different from other areas and no tool really is integrated into a Web portal. This article has described the challenges involved and shown the feasibility to do sophisticated Web analytics within a portal, so that the next generation of portal evolution can arise. These will be portals that will automatically adapt to changing user contexts.



## References

- [CMS99] R. Cooley, B. Mobasher, and J. Srivastava. Data Preparation for Mining World Wide Web Browsing Patterns. *Knowledge and Information Systems*, 1(1):5–32, 1999.
- [CSM97] R. Cooley, J. Srivastava, and B. Mobasher. Web Mining: Information and Pattern Discovery on the World Wide Web. In *9th IEEE International Conference on Tools with Artificial Intelligence (ICTAI'97)*, November 1997.
- [FPSS96] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth. From data mining to knowledge discovery in databases. *Ai Magazine*, 17:37–54, 1996.
- [LU06] S. Liesche and S. Uhlig. Using portal analytics with open-source reporting tools: Monitor, manage, and adapt your portal using WebSphere Portal log data. Technical report, IBM Corporation, 2006.
- [Mas06] N. Mason. The Ideal Web Analytics Tool. Web column, April 2006.
- [MDH01] A. Malacinski, S. Dominick, and T. Hartrick. Measuring Web traffic, Part 1: Getting to know your audience is the key to online success – make the numbers work for you. Technical report, IBM Corporation, 2001.
- [Nau07] A. Nauerz. The Contextual Portal. Master’s thesis, University of Hagen, 2007.
- [RCG<sup>+</sup>05] J. Rodriguez, C. Cesario, K. Galvan, B. Gonzalez, G. Kroner, G. Rutigliano, and R. Wilson. *IBM Rational Application Developer V6 Portlet Application Development and Portal Tools*. IBM Corp, 2005.
- [SCDT00] J. Srivastava, R. Cooley, M. Deshpande, and P.-N. Tan. Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data. *SIGKDD Explorations*, 1(2):12–23, 2000.

# Using Unstructured Data Analysis and Semantic Tagging for the Adaptation of Portal Environments

René Pietzsch<sup>1</sup>, Andreas Nauerz<sup>2</sup>

renepiet@de.ibm.com, andreas.nauerz@de.ibm.com

<sup>1</sup>Technische Universität Dresden  
Fakultät Informatik  
Nöthnitzer Str. 46, 01187 Dresden

<sup>2</sup>IBM Deutschland Forschung &  
Entwicklung GmbH  
Schönaicher Str. 220, 71032 Böblingen

**Abstract:** Today's Web Portals suffer from information overload. This thesis [10] outlines how *tagging* can help to overcome this drawback. Leveraging collective intelligence and gaining insight into the users' characteristics can be used to issue content- and link-level adaptation as well as for sophisticated recommendations.

## 1 Introduction

Recently Enterprise Information Portals (EIP) gained importance in many companies. They offer a single point of access to personalized content, services, applications and processes via one seamless user interface. They become so successful that their constant growing is becoming a serious problem and more and more a productivity threat. EIP users need to find task and role specific information very quickly, but they face information overload and feel "lost in hyperspace". The default hierarchical structures have been designed to satisfy the majority of users not to serve a particular user best. This leads to long navigation paths and causes significant effort to find the information needed. Due to these reasons the Contextual Portal project [4] has been started to overcome the mentioned drawbacks and to evolve the next generation of Portals. The envisioned Portal should become more adaptive, adaptable, dynamic and context-aware. Therefore, content- and link-level adaptation methods [2] are applied to Portals as they are very promising to reduce the problems outlined above but have not yet been applied to EIP.

This work focuses on the utilization of the collaborative tagging pattern for the adaptation of Portals. Thereby, "tagging" is denoted as the association of words or phrases with a Portal resource. Beside the obvious, widely-agreed-upon use of tagging [3], it can add valuable meta-information and even lightweight semantics to web resources as the majority of tags are rather factual [1, 11]. We present the architecture of a tagging and adaptation layer that integrates tagging and rating functionalities into IBM WebSphere Portal and allows for recommendations and adaptations based on the applied tags. It further allows arbitrary annotators, e.g. human users or analysis components, to annotate arbitrary Portal resources (pages, portlets, users, etc). The provided tags build a basis for user and context modelling as well as for recommendation and adaptation components that tailor the Portal to the user and his context. This includes the recommendation of similar resources, content- and link-level adaptation and more.

## 2 Conceptual Design

A high level visualisation of the conceptual design of our approach is pictured in figure 1; each of the *Tagging and Adaptation Frameworks* layer are covered in more detail in the following sections. This approach refines the ideas initially proposed in [7].

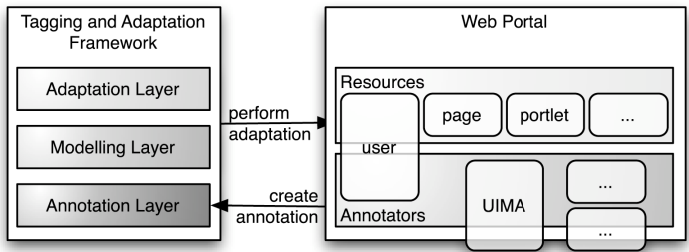


Figure 1 Three-Layered Adaptation Framework

### 2.1 Annotation Layer

Today’s web applications supporting semantic annotations limit their associations to one specific resource type and do not distinguish between different annotators. The presented approach introduces a generic annotation layer which can be used by any annotator type. Thereby, user generated annotations and automatically extracted semantics [6] are stored in the same database along with the type of the annotator. This allows handling annotations by different annotators each in a separate way (e.g., because of varying accuracy, quality, etc.). Integrating programmatic annotators can enrich the data collection in rich but small scaled content like email or chat conversations, wiki or blog posts. The annotation layer needs to be equipped with user interface components to allow users applying and editing tags as well as ratings and to use tag-based navigation. A tag-resource-matrix is defined and serves as a compact data level representation, which is the prerequisite for the subsequent semantic distance calculations. It represents the relations of tags and resources and can easily be constructed. A single column of the matrix contains a user driven description of a resource, as the assigned tags depict the direction of this column vector.

### 2.2 Modelling Layer

The modelling layer aims to derive “higher knowledge” from tags and the tag-resource-matrices. In addition to web mining based modelling [9]; this thesis proposes a pragmatic approach to compare tags and resources: the semantic distance calculation. Based on the tag-resource-matrix the relatedness of tags and resources can be determined as related (column- or row-) vectors have a similar direction. Thus, cosine similarity (or other metrics) can be used to quantify the relatedness. A modification of this method is introduced by applying singular value decomposition before calculating the relatedness to disclose latent structure and to lower the impact of “noisy” tags.

## 2.3 Adaptation Layer

Based on the calculated distances various recommendations and adaptations can be performed. Recommendations may be issued for tags and resources. The semantic distance calculations allow for an estimation of the relation between tags or resources. Additionally, by allowing users to be tagged as a resource, recommendations for experts in a specific field may be produced [5]. A colour scheme is introduced that enables a visual depiction of the level of relatedness. The knowledge about semantic distances between tags and resources can also be used to improve well-known link and content level adaptations from the field of Adaptive Hypermedia [2]. Transformation can be applied to the navigation topology or the page layout to produce a more user-tailored Portal UI. Ratings, tags and automatically analysed content are evaluated and interesting parts are presented while irrelevant items are hidden.

## 3 System and Solution

The prototype realized along with this thesis features a database design allowing to store arbitrary tag and resource types and an interface for UI components as well as for further processing (e.g. the modeling layer). Various UI components have been realized enabling to apply and edit tags and ratings of resources (tagging and rating widget) as well as to use tags for navigation purposes (tag cloud widget). The tagging and the tag cloud widget visualize related tags and resources respectively utilizing the semantic distances calculated in the modeling layer. Another component applies the concept of dynamic assembly transformation [4, 9] to perform a user-scoped dynamic modification of the navigation topology at runtime. We used a subset of resources, the pages and created a minimum spanning tree from a graph like representation of those (each edge between two pages is weighted using their respective semantic distance) to adapt the navigation structure based on the applied tags.

## 4 Conclusion and Future Work

Summarizing the present work several concepts have been explained that allow the utilization of tags and ratings for recommendations of resources and other users (experts) as well as to adapt the layout or the navigation structure of a Web Portal. Additionally, a concept for the integration of automated tagging is exploited and future work in this area is very promising [6]. Furthermore, future efforts include the extension of our recommendation and adaptation techniques. We will investigate improvements to our calculation models and “behavioral clustering” of users. Additionally, we want to generalize our concepts to recommend others than related resources only, e.g. portlet content and documents, and to adapt more than just navigation topologies [8].

*IBM and WebSphere are trademarks of International Business Machines Corporation in the United States, other countries or both. Other company, product and service names may be trademarks or service marks of others.*

## Bibliography

- [1] H. S. Al-Khalifa and H. C. Davis. Towards better understanding of folksonomic patterns. In Proc. of the 18th Conf. on Hypertext and hypermedia, pages 163-166, NewYork, NY, USA, 2007. ACM Press.
- [2] P. Brusilovsky. Adaptive Hypermedia. User Modeling and User-Adapted Interaction, 11(1-2):87-110, 2001.
- [3] C. Marlow, M. Naaman, D. Boyd, and M. Davis. Position Paper, Tagging, Taxonomy, Flickr, Article, ToRead. In Collaborative Web Tagging Workshop at WWW2006, May 2006.
- [4] A. Nauerz. The Contextual Portal. Master's thesis, University of Hagen, January 2007.
- [5] A. Nauerz and G. Groh. Implicit Social Network Construction and Expert User Determination in Web Portals. In Proc. of the AAAI Spring Symposium on Social Information, Stanford University, California, USA, March 2008.
- [6] A. Nauerz, M. Junginger and S. Zhao. A Recommender based on Automatic Metadata Extraction and User-Driven Collaborative Annotating. Proceedings of the International Workshop on Recommendation and Collaboration (Recoll2008), Canary Islands, Spain, January 2008. In Conjunction with 2008 International ACM Conference on Intelligent User Interfaces (IUI2008).
- [7] A. Nauerz, S. Pietschmann and R. Pietzsch. Collaborative Annotation-Driven Adaptation in Web Portals. In Proceedings of the 18th Conference on Hypertext and Hypermedia (Manchester, UK, September 10 - 12, 2007). HT '07. ACM, New York, NY, 155-156, 2007.
- [8] A. Nauerz, S. Pietschmann, R. Pietzsch and S. Zhao. A Framework for TagBased Adaptation in Web Portals. In Proceeding of the 17th International World Wide Web Conference. WWW2008. Beijing, China, April 2008
- [9] A. Nauerz and M. Welsch. (Context) Adaptive Navigation in Web Portals. In Proc. of the Intl. IADIS WWW/Internet Conference 2007, Vila Real, Portugal, October 2007.
- [10] R. Pietzsch. Using Unstructured Data Analysis and Semantic Tagging for the Adaptation of Portal Environments. Diplomarbeit, Technische Universität Dresden, November 2007
- [11] S. Sen, S. K. Lam, A. M. Rashid, D. Cosley, D. Frankowski, J. Osterhouse, F. M. Harper, and J. Riedl. tagging, communities, vocabulary, evolution. In Proc. of the 20th Conf. on Computer supported cooperative work, pages 181-190, New York, NY, USA, 2006. ACM Press.

# Shareclipse - Ein Eclipse-Plugin für kollaborative Softwareentwicklung auf Basis der P2P-Technologie JXTA

Dipl.-Inform. Med. Martin Wiesner

Universität Heidelberg/Hochschule Heilbronn: Abteilung Medizinische Informatik  
mwiesner@hs-heilbronn.de

Der fortschreitende Ausbau breitbandiger Hochgeschwindigkeitsnetze zwischen Forschungseinrichtungen, Konzernen und zunehmend auch Heimanwendern bietet neue Möglichkeiten zur Nutzung von Rechnerkapazitäten. Das als *'death of distance'* bezeichnete Phänomen, dass es quasi keine Rolle mehr spielt, wo auf der Welt eine nutzbare Ressource zur Verfügung steht, eröffnet eine neue Dimension des kollaborativen Arbeitens. Zeitliche Verzögerungen, die bisher ein effizientes Zusammenarbeiten stark eingeschränkt haben, werden zur Nebensächlichkeit. Zugleich entstehen in Unternehmen, bedingt durch die voranschreitende Globalisierung der Wirtschaftswelt, neue Notwendigkeiten zur globalen Vernetzung von Mitarbeitern und der dynamischen Bildung von Teams. Selbige bestehen häufig aus Personen, die auf ihrem Fachgebiet hochgradig spezialisiert, jedoch weltweit verteilt sind.

Eine effiziente Möglichkeit, beiden Notwendigkeiten gerecht zu werden, bietet das Konzept des *P2P-Computing*<sup>1</sup>. Da flexible Gruppenbildung die Grundidee darstellt und *P2P-Computing* in heterogenen Netzstrukturen eingesetzt werden kann, zieht dies zwangsläufig Software nach sich, die neue Möglichkeiten in der Arbeitswelt hervorbringt. Sowohl die Bildung von ad hoc zusammengestellten Teams, als auch das Arbeiten an gemeinsamen 'Objekten' in Echtzeit haben viel Potential.

Tele-Arbeit in Unternehmen von jedem beliebigen Punkt der Welt aus ist mittlerweile in einigen Feldern wie dem Software-Engineering schon weit verbreitet. Jedoch besteht speziell auf diesem Gebiet noch Bedarf an Echtzeit-Tools zur gemeinsamen Arbeit an Software-Quellcode. Auch im Hochschulbereich ist der Einsatz derartiger Hilfsmittel üblicherweise noch nicht verbreitet. Die Nutzung eines unterstützenden Tools zu Lehr- und Korrekturzwecken eröffnet Möglichkeiten zur Betreuung von Studenten außerhalb der Hochschule. Tele-Teaching ist für diese Form der Lehre eine treffende Umschreibung.

## 1 Ideen & Konzeption

Die Thematik der vorliegenden Arbeit (siehe [Wie07]) umfasst den Entwurf und die Realisierung eines Groupware-Tools, welches Tele-Teaching im Hochschulumfeld oder auch im Unternehmensbereich unterstützen soll. Dabei spielt die Mensch-Maschine-Interaktion eine

---

<sup>1</sup>P2P: Peer to peer - Peer (engl.): Gruppenmitglied

wichtige Rolle. Über die üblichen Awareness-Funktionen (vgl. [BS99]) hinaus erhöhen unterstützende Audio- und Video-Konferenz Funktionen zudem den Akzeptanzgrad beim Anwender, da ein menschliches Gegenüber einen vertrauteren Eindruck vermittelt als nur ein blankes Software-System.

Als Basis des zu konzipierenden Software-Tools wurde das Framework JXTA<sup>2</sup> ausgewählt. Dieses bildet die technologische Grundlage für die Kommunikation von Gruppenmitgliedern im Stile eines P2P-Netzes. Die Konzepte des von Sun Microsystems 2001 veröffentlichten Frameworks sind dank einer weltweiten Community, guter Dokumentation und diverser Fachbücher ([Wil02], [OG02], [Gra02]) schnell erlernbar.

Da der Schwerpunkt der zu realisierenden Groupware im gemeinsamen, synchronen Bearbeiten von Quellcode liegt und die Integration in Eclipse eine Vorgabe war, wurde das entwickelte Plugin `SHARECLIPSE` getauft. Dadurch wird die Verbindung der Kernidee - etwas in einem virtuellen Team zu teilen - mit der Entwicklungsplattform deutlich. Dabei ist der 'Share'-Aspekt nicht allein auf Quellcode beschränkt. Theoretisch ist auch das verteilte Modellieren von UML-, Datenbank- oder ähnlichen Modellen denkbar (vgl. [Jun03]). Auf dem jetzigen Stand der Implementierung ist jedoch nur die gemeinsame Bearbeitung von (Java-)Quellcode möglich.

Ein Einsatz in anderen Bereichen, wie z.B. der Medizin zur Besprechung schwieriger Fälle, ist in einer modifizierten Version ebenfalls denkbar. Weltweit verteiltes medizinisches Expertenwissen könnte in Echtzeit in einer Tele-Konsultation hinzu gezogen werden, um Fallbesprechungen durchzuführen und somit schneller zu einer Diagnose bzw. Entscheidung zu gelangen.

Existierende fallbasierte Systeme zur computergestützten Aus- und Weiterbildung in der Medizin - wie z.B. CAMPUS ([HBHL04]) - bieten hierfür interessante Anknüpfungspunkte.

## 2 Ergebnisse

Abbildung 1 gibt einen Überblick über die Benutzeroberfläche der Hauptkomponenten des Shareclipse-Plugins. Dabei kennzeichnen die durch Buchstaben markierten farblichen Umrandungen die Hauptbestandteile A - E, welche in Tabelle 1 zugeordnet und beschrieben werden.

Die verteilte Bearbeitung von Quellcode findet in einer Instanz des sogenannten *Shareclipse-Editors* (D) statt. Dabei werden idealerweise alle Teilnehmer einer *SharedSession* nahezu in Echtzeit durch Anpassung ihrer lokalen Sicht auf das Dokument über jegliche Veränderung informiert. Shareclipse integriert sich nahtlos in die Eclipse-Workbench. Die User-Interaktion wird durch einen Konfigurations-Wizard sowie diverse - hier nicht diskutierte - Dialoge erleichtert. Zusätzlich zum Aspekt des Arbeitens an verteilten Dokumenten ist durch die Arbeit von Kroll (siehe [Kro07]) zudem Multimedia-Unterstützung realisiert worden. Audio- und Video-Kommunikation sind zwischen den verschiedenen Teilnehmern bei vorhandener Hardware und korrekter Konfiguration möglich.

---

<sup>2</sup><http://www.jxta.org>

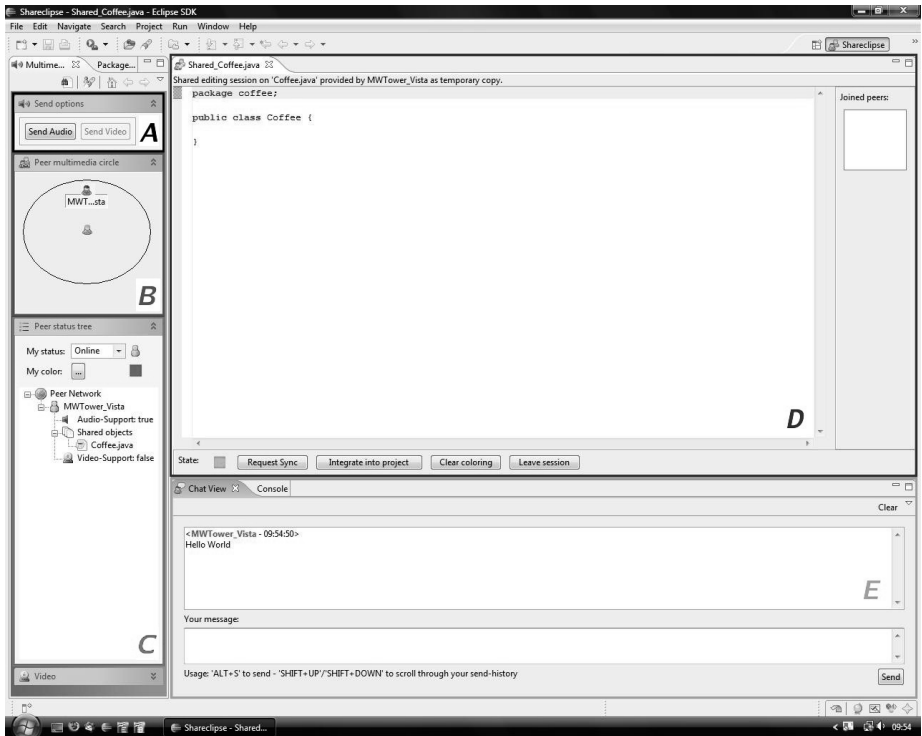


Abbildung 1: Überblick über die Benutzeroberfläche von Shareclipse (Quelle: Eigenerstellung d. A.)

Buchstabe/Bedeutung	
A	<i>Send options</i> : Ermöglicht die Bedienung der Multimedia-Komponenten von Shareclipse
B	<i>Peer multimedia circle</i> : Ermöglicht das Verbinden auf Audio- oder Video-Streams, welche von anderen Peers angeboten werden. Das Icon im Zentrum repräsentiert den lokalen Peer.
C	<i>Peer status tree</i> : Darstellung der momentan im Peernetz aktiven Peers, sowie Zustandsinformationen über diese. Veränderungen von Zuständen werden in Echtzeit abgebildet.
D	<i>ShareclipseEditor</i> : Ermöglicht das synchrone verteilte Editieren von beliebigen (Java) Quellcode-Dateien. Rechtsseitig erfolgt die Darstellung von Session-Teilnehmern.
E	<i>ChatView</i> : Ermöglicht die textuelle Kommunikation des virtuellen Teams.

Tabelle 1: Beschreibung und Zuordnung der umrandeten Komponenten aus Abbildung 1

### 3 Fazit & Ausblick

Das Ergebnis der Diplomarbeit ist ein flexibles Eclipse-Plugin, welches über eine sogenannte Update-Site<sup>3</sup> bezogen und aktualisiert werden kann. Zur Zeit befindet sich Shareclipse in einer Evaluations- und Weiterentwicklungsphase. Das technologische Fundament des realisierten Systems - das *SharedObject*-Konzept - bietet Möglichkeiten, weitere Funktionalität

<sup>3</sup><http://update.shareclipse.de>



zu entwickeln und diese in das bestehende System zu integrieren. Realistisch wäre die Implementierung einer Komponente, welche das gemeinsame Editieren von UML-Diagrammen ermöglicht.

Ein weiterer interessanter Aspekt betrifft die Implementierung eines Rollenkonzepts, wodurch bestimmte Gruppenmitglieder in die Rolle eines Gruppenleiters schlüpfen könnten. Besonders relevant wäre dies im Bereich der Lehre, um Teamsitzungen leiten und moderieren zu können. Eine Standalone-Variante könnte zudem andere Gebiete des P2P-Computings erschließen. Die Idee ein Tool zu entwickeln, welches es Health-Professionals ohne Rücksicht auf räumliche Distanzen ermöglicht, gemeinsam in „virtuellen Visiten“ Fälle durchzusprechen und ihre Diagnosefindung abzustimmen, wurde bereits angesprochen.

Solange sowohl in den Bereichen der mittelständischen Unternehmen, der Hochschulen, als auch in der öffentlichen Verwaltung sich Groupware-Lösungen suboptimal in vorhandene Arbeitsabläufe integrieren lassen (vgl. [Len07]) und zurückhaltend eingesetzt werden, sollten Software-Entwickler weiterhin versuchen (neue) Systeme zu entwickeln, welche die Bildung virtueller Teams fördern und diese besser unterstützen. Aus heutiger Sicht kann zwar keine konkrete Vorhersage über die Verbreitung von Groupware im täglichen Arbeitsleben in den nächsten 20 Jahren getroffen werden, von einem globalen Trend zur „Virtualisierung der Arbeitswelt“ ist der Autor jedoch fest überzeugt.

## Literatur

- [BS99] Thomas Berlangue und Markus Sohlenkamp. Visualizing Common Artefacts to Support Awareness in Computer-Mediated Cooperation. *Computer Supported Cooperative Work*, 8:207–238, 1999.
- [Gra02] Joseph Gradecki. *Mastering Jxta: Building Java Peer-to-Peer Applications*. John Wiley & Sons, Inc., New York, NY, USA, 2002.
- [HBHL04] M Haag, M Bauch, J Heid und FJ Leven. Fallbasiertes und interaktives Lernen in der Medizin mit CAMPUS. *Aktuelle Trends in der Softwareforschung. Tagungsband zum doIT-Software-Forschungstag am 29. Oktober 2004*, Seiten 202–211, 2004.
- [Jun03] Markus Oliver Junginger. Peer-to-Peer mit JXTA, Teil2. *Java Magazin - Internet & Enterprise Technology*, 5:26–31, 2003.
- [Kro07] Sebastian Kroll. Konzeption und Realisierung einer multimedialen kollaborativen Umgebung mit Java. Diplomarbeit, Medizinische Informatik - Universität Heidelberg/Hochschule Heilbronn, Deutschland, 2007.
- [Len07] Thorsten Lenk. Kapitel 2.14 - Virtuelle Teamarbeit. *Das flexible Unternehmen - Digitale Fachbibliothek auf CD-ROM über 3.800 Seiten*, 2007.
- [OG02] Scott Oaks und Li Gong. *JXTA in a Nutshell*. O'Reilly & Associates, Inc., Sebastopol, CA, USA, 2002.
- [Wie07] Martin Wiesner. Design und Implementierung einer kollaborativen Entwicklungsplattform auf Basis der P2P-Technologie JXTA. Diplomarbeit, Medizinische Informatik - Universität Heidelberg/Hochschule Heilbronn, Deutschland, 2007.
- [Wil02] Brendon Wilson. *JXTA*. Pearson Education, 2002.

# Mehrbenutzersynchronisation für XQuery in nativen XML-Datenbankmanagement-Systemen

Sebastian Bächle

AG Datenbanken und Informationssysteme  
Fachbereich Informatik  
Technische Universität Kaiserslautern  
D-67653 Kaiserslautern, Germany  
baechle@informatik.uni-kl.de

## 1 Motivation

Native XML-Datenbankmanagement-Systeme (*XDBMS*) sind die Antwort der Informatikforschung und der Software-Industrie auf den steigenden Bedarf an geeigneten Speicherlösungen für die stetig wachsenden Dokumentenbestände. Sie sollen die Einschränkungen konventioneller relationaler DBMS aufheben, indem sie gezielt auf die Bedürfnisse der Verarbeitung und persistenten Speicherung baumstrukturierter Daten zugeschnitten werden. Die effiziente Unterstützung der deklarativen XML-Anfragesprachen XPath und XQuery ist dabei ein zentraler Aspekt, weshalb man sich zur Zeit fast ausnahmslos auf die schnelle Beantwortung von XML-Anfragen konzentriert.

Neben weiteren Verbesserungen in diesem Bereich wird es darüber hinaus zunehmend attraktiver und auch wichtiger, die Systeme auch für „lebende“ Dokumentenbestände einsetzen zu können. Die Fähigkeit zur transaktionsorientierten Verarbeitung und Modifikation von Dokumenten in einer echten Mehrbenutzerumgebung wird deshalb zu einer Kernanforderung an ein XDBMS werden.

Die dafür bislang vorgestellten Sperrverfahren für XDBMS verfolgen zwei sehr unterschiedliche Strategien. Ein Teil der Ansätze setzt die Sperren direkt auf den Knoten der Baumrepräsentation der Dokumente [HKM04, HH04, JC06]. Dadurch erreichen sie zwar eine hohe Auflösung und eine gute Parallelität, können Phantome bei mengenorientierten Zugriffen jedoch nicht mit vertretbarer Parallelität verhindern. Viele dieser Verfahren benötigen zudem sehr viele Sperren und beeinträchtigen so das Leistungsverhalten der Systeme sehr stark.

Die anderen Ansätze benötigen vergleichsweise wenig Sperren, da sie die Zugriffe durch das Sperren betroffener Pfade oder ganzer Teilbereiche im Dokument synchronisieren [GBS02, PCK05, DH02]. Da die Sperren jedoch vor der eigentlichen Verarbeitung angefordert werden müssen, benötigen sie allgemeine Prädikatsperren, um eine hinreichend gute Parallelität zu ermöglichen. Des Weiteren benötigen sie stets vollständige und aktuelle Metadaten über das Dokument, wodurch ebenfalls zusätzlicher Aufwand entsteht und eine feingranulare Synchronisation nicht praktisch umsetzbar ist. Zusammenfassend bleibt festzuhalten, dass bislang kein effizient realisierbares Sperrverfahren existiert, das XQuery in der Konsistenzstufe *Serializable* ermöglicht.

## 2 Zielsetzung

Ein großer Kritikpunkt bei nahezu allen Ansätzen ist ihre mangelnde Orientierung an realen XDBMS-Architekturen. Weder das Zusammenspiel der einzelnen Komponenten der Anfrageverarbeitung noch die Integration der Verfahren in die verwendeten Operatoren werden hinreichend berücksichtigt. Die Erfahrungen mit relationalen DBMS haben allerdings gezeigt, dass sich gerade durch die Berücksichtigung systeminterner Strukturen und Operatoren häufig Optimierungsmöglichkeiten für viele Problemstellungen der Transaktionsisolation ergeben [Moh90, Lom93, Gra07]. Durch die getrennte Betrachtung der Mehrbenutzersynchronisation und der eigentlichen Anfrageverarbeitung werden somit viele Optimierungspotentiale schon vorab außer Acht gelassen. Eine mögliche Erklärung hierfür ist die bislang noch fehlende Referenzarchitektur.

Verglichen mit XDBMS ist die Transaktionsisolation in relationalen Systemen aufgrund der Einfachheit des Relationenmodells relativ unkompliziert. Das Synchronisationsmodell ergab sich dort fast auf natürlichem Wege, da alle Synchronisationsprobleme direkt auf die wenigen Grundoperationen abgebildet werden konnten. Diese Beobachtung zeigt, wie wichtig es für eine effiziente und umfassende Mehrbenutzersynchronisation ist, die Maßnahmen aus den Operationen des verwendeten Datenmodells abzuleiten. Im XML-Bereich wird dies bislang nur von den Protokollen der taDOM-Familie umgesetzt, wobei diese speziell für navigierende Zugriffe entwickelt wurden und ihnen deshalb die Unterstützung für mengenorientierte Zugriffe fehlt [HH04]. Bei ihnen liegt das Problem also nicht in einer grundsätzlich falschen Herangehensweise, sondern schlicht darin, dass ein rein navigierendes Verarbeitungsmodell nicht optimal für die Auswertung von deklarativen Anfragen geeignet ist.

Bei der Entwicklung eines XDBMS für den Mehrbenutzerbetrieb muss folglich zunächst geklärt werden, welches verbindliche Datenmodell der transaktionsorientierten Anfrageverarbeitung auf XML-Dokumenten zugrunde gelegt werden soll. Für dieses kann dann ein korrektes und optimal auf die internen Strukturen des Systems abgestimmtes *datenmodellbasiertes Sperrverfahren* entwickelt werden.

## 3 Das Synchronisationskonzept

Ein Datenmodell besteht aus einer Menge von Datenstrukturen, einer Menge von Operationen auf diesen Datenstrukturen und einer Menge von Integritätsbedingungen [Cod80]. Obwohl im Umfeld von XML und der vielen anhängigen Standards oft von Datenmodellen gesprochen wird, existiert für XML-Dokumente keine solche Modellierung beziehungsweise keine, die für die Verwendung in XDBMS geeignet ist [Wid99]. Für die gängige Interpretation der Dokumente als eine serialisierte Baumstruktur existiert mit der DOM-API zumindest eine Anwendungsschnittstelle, die definiert, wie innerhalb dieser Baumstruktur navigiert werden kann und welche Manipulationen möglich sind. Für die effiziente Verarbeitung von deklarativen Anfragen benötigt ein XDBMS neben solchen elementaren Navigationsoperationen jedoch auch die Möglichkeit, relevante Bereiche eines Dokumentes direkt über zusätzliche Indexstrukturen zu identifizieren. Ein Datenmodell muss daher auch festlegen, welche Art von Struktur-Wert-Kombinationen als Indexeinträge möglich sind.

Als Ergebnis dieser Herangehensweise erhält man eine Modellierung bestehend aus einer Menge von kontextsensitiven Operationen zum Navigieren und Modifizieren der Baumstruk-

tur sowie einer Menge kontextloser Leseoperationen, mit denen einzelne Knoten direkt über ihren Namen, ihren Textwert oder ihren Pfad im Dokument gelesen werden können.

Jede Operation stellt nun einen Zugriffsweg auf einen Dokumentknoten dar, der zu einer Verletzung der Isolationsbedingungen führen kann. Art und Anzahl der Zugriffswegen hängen dabei vom Typ des jeweiligen Knotens ab. In der gewählten Modellierung können Attributknoten so über sieben, Textknoten über zehn und Elementknoten sogar über vierzehn verschiedene Varianten erreicht werden. Abbildung 1 zeigt als Beispiel die Zugriffsmöglichkeiten auf einen Elementknoten.

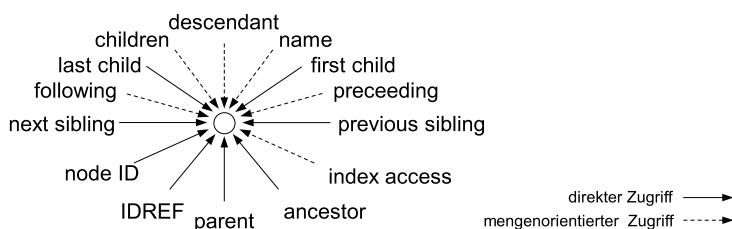


Abbildung 1: Zugriffsmöglichkeiten auf einen Elementknoten

Jeder Zugriffsweg des Datenmodells repräsentiert eine Eigenschaft (Prädikat) eines Dokumentknotens. Ein Sperrverfahren muss diese Prädikate nun einfach in einem gerichteten azyklischen Graphen anordnen und den gewählten Zugriffsweg vor einem Zugriff ausgehend von den Einstiegspunkten sperren [Gra78]. Zum Lesen eines Knotens genügt es dann, mindestens einen Zugriffsweg durch eine Lesesperre abzusichern, während zum Schreiben stets alle Zugriffswegen durch eine exklusive Schreibsperre abgesichert werden müssen.

Da die Verwendung eines einfachen hierarchischen IRIX-Sperrprotokolls zu einer geringen Parallelität und einer hohen Deadlock-Wahrscheinlichkeit bei gleichzeitigen Navigationsoperationen und Indexzugriffen führen würde, wird taDOM als Basis-Protokoll gewählt. Obwohl es bereits einen effizienten Schutz bei der navigierenden Verarbeitung von XML bietet [HHL06], müssen Erweiterungen und Anpassungen vorgenommen werden, um das Navigationsprotokoll mit Indexzugriffen abzustimmen. Dazu werden wie schon bei der Entwicklung des ursprünglichen taDOM einzelne Sperrmodi speziell an die neuen Anforderungen angepasst. Danach können die XML-spezifischen Knoten- und Kantensperren von taDOM relativ einfach mit bewährten Indexsperrverfahren aus relationalen DBMS (z. B. [Moh90, Lom93]) zu einem Gesamtkonzept integriert werden, das die Zugriffswegen des für XQuery konzipierten Datenmodells durch effiziente Routinen vollständig abdeckt.

## 4 Ausblick

Nach der Konzeptionen des Sperrverfahrens wird der vorgestellte Ansatz gerade in den XTC-Prototyp [XP] integriert, um Leistungsmessungen vorzunehmen und die Vorteile einer Harmonisierung von Systemarchitektur und Synchronisationsverfahren herauszustellen. Im Anschluss sind weitere Anpassungen geplant, um die Skalierbarkeit des Ansatzes weiter zu erhöhen.

## Literatur

- [Cod80] E. F. Codd. Data Models in Database Management. In *SIGMOD '80: Proc. Workshop on Data Abstraction, Databases and Conceptual Modelling, Pingree Park, Colorado*, Seiten 112–114, 1980.
- [DH02] Stijn Dekeyser und Jan Hidders. Path Locks for XML Document Collaboration. In *WISE 2002: Proc. 3rd International Conference on Web Information Systems Engineering, Singapore*, Seiten 105–114, 2002.
- [GBS02] Torsten Grabs, Klemens Böhm und Hans-Jörg Schek. XMLTM: Efficient Transaction Management for XML Documents. In *CIKM 2002: Proc. ACM CIKM International Conference on Information and Knowledge Management, McLean, VA, USA*, Seiten 142–152, 2002.
- [Gra78] Jim Gray. Notes on Data Base Operating Systems. In *Advanced Course: Operating Systems*, Seiten 393–481. Springer Verlag, 1978.
- [Gra07] Goetz Graefe. Hierarchical Locking in B-tree Indexes. In *BTW 2007: 12. Fachtagung Datenbanksysteme in Business, Technologie und Web des GI-Fachbereichs „Datenbanken und Informationssysteme“ (DBIS), Aachen, Germany*, Seiten 18–42, 2007.
- [HH04] Michael Peter Haustein und Theo Härder. Adjustable Transaction Isolation in XML Database Management Systems. In *XSym 2004: Proc. 2nd Second International XML Database Symposium on Database and XML Technologies, Toronto, Canada*, Seiten 173–188. Springer Verlag, 2004.
- [HHL06] Michael Peter Haustein, Theo Härder und Konstantin Luttenberger. Contest of XML Lock Protocols. In *VLDB '06: Proc. 32nd International Conference on Very Large Data Bases, Seoul, Korea*, Seiten 1069–1080, 2006.
- [HKM04] Sven Helmer, Carl-Christian Kanne und Guido Moerkotte. Evaluating lock-based Protocols for Cooperation on XML Documents. *SIGMOD Record*, 33(1):58–63, 2004.
- [JC06] Kuen-Fang Jea und Shih-Ying Chen. A High Concurrency XPath-based Locking Protocol for XML Databases. *Information & Software Technology*, 48(8):708–716, 2006.
- [Lom93] David B. Lomet. Key Range Locking Strategies for Improved Concurrency. In *VLDB '93: Proc. 19th International Conference on Very Large Data Bases, Dublin, Ireland*, Seiten 655–664. Morgan Kaufmann, 1993.
- [Moh90] C. Mohan. ARIES/KVL: A Key-Value Locking Method for Concurrency Control of Multi-action Transactions Operating on B-Tree Indexes. In *VLDB '90: Proc. 16th International Conference on Very Large Data Bases, Brisbane, Queensland, Australia*, Seiten 392–405. Morgan Kaufmann, 1990.
- [PCK05] Peter Pleshchakov, Petr Chardin und Sergei Kuznetsov. XDGL: XPath-Based Concurrency Control Protocol for XML Data. In *BNCOD 2005: Pro. 22nd British National Conference on Databases, Sunderland, UK*, Seiten 145–154. Springer Verlag, 2005.
- [Wid99] Jennifer Widom. Data Management for XML: Research Directions. *IEEE Data Engineering Bulletin*, 22(3):44–52, 1999.
- [XP] XTC-Project. AG DBIS, Technische Universität Kaiserslautern, <http://www.xtc-project.de>.

# DyOgeneS: Generierte Oberflächen in einem verteilten automatisierten System

**Thomas Bremes**

Universität Paderborn  
Institut für Informatik  
bremes@uni-paderborn.de

**Lars Scheuermann**

Universität Paderborn  
Institut für Informatik  
lars.scheuermann@uni-paderborn.de

**Abstract:** Im Zuge der Entwicklung von grafischen Oberflächen wird von Benutzern oft die Möglichkeit verlangt, Anordnungen und Menüstrukturen flexibel zu gestalten. Auch das Ändern der Struktur bei und besonders während der Nutzung der grafischen Oberflächen ist ebenfalls zeitaufwendig und fehlerträchtig. Seit einiger Zeit ist XML für diesen Bereich populär geworden. Vor allem wird XML für die Datenmodellierung und den Datenaustausch verwendet. Wegen vieler Vorteile, wie unter anderem dem Klartext und die gute Strukturierung kommt XML bei der Entwicklung von Benutzeroberflächen in Frage.

## 1 Das Problem

Eine Vielzahl von Softwarelösungen, die dem Wissens- und Informationsmanagement dienen, erleichtern dem Anwender die Bearbeitung von Arbeitsprozessen. Jedoch weisen diese Softwarelösungen erhebliche Schwächen im Bereich der Gewinnung und Archivierung von Wissensobjekten auf. Die bisher bekannten Ablagesysteme bieten hierfür schlechte Voraussetzungen. Ein mögliches Konzept ist das *Single Point of Information* (SPI) [HHM<sup>+</sup>06]. Bei diesem Konzept liegt der Schwerpunkt auf dem Erzeugen einheitlicher Speicherung und Suche von Wissensobjekten. Im Rahmen der Projektgruppe MoKEx III wird eine Software entwickelt, die auf dem SPI- Konzepts basiert. Um Wissensobjekte einheitlich ablegen zu können, werden diese mithilfe einer Benutzeroberfläche an ein zentrales System übergeben, welches die Wissensobjekte analysiert und ablegt. Bei der Analyse werden Metadaten aus dem Wissensobjekt gewonnen, mit Hilfe derer ein späteres Wiederfinden gewährleistet werden kann.

Das entwickelte Projekt setzte sich aus 6 Teilkomponenten zusammen. Diese teilen sich wie folgt auf:

Metadatenextraktion sowie semantische Analyse (MetaXsA) [SMR08] [MS08], Ablage der gewonnenen Metadaten und Semantiken (MeduSA) [DG08], einem Dokumenten-Management-System (DMS), einem Benutzerverzeichnis, einem Kommunikationsmanager (KoMa) [MMS08], sowie einer clientseitigen Benutzeroberfläche (DyOgeneS<sup>1</sup>). Jedes der Teilsysteme ist ein in sich geschlossenes System, welches zentral auf einem Server oder

---

<sup>1</sup>Dynamische Oberflächen und generierte Sichten

verteilt auf mehreren Servern initialisiert werden kann. Die Kommunikation zwischen den Teil- Systemen erfolgt über Web- Services [DKC05].

Im Folgenden wird die Problemstellung und der Lösungsansatz der im MoKEx III entstandenen Benutzeroberfläche näher erörtert.

DyOgeneS stellt eine grafische Benutzeroberfläche (GUI) bereit, mittels der die folgenden Funktionalitäten bedient werden können. Es bietet eine Suchoberfläche, eine generierte Oberfläche zur Anzeige der Suchergebnisse und eine Validations- und Modifikationsmöglichkeit für extrahierte Metadaten und Semantiken. Der Datentransfer zu DyOgeneS findet im plattform- und programmiersprachenunabhängigen XML-Format statt. Die Oberfläche besitzt statische und dynamische Elemente. Dabei sind Oberflächenelemente, wie zum Beispiel die Initial-GUI und die Auswahl-GUI, welche die allgemeinen Funktionalitäten bereitstellt statisch. Die Fehlermeldungs-GUIs die von externen Komponenten erstellt werden, werden dynamisch erzeugt. Möchte der Benutzer ein Dokument in das System hinzufügen, speichert er dieses Dokument in einem speziellen Verzeichnis, welches von DyOgeneS überwacht wird. Wenn eine Datei in dem Verzeichnis gespeichert wird, überträgt DyOgeneS die Datei an den KoMa. Neben dem Hinzufügen, Suchen und Bearbeiten von Dokumenten ist die Administration ein weiterer Bestandteil von DyOgeneS. Die Administration unterteilt sich hierbei in zwei Module. Zum einen soll die Möglichkeit bestehen bearbeitbare Metadaten zu wählen und zum anderen soll die Möglichkeit der Moduladministration geschaffen werden. Wird während der Laufzeit eine administrative Änderung an einem DyOgeneS-Client vorgenommen, so soll diese Änderungen an alle Clients übermittelt.

## 2 Der Lösungsansatz

Der Start von DyOgeneS erfolgt mit Hilfe der Java Web-Start-Technologie, die von einem Webserver auf den Client übertragen wird. Serverseitig wird eine “.jnlp”-Datei zum Download bereitgestellt. Diese Datei ist eine spezielle XML-Datei, die Informationen und Verweise auf die auszuführende “.jar”-Datei beinhaltet. Dies ermöglicht die entfernte Java-Anwendungen in webbasierten Umgebungen einfach zu starten. Im Gegensatz zu Java-Applets benötigen Java Web-Start-Anwendungen keinen Browser, um ausgeführt werden zu können.

Die Ordnerüberwachung wird mit Hilfe der in dem Paket JPoller<sup>2</sup> gekapselten Überwachungsfunktionalität umgesetzt. Wird ein neues Dokument durch den JPoller im Upload-Verzeichnis erkannt, so wird dieses Dokument mittels eines angebotenen Web-Service vom KoMa zum KoMa übertragen.

Die grafische Benutzeroberfläche dient dem Benutzer zur Interaktion mit dem KoMa, beziehungsweise mit der Server-Seite des Gesamtsystems. Um den angestrebten Funktionsumfang zu leisten sind unter anderem folgende Oberflächenelemente implementiert worden:

---

<sup>2</sup><http://jpoller.sourceforge.net/>

- **Login-Fenster:** Als erste Benutzerinteraktion nach dem Aufruf der `.jar`-Datei wird der Benutzer aufgefordert seinen Benutzernamen und sein Passwort zur Identifikation einzugeben. Dabei ruft DyOgeneS den Webservice `koma_login` auf
- **Suchanfrage:** Dabei gibt der Benutzer ein oder mehrere Suchbegriffe und die Suchmaske ein. Wird die Suche abgeschickt wird der KoMa-Web-Service `koma_search` aufgerufen. Als Rückgabe wird eine Liste von Metadaten geliefert mit der die Ergebnisoberfläche generiert wird.
- **Ergebnisoberfläche:** Für jedes Ergebnis hat der Benutzer folgende Auswahl:
  - Dokument bearbeiten: Das Dokument wird heruntergeladen und zum Bearbeiten geöffnet.
  - Dokument lesen: Das Dokument wird heruntergeladen und nur zum Lesen geöffnet.
  - Metadaten modifizieren: Die Metadaten eines Dokuments werden heruntergeladen und in DyOgeneS bearbeitet.
- Die Administrationseinstellungen von DyOgeneS bieten dem Administrator die Möglichkeit, Einstellungen der einzelnen Komponenten im laufenden Betrieb zu modifizieren. Hierbei wird von jeder Komponente eine Konfigurationsdatei bereitgestellt, die mittels einer Konfigurationsschnittstelle über Web-Services an die Komponenten verteilt wird. Die Konfigurationen werden von dem KoMa entgegengenommen, sowie an die einzelnen Komponenten verteilt. Einen Sonderfall stellt DyOgeneS hierbei selbst dar. Die Konfigurationsdatei wird analog zu den anderen Komponenten an den Kommunikationsmanager übermittelt. Durch diesen Mechanismus wirken sich veränderte Einstellungen in DyOgeneS *sofort* auf alle aktiven GUIs aus.

Für die Generierung von Oberflächen wird *SwiX<sup>ml</sup>*<sup>3</sup> verwendet. *SwiX<sup>ml</sup>* ist ein Java Projekt zur Beschreibung und Generierung von grafischen Benutzeroberflächen mittels XML-Dokumenten. Das *SwiX<sup>ml</sup>*-Paket konvertiert zur Programmlaufzeit diese XML-Dokumente in eine Swing-Benutzeroberfläche. Der XML-Dialekt orientiert sich an denen der Java-API `javax.swing`. Hierbei werden Objekte in Tags und Methodennamen in Attributen abgebildet. Durch die Verwendung von *SwiX<sup>ml</sup>* ist eine dynamische Generierung von Fehlermeldungsflächen möglich. Dies hat den Vorteil, dass von einzelnen Komponenten des Systems ein Fehlerfenster im *SwiX<sup>ml</sup>*-Dialekt generiert werden kann, diese XML-Beschreibung an Dyogenes geschickt wird und zur Laufzeit aus diesem Code ein Fenster mit unbekannten Inhalt erstellt.

---

<sup>3</sup><http://www.swixml.org/>



### 3 Zusammenfassung und Ausblick

DyOgeneS wurde nach den in der Problematik vorgestellten Voraussetzungen umgesetzt und fügt sich somit in das vorgestellte Verbundprojekt MoKEx III optimal ein. Dadurch wird dem Benutzer die Möglichkeit gegeben seine Dokumente in ein neuartiges System zu archivieren, diese Dokumente anhand seiner zuvor extrahierten Metadaten zu suchen um auf diese aufs Neue zugreifen zu können. Durch den Einsatz von Web-Services als Kommunikationsgrundlage, der Oberflächenbeschreibung im XML-Dialekt und die Umsetzung in Java ist es gelungen die Funktionalität von DyOgeneS Plattformunabhängig zu gestalten. DyOgeneS bietet in dieser Hinsicht den Benutzer eine ansprechende GUI mit einer auf das Projekt abgestimmten Funktionalität mit der der Benutzer mit dem System problemlos agieren kann und dient als erster Schritt zu einem nutzbaren Single Point of Information.

### Literatur

- [DG08] Steffi Döbbelin und Senol Gül. MeduSA: Umgang mit Metadaten in dynamischen Softwareumgebungen. In *eingereicht bei den Informatiktagen 2008*, 2008.
- [DKC05] Jean Dollimore, Tim Kindberg und George Coulouris. Verteilte Systeme. In *Pearson Studium*, 3. Auflage, 2005.
- [HHM<sup>+</sup>06] Knut Hinkelmann, Kai Holzweißig, Johannes Magenheim, Fabian Probst und Wolfgang Reinhardt. Linking Communities of Practice with Learning Communities in Computer Science Education. In *Proceedings of the 19th IFIP World Computer Congress (WCC2006)*, Santiago de Chile, August 20–25 2006.
- [MMS08] Robin Marterer, Matthias Moi und Tobias Schauerte. KoMa: Koordination verketteter Webservice-Aufrufe. In *eingereicht bei den Informatiktagen 2008*, 2008.
- [MS08] Christian Mletzko und Benedikt Schmidt. Automatische Extraktion von Metadaten aus verschiedenen Dateiformaten. In *eingereicht bei den Informatiktagen 2008*, 2008.
- [SMR08] Benedikt Schmidt, Christian Mletzko und Wolfgang Reinhardt. MetaXsA: Modulares Framework zur Metadatenextraktion durch semantische Dokumentanalyse auf unstrukturierten Textdokumenten. In *eingereicht bei den Informatiktagen 2008*, 2008.

# MeduSA: Umgang mit Metadaten in dynamischen Softwareumgebungen

Steffi Döbbelin

Senol Gül

Universität Paderborn  
Institut für Informatik  
steffid@mail.upb.de

Universität Paderborn  
Institut für Informatik  
senolg@upb.de

**Abstract:** Dieser Artikel beschreibt die MeduSA-Komponente aus der MoKEx-Projektserie, die in MoKEx III entstanden ist. Die Aufgabe dieser Komponente ist zum Einen die Speicherung und Suche der von der Komponente MetaxSA extrahierten Metadaten. Des Weiteren führt die Komponente bei der Initialisierung eine Art Selbstüberprüfung aus, bei der festgestellt wird, ob Erweiterungen am Speicherschema erfolgt sind. In diesem Fall kann MeduSA (Metadaten- und Semantiken-Ablage) vollständig autonom alle nötigen Änderungen in der zugrunde liegenden Datenbank vornehmen.

## 1 Das Problem

Im heutigen Medien-Zeitalter gibt es das bekannte Problem einer immer weiter zunehmenden Datenflut. Dabei wird es für einen Benutzer immer schwieriger zu wissen welche Informationen ihm direkt zur Verfügung stehen. Das Verbundprojekt MoKEx ([HHM<sup>+</sup>06]) hat es sich zur Aufgabe gemacht, dieses Problem mit einer automatisierten Extraktion von Metadaten zu begegnen. In der dritten Runde der MoKEx-Projektserie an der Universität Paderborn ist eine Gesamtanwendung entstanden, in der ein Benutzer nur noch die Dokumente einstellt, die er aufheben möchte. Die eigentliche Ablage der Dokumente, sowie die Speicherung der Metadaten wird vor ihm versteckt.

Dieses Paper wird sich im folgenden mit dieser Speicherung der Metadaten und den Herausforderungen, die sich aus dem dynamischen Kontext ergeben, beschäftigen. Die Metadaten werden in Form von XML-Dokumentenstrukturen definiert, welche im Fall der Weiterentwicklung der Extraktion mehr Metadaten, XML-Elemente, zurückliefert. Beispielsweise muss dann die Komponente selbstständig reagieren und Erweiterungen am Speichermodell vornehmen.

Die Kommunikation der einzelnen Komponenten im Architekturverbund erfolgt über Webservices (vgl. dazu [MSM08]), die XML-Daten mit den extrahierten Metadaten verschicken. Diese Daten sollen sowohl im XML-Original als auch, für eine schnellere Suche, relational gespeichert werden.

Die interne Bezeichnung der betreffenden Komponente ist MeduSA (Metadaten- und Semantiken-Ablage). Hierbei ist es wichtig zu betonen, dass es sich bei MeduSA nicht nur um

eine reine Datenbankanwendung handelt, da die Komponente auch Aufgaben der autonomen Datenbank-Administration übernimmt.

## 2 Ein Lösungsansatz

Zusammenfassend kann die Aufgabenstellung der MeduSA-Komponente in folgenden Punkten dargestellt werden:

Da die zu speichernden Metadaten, welche von MetaxSA extrahiert wurden, im XML-Format geliefert werden und auch in diesem abgelegt werden sollen, muss die Datenbank **XML-fähig** sein. Um die Suche in den Metadaten zu beschleunigen sollen diese zusätzlich noch **relational** in der Datenbank gespeichert werden. Für eine einheitliche Sicherung der Metadaten kommen damit nur Hybriddatenbanken in Frage.

Als Softwarelösung der MeduSA-Komponente wurde die DB2 von IBM in der Version 9<sup>1</sup> gewählt, da diese ein xml-fähiges RDBMS besitzt. Diese Hybrid-Datenbank unterstützt sehr einfacher Art und Weise die Sicherung von XML-Daten im Original und gleichzeitig mit einem Dekompositionsmechanismus die vollständige relationale Ablage.

Für den Dekompositionsmechanismus werden, durch die MeduSA-Komponente, Tabellenspezifische Informationen, wie Tabellennamen und Spalten, einem XML-Schema beigefügt. Das Schema wird dann in der Datenbank registriert. Der Mechanismus prüft die Metadaten im XML-Format auf die Gültigkeit mit dem XML-Schema. Nach einer erfolgreichen Validierung wird jedes XML-Element auf eine Spalte abgebildet. Zusätzlich wird die XML-Datei in einer xml-fähigen Spalte gespeichert.

Wenn die Metadatenextraktion [MS08] und die semantische Analyse [SMR08] in der Weiterentwicklung mehr Metadaten liefern, muss MeduSA damit zurecht kommen. Das heißt, MeduSA bekommt in der Initialisierungsphase von der KoMa-Komponente ein XML-Schema und überprüft damit, ob alle Datenbanktabellen noch "ausreichend" sind. Gegebenenfalls werden noch fehlende Spalten angelegt und die vorhandene Referenztabelle erweitert. Die Referenztabelle enthält Informationen darüber welche XML-Pfade in welchen Spalten abgebildet wurden. Existieren alle benötigten Spalten werden die Beziehungen zwischen den einzelnen XML-Elementen und den Spalten der Datenbanktabellen im XML-Schema selbst vermerkt und das neu annotierte XML-Schema wird in der Datenbank registriert. Im laufenden Betrieb kann so leicht festgestellt werden, wo was relational abgelegt werden soll.

Eine Eigenschaft von XML-Elementen ist es, dass sie mehrfach auftreten können. So auch in den XML-Daten, die vom KoMa kommen. Dies muss besonders in dem "Mapping" auf die Spalten der Datenbanktabellen berücksichtigt werden. Als Lösung für MeduSA wurde eine Auslagerung in weitere Tabellen umgesetzt, bei der für jedes einzelne Auftreten jeweils ein eigener Datensatz in diesen Tabellen angelegt wird.

---

<sup>1</sup> siehe dazu <http://www-306.ibm.com/software/data/db2/9/>

Mit Hilfe einer intern definierten Suchsyntax kann eine allgemeine Suche oder eine erweiterte Suche im XML-Format beschrieben werden. Diese Suchanfrage im XML-Format wird über ein Webservice an die MeduSA-Komponente geschickt. Die Syntax liefert Informationen über die zu suchenden XML-Pfade, Suchbegriffe und falls vom Benutzer eingegeben auch boolesche Operatoren. Die MeduSA-Komponente wertet die Suchanfrage mit Hilfe der Referenztabelle aus und generiert daraus eine SQL-Anfrage. Hierbei werden die booleschen Operatoren mitbetrachtet. Eine erfolgreiche Suche liefert, wieder über einen Webservice, eine Liste mit den Metadaten im XML-Format.

Da eventuell nicht immer alle Metadaten extrahiert werden, um zu einem positiven Suchergebnis zu führen, speist MeduSA zusätzlich noch den Klartext der Originaldatei im Anschluss der Dekomposition in Lucene<sup>2</sup> ein. Wenn eine Suchanfrage in den reinen Metadaten nicht zum Erfolg führt, wird diese Suchanfrage an Lucene und damit einem Volltextindex weitergeleitet. Der Volltextindex wird nach jeder Dekomposition um den jeweiligen Klartext des Dokumentes erweitert. Dies dient lediglich der Sicherstellung, dass ein einmal eingestelltes Dokument immer wiedergefunden werden kann. Im Normalfall reicht eine Suche in den relational gespeicherten Daten aus.

Neben Anlegen neuer Datensätze bietet MeduSA auch das Abrufen und Ändern einzelner Metadaten über entsprechende Webservices.

### 3 Zusammenfassung und Ausblick

In MoKEx III ist mit MeduSA eine komplexe Datenbankanwendung entstanden, die in der Lage ist eine mit Metadaten befüllte XML-Datei in einem RDBMS abzulegen. MeduSA kann autonom auf Änderungen reagieren und leitet Suchanfragen zuerst intern an relationale Datenbanktabellen. Bei Misserfolg wird diese Anfrage an Lucene weitergeleitet.

In der Perspektive kann MeduSA auch auf andere Datenbanken ausgeweitet werden. Dazu muss der Teil, den die DB2 mit dem Dekompositionsmechanismus übernimmt, vom Rest von MeduSA getrennt werden.

### Literatur

- [BS08] Thomas Bremes und Lars Scheuermann. DyOgeneS: Generierte Oberflächen in einem verteilten automatisierten System. In *eingereicht bei den Informatiktagen 2008*, 2008.
- [HHM<sup>+</sup>06] Knut Hinkelmann, Kai Holzweißig, Johannes Magenheimer, Fabian Probst und Wolfgang Reinhardt. Linking Communities of Practice with Learning Communities in Computer Science Education. In *Education for the 21st Century — Impact of ICT and Digital Resources*, IFIP Computer Science, Seiten 83–92. Springer Berlin / Heidelberg, 2006.

---

<sup>2</sup> siehe dazu <http://lucene.apache.org/java/docs/index.html>

- [MS08] Christian Mletzko und Benedikt Schmidt. Automatische Extraktion von Metadaten aus verschiedenen Dateiformaten. In *eingereicht bei den Informatiktagen 2008*, 2008.
- [MSM08] Robin Marterer, Tobias Schauerte und Matthias Moi. KoMa: Koordination verketteter Webservice-Aufrufe. In *eingereicht bei den Informatiktagen 2008*, 2008.
- [SMR08] Benedikt Schmidt, Christian Mletzko und Wolfgang Reinhardt. MetaXsA - Modulares Framework zur Metadatenextraktion durch semantische Dokumentanalyse auf unstrukturierten Textdokumenten. In *eingereicht bei den Informatiktagen 2008*, 2008.

# **Laufzeitadaption in einem serviceorientierten Komponentenmodell**

Thomas Fischer

University of Erlangen-Nuremberg  
Department of Computer Science  
Computer Science 6 (Data Management)  
Martensstrasse 3  
D-91058 Erlangen  
k.p.t.fischer@gmx.de

## **1 Einleitung**

In den vergangenen Jahren erlangte der Begriff der „serviceorientierten Architektur“ (SOA) [PH07, HS05, Tho05, Jon05] immer mehr an Bedeutung. Eine Eigenschaft solcher lose gekoppelten Systeme ist es, dass die Services dynamisch zu jeder Zeit aktiv oder inaktiv werden können. Ein inaktiver Service beeinflusst aber möglicherweise die Verfügbarkeit aller davon abhängigen Applikationen. Dies ist zu der Forderung nach Hochverfügbarkeit, die vor allem im Enterpriseumfeld gestellt wird, kontraproduktiv. Daher ist es das Ziel, die Verfügbarkeit von Anwendungen so hoch wie möglich zu halten.

In dieser Arbeit wird die CoBRA (Component Based Runtime Adaptable) Architektur vorgestellt, welche einen Ansatz verfolgt, der die Verfügbarkeit von Services mittels Laufzeitadaption erhöht. Dies kann Ausfallzeiten zum Beispiel durch Wartungsarbeiten entgegenwirken.

Das entwickelte Framework basiert dabei auf einem sog. *serviceorientierten Komponentenmodell* [CH04]. Dieses Modell stellt eine Verbindung von Komponenten im klassischen Sinne [Szy02, Som95] und den Ansätzen einer SOA dar.

## **2 Laufzeitadaption**

Laufzeitadaption [Blo83, MG05, MMH03] in einem serviceorientierten Komponentenmodell beschreibt den Vorgang, dass ein Service sich zur Laufzeit an neue Anforderungen anpassen kann, ohne dabei die Ausführung abhängiger Applikationen zu beeinflussen. Durch die sich aus der Komponentendefinition ergebende Teilung eines Services in eine *Servicebeschreibung* und eine *Serviceimplementierung* lässt sich die Laufzeitadaption in der gewählten Umgebung durch einen Austausch der Serviceimplementierung realisieren,

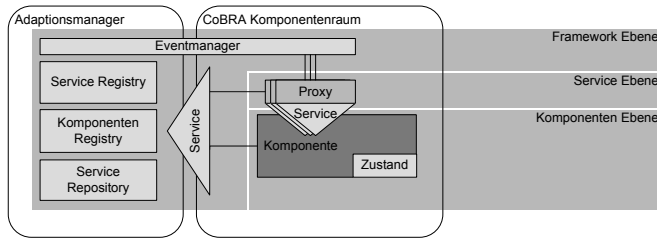


Abbildung 1: CoBRA Architektur

ohne dabei die Gültigkeit der Servicebeschreibung zu verletzen. Daraus ergeben sich eine Reihe von Anforderungen an eine laufzeitadaptierbare Umgebung:

- *Transparenz*: Der Austausch einer Serviceimplementierung muss transparent gegenüber nutzenden Services oder Applikationen sein.
- *Atomarität*: Der Austausch muss atomar erfolgen, um eventuelle Nebenläufigkeiten von zwei Serviceimplementierungen zu vermeiden.
- *Zustandserhaltung*: Der Zustand der Serviceimplementierung muss über den Ersetzungsvorgang erhalten bleiben.
- *Koordination*: Der gesamte Ersetzungsvorgang muss global koordiniert werden um eine Einhaltung der Anforderungen zu gewährleisten.

Diese Anforderungen wurden im Rahmen der CoBRA-Architektur umgesetzt, wohingegen Ansätze wie [Blo83] oder [MG05] die geforderte Transparenz nicht vollständig umsetzten.

### 3 Die CoBRA Architektur

Die OSGi-Service-Plattform [OSG05] dient der Umsetzung des CoBRA-Ansatzes als Basis und stellt mittels eines serviceorientierten Komponentenmodells die notwendigen Voraussetzungen für die Umsetzung einer Laufzeitadaptionsumgebung zur Verfügung. Aufbauend auf der OSGi-Schichtenarchitektur, wurde eine weitere Schicht entwickelt, welche das OSGi-Modell um die Laufzeitadaptionsfunktionalität erweitert, aber auch mit klassischen OSGi-Komponenten interagiert.

Abbildung 1 zeigt den Aufbau der CoBRA-Architektur. Eine CoBRA-Komponente besteht dabei aus einem Zustand und einer Anzahl von Services. Eine Applikation setzt sich aus mehreren über Services miteinander interagierenden Komponenten zusammen.

Die eingeführte Transparenzbedingung wird dabei in der CoBRA Architektur mittels der Java Dynamic Proxies [Sun97] umgesetzt. Diese für den Servicenutzer vollkommen transparent generierten Proxies replizieren jeden Service über den Komponenten miteinander

interagieren. Fordert nun ein Servicenutzer einen Service über die *Service Registry* [HS05] an so erhält er ein Proxyobjekt statt des eigentlichen Serviceobjekts. Die daraus resultierende Indirektion ermöglicht die Sicherung der Atomarität mittels der Umsetzung des *Protection Proxy Patterns* [GHJ95]. Die Notwendigkeit der Einführung dieses in [Blo83] auch *Guardian Objekt* genannten Mechanismus ergibt sich aus der Forderung nach einem atomaren Vorgang, aber auch aus der gewählten Zustandsübergangsstrategie der *Weak Migration* [IKWK00]. Das bedeutet, dass nur zu Zeitpunkten an welchen sich eine Serviceimplementierung in einem konsistenten Zustand befindet eine Adaption durchgeführt werden kann. In der CoBRA Architektur kommt dies einem blockierenden Verhalten von Serviceaufrufen gleich, sobald ein Adaptionsvorgang initiiert worden ist. Befindet sich gerade ein Service in Benutzung, so wird der Adaptionsvorgang bis zu dessen Beendigung blockiert. Demnach definiert sich der konsistente Zustand eines Services darin, dass sich keine seiner Methoden in Ausführung befindet.

Der gesamte Vorgang einer Adaption wird durch einen *Adaptionsmanager* (siehe Abbildung 1) koordiniert. Dieser sichert sowohl den Zustandsübergang zwischen den an der Adaption beteiligten Serviceimplementierungen als auch die Synchronisation der am Vorgang beteiligten Proxies zur Gewährleistung der Transparenz und der Atomarität. Diese Koordination findet auf drei Ebenen statt: Auf Serviceebene wird jeder Service durch einen Proxy repliziert, der durch ein Eventsystem vom Adaptionsmanager gesteuert wird. Auf Komponentenebene wird ein Zustandsobjekt gehalten welches bei einem Adaptionsvorgang durch das Framework extrahiert wird und in die adaptierte Komponente injiziert wird. Auf Framework Ebene wird aus diesem Grund eine *Komponenten Registry* gehalten, welche die Zustandsobjekte der Komponenten verwaltet. Darüberhinaus wird ein Service Repository eingeführt (vgl. [Obj02]), welches die Servicebeschreibungen aller zur Verfügung stehenden Services enthält, auch die der momentan nicht aktiven Services. Damit ist der Adaptionsmanager in der Lage den gesamten Vorgang zu koordinieren.

Der Ablauf eines Adaptionsvorgangs ist dabei in drei Phasen gegliedert: die *Sicherungsphase*, die *Umschaltphase* und die *Wiederherstellungsphase*. In der Sicherungsphase werden zunächst alle Proxy-Instanzen der angebotenen Services invalidiert. Damit blockieren alle Methodenaufrufe auf diesen Proxies. Anschließend werden die Zustandsobjekte extrahiert und vom Framework zwischengespeichert. In der folgenden Umschaltphase wird die alte Komponente aus dem Framework entfernt und die neue Komponente installiert. Damit beginnt die Wiederherstellungsphase. In dieser letzten Phase wird der Zustand in die neue Komponente eingebracht. Der Adaptionsvorgang endet mit der Reaktivierung aller blockierten Proxies und deren Methoden.

## 4 Zusammenfassung

Die vorgestellte CoBRA Architektur, die im Rahmen dieser Arbeit prototypisch implementiert wurde, setzt einen möglichen Ansatz zur Laufzeitadaption um: Adaption durch Ersetzung. Dieser Ansatz ermöglicht es die Verfügbarkeit von Services stark zu erhöhen, da im Gegensatz zu anderen Komponenten- und SOA-Architekturen nur sehr wenige Wartungs- und Änderungsvorgänge einen Neustart der Applikation erfordern.



## Literatur

- [Blo83] Toby Bloom. *Dynamic Module Replacement in a Distributed Programming System*. Dissertation, Massachusetts Institute of Technology, März 1983.
- [CH04] Humberto Cervantes und Richard S. Hall. Autonomous Adaptation to Dynamic Availability Using a Service-Oriented Component Model. In *ICSE '04: Proceedings of the 26th International Conference on Software Engineering*, Seiten 614–623, Washington, DC, USA, 2004. IEEE Computer Society.
- [GHJ95] Erich Gamma, Richard Helm und Ralph Johnson. *Design Patterns. Elements of Reusable Object-Oriented Software*. Addison-Wesley Professional Computing Series. Addison-Wesley, 1995. GAM e 95:1 1.Ex.
- [HS05] Michael N. Huhns und Munindar P. Singh. Service-Oriented Computing: Key Concepts and Principles. *IEEE Internet Computing*, 9(1):75–81, 2005.
- [IKWK00] Torsten Illmann, Frank Kargl, Michael Weber und Tilmann Krüger. Migration of Mobile Agents in Java: Problems, Classification and Solutions. In *Proceedings of the MAMA'00, Wollongong, Australia*, 2000.
- [Jon05] Steve Jones. Toward an Acceptable Definition of Service. *IEEE Software*, 22(3):87–93, 2005.
- [MG05] A. Mukhija und M. Glinz. Runtime Adaptation of Applications through Dynamic Re-composition of Components. In *Proc. of 18th International Conference on Architecture of Computing Systems*, 2005.
- [MMH03] J. Matevska-Meyer und W. Hasselbring. Enabling reconfiguration of component-based systems at runtime. In J. van Gurp und J. Bosch, Hrsg., *Proceedings of Workshop on Software Variability Management*, Seiten 123–125. University of Groningen, Februar 2003.
- [Obj02] Object Management Group. CORBA/IIOP Specification, Version 3.0, März 2002.
- [OSG05] OSGi Alliance. OSGi Service Platform Core Specification, Release 4, August 2005.
- [PH07] Mike P. Papazoglou und Willem-Jan Heuvel. Service oriented architectures: approaches, technologies and research issues. *The VLDB Journal*, 16(3):389–415, 2007.
- [Som95] Ian Sommerville. *Software engineering (5th ed.)*. Addison Wesley Longman Publishing Co., Inc., Redwood City, CA, USA, 1995.
- [Sun97] Sun Microsystems. Java Core Reflection API and Specification, Februar 1997.
- [Szy02] Clemens Szyperski. *Component Software: Beyond Object-Oriented Programming*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 2002.
- [Tho05] Erl Thomas. *Service-Oriented Architecture*. Prentice Hall PTR, Upper Saddle River, 2005.

# Redaktionsprozesse in einer Serviceorientierten Architektur

Thomas Fitting

Fachhochschule Hannover

Fakultät I

tfitting@freenet.de

Betreuer/in der Arbeit: Prof. Dr. Volkert Brosda

**Abstract:** Serviceorientierte Architekturen (SOA) sind mittlerweile kein neues Paradigma mehr. Allerdings kristallisiert sich immer mehr heraus, welche Vorteile die lose Kopplung durch Webservices in Hinblick auf die Geschäftsprozessmodellierung (BPM) bringt. An zentraler Stelle können Geschäftsprozesse eingeführt und optimiert werden. Ziel dieser Arbeit soll es sein, die Geschäftsprozesse für ein topicbasiertes Redaktionssystem \textbf{SOA~RS} zu modellieren. Nach diesem Ansatz werden Struktur, Inhalt und Layout strikt getrennt um, die Wiederverwendung zu ermöglichen. Gleichzeitig werden diese Bestandteile jedoch generisch als Assets verwaltet.

## XML als Basis

Basis für die Serviceorientierte Architektur dieser Arbeit ist das neutrale Austauschformat XML. Daher soll dies auch die Grundlage für die in der Arbeit modellierten Prozesse sein. XML wird über dies auch als Format für Dokumentationen genutzt. Vorteile dieses Formates sind u.a.:

- **Vom Layout unabhängige Daten** - Die ausgegliederten Layout-Angaben können wieder verwendet und getrennt von den Daten weiterentwickelt werden.
- **Einfacher Zugriff** - Mit der Sprache XPath kann mit einem Pfadausdruck durch den XML-Baum navigiert werden. Regelbasierte Verarbeitung ist durch XSLT möglich, sequenzieller Zugriff durch XQuery.
- **Überprüfbarkeit der Daten** - Durch ausgelagerte Beschreibungen (in den Sprachen DTD, XML Schema und RELAX NG) lässt sich die Gültigkeit der XML-Daten prüfen.

- **Unabhängigkeit** - XML ist durch vereinheitlichte Schnittstellen (DOM, SAX) mit jeder Programmiersprache les- und schreibbar. XML ist also unabhängig von Betriebssystemen und Programmiersprachen. Als reines Textformat kann XML prinzipiell auch mit jedem Texteditor geschrieben werden.

## SOA als Technik

In Serviceorientierten Architekturen lösen modulare, funktionale Dienste monolithische Anwendungen ab. Jeder Dienst stellt in einem solchen System seine eigene Schnittstellenbeschreibung (in Form eines WSDL-Dokumentes) bereit. Durch diesen Aufbau werden die Komponenten der Architektur austauschbar und Geschäftsprozesse flexibler.

Serviceorientierung ist außerdem eine Architektur auf einer sehr hohen Abstraktionsstufe. Das erlaubt maschinenlesbare Geschäftsprozesse explizit und isoliert in BPEL-Beschreibungen zu definieren. Für die Modellierung ist BPMN geeignet, weil die Sprachmittel speziell für Prozessmodellierung in SOA ausgelegt sind.

SOAP ist ein Protokoll für Webservices. Webservices sind nicht auf den Standard SOAP beschränkt, doch obwohl SOAP das umfangreichste Protokoll für Webservices ist, besteht für diesen Standard die meiste Unterstützung in den (teilweise freien) Tools. Außerdem lassen sich in SOAP sowohl andere XML-Dialekte als auch binäre Daten einbetten.

## Topics und Assets

Das Redaktionssystem SOA~RS basiert auf dem Paradigma Topicorientierung. Nach diesem Paradigma besteht ein Dokument aus

- **Topics**, XML-Fragmente deren internes Schema ihrer Funktion entspricht, und
- **Topicmaps**, hierarchischen Informationsgebilde, die Topics referenzieren und so die Superstruktur der Dokumentation repräsentieren.

Der Content wird in SOA~RS unabhängig vom Dokument erstellt, wodurch Inhalte wiederverwendet werden können und das Erzeugen des Dokumentes in den Hintergrund gerückt wird. Ein XML-Schema diktiert die semantische Struktur der Topics. In SOA~RS werden Topicmaps, Topics und Grafiken als Assets abstrahiert. Das Resultat ist Wiederverwendung von Metadaten und sogar von Prozessen.

## Geschäftsprozesse

Der Schwerpunkt der Arbeit sind die der Technischen Redaktion entstammenden Geschäftsprozesse. Szenario ist die Erstellung einer Dokumentation unter Berücksichtigung verschiedener Rollen. Diese Redaktionelle Prozesse unterscheiden sich zwar abhängig von Medium und Dokumenttyp, haben jedoch auch viele Gemeinsamkeiten. Die entwickelten Prozesse sind daher zwar exemplarisch, können aber weiterentwickelt und optimiert werden.

## Ergebnisse der Arbeit

Ergebnis der Arbeit sind

- die Konzeption des Redaktionssystems SOA~RS,
- die Modellierung der Prozesse in der Notation BPMN und
- ein praktischer Teil mit exemplarischen Implementierungen.

Bei der Konzeption gehe ich auf Kommunikation, Metadaten, das Rollenkonzept, Sperrprinzipien und Qualitätssicherung ein. Die Modellierung der Prozesse richtet sich nach dem „Content Life Cycle“ und bildet dessen Phasen, von dem Entdecken des Bedarfs bis zur Archivierung der erstellten Dokumentation, ab. Den Abschluss bildet die Identifikation der für dieses System benötigten Komponenten.

Der Praktische Teil der Arbeit besteht aus dem selbst implementierten Messaging Server, exemplarisch generiertem Code für Client und Server aus der WSDL-Beschreibung dieses Messaging Servers und ein XML-Schema für die Metadaten. Der Messaging Server selbst ist eine in der Scriptsprache Python entwickelten Webanwendung, welche auf dem Web Framework **TurboGears** dessen Erweiterung **TGWebServices** und der Bibliothek **xmppy** besteht.

## Literaturverzeichnis

- [Bar] Wachsen die BPM- und Workflow-Lager zusammen?,  
<http://www.bpm-guide.de/articles/66>, 10.10.2002
- [BCS05] BEN COLLINS - SUSSMANN, Brian W. Fitzpatrick & C. Michael P.:  
Versionskontrolle mit Subversion. O'Reilly, 2005
- [Bec] BECKETT, Dave: Ressource Description Framework.  
<http://www.w3.org/RDF/>, Abruf: 31.09.2007
- [Chr] CHRISTENSEN, Erik: Web Services Description Language 1.1.  
<http://www.w3.org/TR/wsdl>, Abruf: 04.09.07
- [Clo07] CLOSS, Sissi: Single Source Publishing.  
Software & Support Verlag GmbH, 2007
- [Gro] GROUP, W3C HTML W.: XHTML 1.0. <http://www.w3.org/TR/xhtml1>,  
Abruf: 26.09.2007

- [Ini] INITIATIVE, Dublin Core M.: Dublin Core Metadata Element Set, Version 1.1.  
<http://dublincore.org/documents/dces/>, Abruf: 21.09.2007
- [Kro03] KRONZ, Sabine: Content Management. EUL Verlag, 2003
- [Mel06] MELZER, Ingo: Ser vice Orientier te Architekturen. Spektrum Akademischer Verlag, 2006 (2)
- [MP] MICHAEL PRIESTLEY, Johann H. Rober t D. Anderson: DITA Specification.  
<http://docs.oasis-open.org/dita/v1.1/CS01/over view/over view.html>, Abruf: 17.09.2007
- [N.F] N. FREED: Multipurpose Internet Mail Extensions (MIME).  
[www.ietf.org/rfc/rfc2045.txt](http://www.ietf.org/rfc/rfc2045.txt), Abruf: 20.09.2007
- [NM] NILOMITRA, Yves L.: Simple Object Access Protocol 1.1.  
<http://www.w3.org/TR/soap>, Abruf: 04.09.07
- [spe] Business Process Modeling Notation Specification.  
<http://www.omg.org/docs/dtc/06-02-01.pdf>, Abruf: 04.09.07
- [TF07] THILO FROTSCHER, Dapeng W. Marc Teufel T. Marc Teufel:  
Java Web Services mit Apache Axis2. Software & Support Verlag GmbH, 2007
- [Tha] THATTE, Satish: Business Process Execution Language for Web Services.  
<http://download.boulder.ibm.com/ibmd/pub/software/dw/specs/ws-bpel/ws-bpell.pdf>,  
Abruf: 04.09.07
- [Tho] THOMPSON, Henry S.: XML Schema Part 0: Primer Second Edition.  
[www.w3.org/TR/xmlschema-0/](http://www.w3.org/TR/xmlschema-0/), Abruf: 25.09.2007
- [Ver02] VERSTEEGEN, Gerhard: Management-Technologien. Springer-Verlag  
Berlin Heidelberg New York, 2002
- [AB00] Abel, K.; Bibel, U: Formatierungsrichtlinien für Tagungsbände. Format-Verlag, Bonn, 2000.
- [ABC01] Abraham, N.; Bibel, U.; Corleone, P.: Formatting Contributions for LNI. In (Glück, H.I. Hrsg.): Proc. 7th Int. Conf. on Formatting of Workshop-Proceedings, New York 1999. Noah & Sons, San Francisco, 2001; S. 46-53.
- [Az99] Azubi, L. et.al.: Die Fußnote in LNI-Bänden. In (Glück, H.I.; Gans, G., Hrsg.): Formattierung leicht gemacht – eine Einführung. Format-Verlag, Bonn, 1999; S. 135-162
- [Ez99] Ezgarani, O.: The Magic Format – Your Way to Pretty Books, Noah & Sons, 2000.

# Process based Metadata for Semantic Desktop Search

Florian Gerecke

`mail@floriangerecke.de`

**Abstract:** Information about a user's activities that result in modification of resources like documents, emails, or calendar entries are usually not stored permanently. One part of these activities are file system events like closing a file after writing to it. Some of these events are triggered by applications that are invoked by a user. In this work, metadata related to file modifying processes is stored as RDF statements. This metadata enables the user to either search for a file by using the name of the application or to search for process based metadata related to a given file.

## 1 Introduction

Finding previously stored data has become a serious problem for many computer users, because of the increasing amount of data that is stored on a computer. While storing the data is rarely a problem—since in most cases storage capacity can be sufficiently extended—the number of options to find a specific resource does not increase.

The vision of a semantic desktop is, that there are tools, that are able to answer questions like: 'Who sent me this file?' or 'Where is the file that has been created by user *U* with Program *P1* last year and that I modified with Program *P2* last month?'. While the first question is discussed by many semantic desktop projects<sup>1</sup> like e.g. Gnowsis,<sup>2</sup>[SGK<sup>+</sup>06] the latter belongs to the objective of this work.

The approach of this work is to store information about processes that write files and about the user that invoked the process. The corresponding file system event is regarded as a relationship between an application, a person, and a file and has the potential of being related to other resources on a semantic desktop. In the following section the approach and its realisation is explained.

## 2 The approach: process based metadata

Metadata for a semantic repository should be retrieved in the background without the need for user interaction because it would interrupt the users workflow. The approach of this work follows this requirement, and the system is designed to collect the information

---

<sup>1</sup><http://www.semanticdesktop.org/>

<sup>2</sup><http://www.gnowsis.org/>

automatically in the background.

Semantic desktop applications can be divided into user applications that generate metadata about their own activities, and metadata generators that are independent of user applications. While the first is interesting in respect of the potentially strong significance of the metadata, the latter is also important, because most of the applications that are used are not aware of a semantic desktop[CGG<sup>+</sup>05]. The approach of this work is application independent.

A process on a computer is a program in execution. During the time of execution many pieces of information about a process are available. These include the process identification number, the user identification of the process owner, the process name, the commandline, the open file descriptors and information like CPU or memory usage. Modern operating systems also provide the user with the possibility to be notified about file system events. At first glance this seems to be enough to collect the needed information and to store it. Nevertheless mechanisms for monitoring file system events—like 'inotify' in the Linux kernel—do not provide information about the process that caused the event. To obtain the missing information, a new module with almost the same interface like the inotify module was added to the kernel[BC05]. In addition to the functionality of inotify the new module provides the process name and the user ID of that process. The new module is used in the same way as the original inotify module[Lov07].

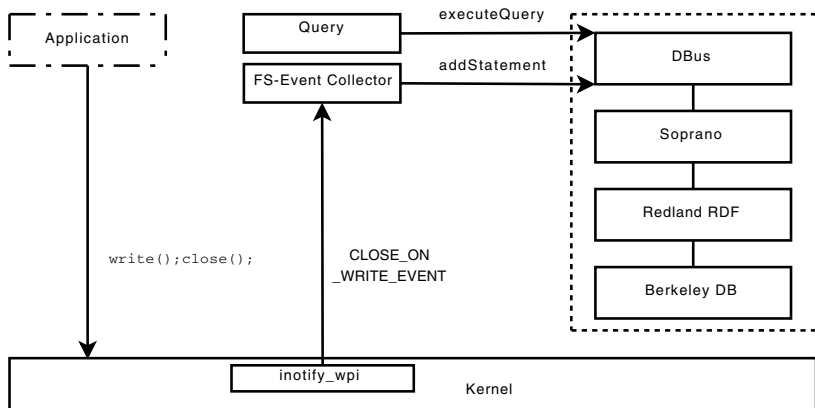


Figure 1: Components of the implemented system and its environment

In figure 2 the notification by the inotify module with process information is shown. The reported file system event is processed by a metadata generator that assembles the information into RDF statements. These statements are added to a central semantic repository. In this work the RDF framework Soprano<sup>3</sup> is used. The Soprano server is accessed via its DBus<sup>4</sup> interface. The Redland RDF<sup>5</sup> backend plugin of Soprano provides a persistent

<sup>3</sup><http://soprano.sourceforge.net/>

<sup>4</sup><http://dbus.freedesktop.org/>

<sup>5</sup><http://librdf.org/>

storage in Berkeley DB<sup>6</sup> format. Soprano is used by Nepomuk-KDE<sup>7</sup>, the semantic parts of the desktop environment KDE4<sup>8</sup>.

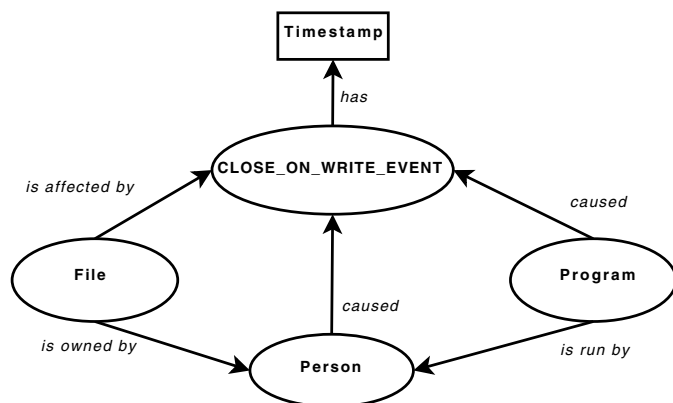


Figure 2: Ontology of a file system event with process information

The statements about the file system events follow the ontology that is shown in figure 2. The predicate 'is owned by' between 'file' and 'person' is not required by the file system event, but illustrates the potential for further relations in a semantic desktop environment. Other applications that are aware of the semantic desktop write information to the repository which may be related to this process based metadata.

The implemented software to handle this process based metadata has been tested in the desktop environment KDE4.

### 3 Conclusion

This work is a contribution to the ideas of the semantic desktop. It addresses a small but up to now unconsidered part of the information that a desktop system could provide. The implemented software provides a history for every watched file that contains information about when a file was modified and who modified it with which program. This process based metadata can be related to other resources and has the potential for an improved desktop search.

<sup>6</sup><http://www.oracle.com/technology/products/berkeley-db>

<sup>7</sup><http://nepomuk-kde.semanticdesktop.org>

<sup>8</sup><http://www.kde.org>



## References

- [BC05] Daniel P. Bovet and Marco Cesati. *Understanding the Linux Kernel*. O'Reilly Media, Inc., 2005.
- [CGG<sup>+</sup>05] Paul A. Chirita, Rita Gavriloaie, Stefania Ghita, Wolfgang Nejdl, and Raluca Paiu. *Activity Based Metadata for Semantic Desktop Search*. 2005.
- [Lov07] Robert Love. *Linux System Programming*. O'Reilly Media, Inc., 2007.
- [SGK<sup>+</sup>06] Leo Sauermann, Gunnar Aastrand Grimnes, Malte Kiesel, Christiaan Fluit, Heiko Maus, Dominik Heim, Danish Nadeem, Benjamin Horak, and Andreas Dengel. Semantic Desktop 2.0: The Gnowsis Experience. In *The Semantic Web - ISWC 2006*, volume Volume 4273/2006, pages 887–900. Springer Berlin / Heidelberg, 2006.

# Availability Guarantees for Data Items in Unstructured P2P Networks

Stefan Holder

Universität des Saarlandes  
FR 6.2 - Informatik  
sholder@mpi-inf.mpg.de

Art der Arbeit: Masterarbeit

Betreuer der Arbeit: Mauro Sozio (PhD)

**Abstract:** A common approach to realize decentralized search in unstructured P2P networks is to use random walks combined with a replication strategy. We investigate the problem of minimizing the number of replicas needed to achieve a given probability of finding an existing data item. In contrast to other approaches, we thereby consider the probability of each peer and communication link to be available. In this setting, the main objective of our research is to propose a replication algorithm together with a worst-case approximation guarantee.

## 1 Introduction

Peer-to-peer (P2P) systems have gained much research interest in the last years. In general, it can be distinguished between structured and unstructured P2P networks. In structured P2P networks, data placement and overlay topology are highly constrained to provide a routing mechanism for queries so that data items can be efficiently looked up. In unstructured P2P networks, peers can freely choose their neighbors and data items can be stored at any peer. Because of this flexibility, unstructured P2P networks can handle a higher amount of churn (peers going online and offline) than structured P2P networks.

Since unstructured P2P networks lack an inherent routing mechanism, decentralized searches are a challenge for this type of network. While flooding (each peer forwards a request to all of its neighbors) is guaranteed to find the requested item, this strategy requires all peers to be visited in the worst-case, which does not scale for larger networks. Therefore, alternative approaches, e.g. [LCC+02, MBS05], are suggested that are based on random walks. In a random walk, a peer forwards a request to a random neighbor. Random walks provide better scalability but data items need to be replicated in order to achieve a sufficient probabilistic guarantee of finding data items.

In our approach, the user specifies a certain availability guarantee for each data item. An availability guarantee means that the probability a data item is accessible by representative queries is always above a given threshold. In P2P networks, the time peers are available (online) can vary considerably. The availability of a peer (probability that a peer is online) makes a difference for replica placement because if a replica of a data item is placed at a peer with low availability, it is likely that the peer will go offline shortly. Other replicas might then need to be placed to maintain the availability guarantee for the data item, which causes the following problems. On the one hand, this requires additional resources for further computations of replica placements and for actually performing the replica placements. On the other hand, when a replica holder goes offline the availability guarantee for the data item can be violated for the time until new replica placements are computed and performed.

The availability of peers should therefore inform the computation of replica placements. In contrast to other approaches like [LCC+02, MBS05], we consider the availability of peers when computing a replica placement that achieves the desired availability guarantee. Furthermore, because communication links between peers are subject to failure, we also consider link availabilities (probability that they are functioning). Since the problem of optimal replica placement is at least as hard as NP-complete problems [JR94], we investigate approximation algorithms and their worst-case approximation guarantees.

In the following two sections, main issues of our approach are discussed in more detail. In Section 2, the model we use for peer and link availability is presented. In Section 3, approximation algorithms for replica placement are discussed. Finally, further research is outlined in Section 4.

## 2 Peer and Link Availability Models

In the following, peers and communication links are referred to as components. A common way to model component availabilities [STP96] is briefly described as follows. The *transient* availability  $A_C(t)$  is the probability that the component  $C$  functions at time  $t$ . Let  $MTTF_C$  be the mean time to failure (mean time until component is unavailable) and let  $MTTR_C$  be the mean time to repair (mean time until component is available again) of a component  $C$ . The *steady-state* availability of a component  $C$  is then [STP96]:

$$A_C = \lim_{t \rightarrow \infty} A_C(t) = \frac{MTTF_C}{MTTF_C + MTTR_C}$$

A typical availability function is shown in Figure 1 [STP96, taken from Rie98]. Note that the steady-state availability  $A_C = 0$  if a component  $C$  never goes online again once it is offline and that the steady-state availability is a tight lower bound of the transient availability. Since we are interested in long-term availability guarantees and because the steady-state availability can be easily computed if  $MTTF_C$  and  $MTTR_C$  are given, our investigation focuses on steady-state availabilities.

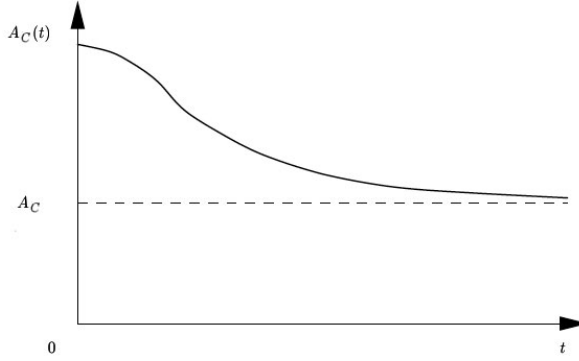


Figure 1: A typical availability function.

### 3 Approximation Algorithms for Optimal Replica Placements

In the case of general network topologies, already the problem of the optimal placement of only one replica is proven to be #P-complete [JR94], i.e. to be at least as hard as NP-complete. It is generally believed that there is no polynomial-time algorithm for any NP-complete problem. We therefore study approximation algorithms with polynomial runtime, which can (and often do) produce suboptimal solutions. Riedewald [Rie98] suggests a heuristic in the context of distributed database systems, which aims to minimize the number of replicas needed to achieve a given availability guarantee for each data item. The algorithm, which we use as a starting point for our investigation, is summarized below.

The algorithm computes the replica placement separately for each data item. Let  $A_d(p)$  be the probability that data item  $d$  is accessible from peer  $p$  in the network and let  $A_d$  be the probability that data item  $d$  is accessible given the request probability of each peer for this data item. It is assumed that the network graph is acyclic because with this assumption  $A_d(p)$  can be computed for each peer with a simple deep-first search algorithm.  $A_d$  is then computed as the weighted sum of the  $A_d(p)$  values of all peers with weights according to the request probability distribution for data item  $d$ . When the algorithm starts, it is assumed that no replica has been placed yet. In each step one more replica is placed by first computing  $A_d$  for each candidate placement of the new replica. Then the candidate placement with the highest increase on  $A_d$  is chosen. Finally, the algorithm stops as soon as the desired availability guarantee is achieved.

Riedewald [Rie98] shows with experiments that the proposed heuristic works well for various network topologies but no worst-case approximation guarantee could be given. For example, it might be possible to show that in the worst case the heuristic needs to place  $k$  times more replicas than the number of replicas in the optimal solution. We are investigating such a worst-case approximation guarantee since it is an important indicator for the quality of the approximation.

The algorithm has an asymptotic runtime of  $O(n^4)$  ( $n$  is the number of peers), which might be critical in practice. We therefore plan to develop an approximation algorithm with better asymptotic runtime. Moreover, Riedewald's heuristic computes the replica placements for each data item in a centralized way. Our goal is to compute this in a distributed way, so that peers do not need to have global knowledge over all replica placements in the network. Finally, our algorithm should not be restricted to acyclic graphs.

## 4 Further Research

In Section 3, possible extensions for Riedewald's heuristic [Rie98] are described. Apart from that, we explore how search techniques based on random walks can be extended to achieve availability guarantees for data items in unstructured P2P networks. We further plan to study how our approach can be extended by taking into account that peers only have a limited storage capacity and by optimizing for other criteria such as minimizing the network traffic for placing the replicas.

## References

- [JR94] Donald B. Johnson and Larry Raab. Complexity of Network Reliability and Optimal Resource Placement Problems. *SIAM Journal on Computing*, 23:510–519, 1994.
- [LCC+02] Qin Lv, Pei Cao, Edith Cohen, Kai Li, and Scott Shenker. Search and replication in unstructured peer-to-peer networks. In *16th international conference on Supercomputing*, pages 84–95, New York, USA, 2002. ACM.
- [MBS05] Ruggero Morselli, Bobby Bhattacharjee, Aravind Srinivasan, and Michael A. Marsh. Efficient lookup on unstructured topologies. In *24th annual ACM symposium on Principles of distributed computing*, pages 77–86, Las Vegas, NV, USA, 2005. ACM.
- [Rie98] Mirek Riedewald. *Replikationsstrategien für verteilte Systeme mit vorgegebenen Verfügbarkeitszielen*. Diplom thesis, Saarland University, 1998.
- [STP96] Robin A. Sahner, Kishor S. Trivedi, and Antonio Puliafito. *Performance and reliability analysis of computer systems: an example-based approach using the SHARPE software package*. Kluwer Academic Publishers, 1996.

# KoMa: Koordination verketteter Webservice-Aufrufe

**Robin Marterer**

Universität Paderborn  
Institut für Informatik  
robbim@upb.de

**Matthias Moi**

Universität Paderborn  
Institut für Informatik  
moisun@upb.de

**Tobias Schauerte**

Universität Paderborn  
Institut für Informatik  
tschauer@upb.de

**Abstract:** Ziel der Projektgruppe MoKEx III (Mobile Knowledge Experience III) ist die Entwicklung eines Wissensmanagementsystems, das von dem Umgang mit Dokumenten im klassischen Sinn abstrahiert. Dazu wird jedes Dokument, das dem System zugeführt wird, nicht nur einer umfassenden Metadatenextraktion mit mehreren Werkzeugen, sondern auch einer semantischen Analyse unterzogen. Ergebnis dieser Prozesse ist ein Wissensobjekt, das nicht mehr aufgrund eines Dateinamens, sondern über seine Eigenschaften (Metadaten), die im LOM-Format gespeichert werden, wiedergefunden werden soll. Wissensobjekte und deren Metadaten sind suchbar und editierbar. Das zugrundeliegende System besteht aus mehreren Teilkomponenten, die ausschließlich über Webservices miteinander kommunizieren. Dabei ergeben sich spezifische Probleme wie die Koordination von verketteten Webservice-Aufrufen und die Sicherstellung von Transaktionssicherheit. Diese Probleme werden von der zentralen Komponente des Projekts, dem Kommunikations-Manager (KoMa), einer konzeptionellen Neuentwicklung des in MoKEx II entwickelten Knowledge Bus (KNS), gelöst.

## 1 Gesamtsystem

Heutzutage existieren viele Softwarelösungen, die das Wissens- und Informationsmanagement unterstützen. Diese sind allerdings bei der Extraktion und der Ablage von Wissensobjekten weniger fortschrittlich als vorhandene Systeme. Für die Speicherung von Wissensobjekten bietet sich das Konzept des „Single Point of Information“ (SPI) [HMR<sup>+</sup>07] an, um der heutigen Informationsflut gerecht zu werden. Es erleichtert dem Benutzer die Arbeit mit Informationen erheblich, da es für alle Arten von Wissensobjekten nur noch eine Anlaufstelle gibt, wo Informationen erzeugt, abgelegt und wiedergefunden werden. Hinter dem SPI steht ein Framework mit strikt service-

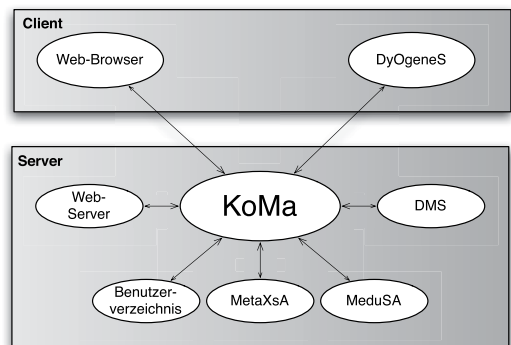


Abbildung 1: Kommunikationsarchitektur

orientierter Architektur (SOA), welches für die Vergaben von Metadaten, die persistente Ablage der Wissensobjekte und die Auffindbarkeit der Informationen über die Metadaten sorgt. Im Rahmen der Projektgruppe MoKEx III wurde ein hoch modulares System [HHM<sup>+</sup>06] entwickelt, welches generell mit Hilfe von Metadaten ein effizientes Ablegen und Auffinden von Wissensobjekten ermöglicht. Bei Wissensobjekten beschränken wir uns in MoKEx III auf Dokumente.

Das Gesamtsystem besteht aus den Komponenten DyOgeneS [BS08], MetaXsA [SMR], MeduSA [DG08], Benutzerverzeichnis, DMS, Web-Server und KoMa. Die entwickelte Architektur des Gesamtsystems entspricht, wie in Abbildung 1 zu sehen, einer Client-Server-Architektur. Jede Server-Komponente kann zur besseren Lastverteilung auf einer separaten Hardwareumgebung laufen. Jegliche Kommunikation zwischen den Komponenten im System erfolgt über den Kommunikations-Manager KoMa. Alle serverseitigen Komponenten werden in einer Initialisierungsphase beim KoMa registriert. Das Benutzerverzeichnis ist für die Verwaltung von Benutzern zuständig. Neben benutzerspezifischen Informationen werden Nachrichten vermerkt, die den Benutzer aufgrund eines vorzeitigen Abmeldens nicht mehr erreicht haben und beim erneuten Anmelden am System nachgeholt werden müssen. Intern nutzt das Benutzerverzeichnis einen LDAP v3 Server. Die Benutzer melden sich über den Client DyOgeneS am System an, welcher sich per Web-Browser als Java-Webstart-Applikation beziehen lässt.

## 2 Webservice-Nutzung

Die Kommunikation der einzelnen Systeme untereinander erfolgt ausschließlich über Webservices. Zur Implementierung wurde die im Java SDK 6 enthaltene Java API for XML Web Services<sup>1</sup> (JAX-WS) genutzt, eine High-Level-API, die dem Programmierer das Versenden und Empfangen der SOAP-Nachrichten nahezu so einfach macht, wie den Aufruf einer entfernten Prozedur. Um Webservices anbieten zu können, wird der in Java SDK 6 integrierte Webserver genutzt, so dass auch auf dem Java-Webstart-Client DyOgeneS Webservices unmittelbar angeboten werden können. Da aus einem Webservice-Aufruf des Benutzer-Clients DyOgeneS auf dem KoMa in den meisten Fällen eine Kommunikationsnotwendigkeit des KoMa mit vielen weiteren Systemen entsteht, sind die über die Webservices angebotenen Funktionalitäten so grobgranular wie möglich. Dies sichert uns, dass die durch die Nachrichtenübertragungen hervorgerufenen Verzögerungen möglichst gering bleiben. Damit die Nachrichtenübertragung durch Webservices die Transaktionssicherheit des Systems nicht gefährdet, wird jede asynchrone Kommunikation durch synchrone Kommunikation simuliert. Somit liefert jeder Aufruf eine Erfolgsantwort oder einen handhabbaren Fehler.

---

<sup>1</sup><https://jax-ws.dev.java.net/>

### 3 Beispiel einer Webservice-Kette

Nun betrachten wir den Webservice *koma\_uploadDoc* mit den Aufrufparametern *con*, *docId*, *uploadDocument*, *storageSystem* und dem optionalen Rückgabeparameter *LomFile*. Dieser Webservice dient dem Hochladen eines Dokuments in das System. Bei *con* handelt es sich um eine Session-ID, die ein Client beim Aufruf des Login-Webservices erhält. Nur mit einer gültigen Session-ID darf ein Client Webservices auf dem MetaXsA ausführen. *docId = 0* bedeutet, dass ein Dokument neu - und *docId > 0*, dass ein Dokument als neue Version des durch die *docId* spezifizierten Dokuments in das System hochgeladen werden soll. Dies ist nur möglich, wenn das Dokument mit dieser *docId* zuvor für den handelnden Benutzer reserviert und somit für alle anderen Benutzer gesperrt wurde. *storageSystem* spezifiziert das Teilsystem, in dem das *uploadDocument* gespeichert werden soll. Im Rahmen von MoKEx III ist dieses immer das DMS. Der Webservice *koma\_uploadDoc* wurde an dieser Stelle gewählt, da er eine der längsten Webservice-Ketten im Kontext von MoKEx III implementiert und dabei alle angeschlossenen Teilsysteme verwendet werden. Wir betrachten nachfolgend den Aufruf des Webservices mit einer *docId > 0*.

*koma\_uploadDoc* wird über den Aufruf eines Clients, z.B. DyOgeneS, angestoßen. Nach der Feststellung der Gültigkeit der Session-ID wird das Dokument per Webservice an das DMS übergeben. Dieses speichert das Dokument und vergibt eine neue *docId\_neu*, welche als synchrone Webservice-Antwort an den MetaXsA zurückgegeben wird. Hier wird durch eine persistente Speicherung der *docId\_neu* zusammen mit den Aufrufparametern von *koma\_uploadDoc* im Transaktions-Manager sichergestellt, dass die bis dahin verfügbaren Informationen nicht verloren gehen können, z.B. um einen wegen eines Systemausfalls nicht beendeten Hochladevorgang erneut anstoßen zu können. Der Client bekommt eine synchrone Webservice-Antwort, die eine Bestätigung enthält, dass das Dokument serverseitig erfolgreich gespeichert wurde. Unmittelbar vor dieser Antwort wird serverseitig ein eigenständiger Thread gestartet, der die Metadatenextraktion und semantische Analyse durch MetaXsA anstößt. Das Ergebnis dieses Prozesses sind eine LOM-Datei, welche die Metadaten enthält und eine Textdatei, die den unformatierten Inhalt des hochzuladenden Dokuments enthält. Diese Textdatei dient der Erstellung eines Lucene-Indexes, dessen Verwendung in [DG08] beschrieben wird.

Global gesehen handelt es sich bei dem Webservice *koma\_uploadDoc* um einen asynchronen Aufruf, dessen Antwort über den synchronen Webservice-Aufruf *koma\_finishedExtractMetadata* durch MetaXsA auf dem MetaXsA realisiert wird. MetaXsA übergibt diesem Webservice die soeben erstellte LOM-Datei sowie die Lucene-Textdatei. Anschließend fordert der MetaXsA die LOM-Datei des Vorgängerdokuments (*docId*) per Webservice von MetaXsA an. Diese, sowie die neue LOM-Datei, werden vom MetaXsA mit bestimmten Metadaten, z.B. Versionierung, Vorgänger- und Nachfolger-Beziehung, angereichert und gemeinsam per Webservice an MetaXsA zurückgegeben. War die Speicherung in MetaXsA erfolgreich, so wird die *docId* aus dem Transaktions-Manager gelöscht. Abhängig davon, ob der Benutzer dies wünscht, muss die LOM-Datei noch an den Client weitergegeben werden, um eine Bearbeitung der extrahierten Metadaten zu ermöglichen. Dies erfolgt durch einen Webservice-Aufruf auf dem Client, dem die LOM-Datei übergeben wird. Sowohl der KoMa als auch der Client bestätigen den Aufruf



des jeweiligen Webservices durch eine positive synchrone Webservice-Antwort. Ist der Client nicht erreichbar, wird die *docId\_neu* für den handelnden Benutzer im Benutzerverzeichnis gespeichert und die zugehörige LOM-Datei beim nächsten Einloggen an diesen übertragen.

## 4 Ausblick

Das Gesamtsystem besteht aus verschiedenen Komponenten, deren Funktion wohldefiniert ist. Diese Komponenten kommunizieren ausschließlich über Webservices. Somit ist es problemlos möglich, beliebige Komponenten durch andere zu ersetzen, sofern sie die konzeptionell gleichen Funktionalitäten über die gleichen Webservices bereitstellen. Weiterhin ist im System konzeptionell berücksichtigt worden, dass in Zukunft neben Dokumenten weitere Wissensobjekte, z.B. E-Mails, aus weiteren Quellen, z.B. einem Mailserver, verarbeitet werden können. Durch die klar definierten Aufgaben der Komponenten, die exakten Schnittstellen in Form von Webservices und die Vorbereitung, weitere Wissensobjekte aus weiteren Datenquellen verarbeiten zu können, bietet das System eine ideale Basis für ein umfassendes Verwaltungssystem von Wissensobjekten.

## Literatur

- [BS08] Thomas Bremes und Lars Scheuermann. DyOgeneS: Generierte Oberflächen in einem verteilten automatisierten System. In *eingereicht bei den Informatiktagen 2008*, 2008.
- [DG08] Steffi Döbbelin und Senol Gül. MeduSA: Umgang mit Metadaten in dynamischen Softwareumgebungen. In *eingereicht bei den Informatiktagen 2008*, 2008.
- [DKC05] Jean Dollimore, Tim Kindberg und George Coulouris. *Verteilte Systeme*. Pearson Studium, 3. Auflage, 2005.
- [HHM<sup>+</sup>06] Knut Hinkelmann, Kai Holzweißig, Johannes Magenheim, Fabian Probst und Wolfgang Reinhardt. Linking Communities of Practice with Learning Communities in Computer Science Education. In *Education for the 21st Century — Impact of ICT and Digital Resources*, IFIP Computer Science, Seiten 83–92. Springer Berlin / Heidelberg, 2006.
- [HMR<sup>+</sup>07] K Hinkelmann, J. Magenheim, W. Reinhard, T. Nelkner, K. Holzweißig und M. Mlynarski. KnowledgeBus - An Architecture to Support Intelligent and Flexible Knowledge Management. In *Duval, E., Klamma, R., Wolpers, M. (Hrsg.) Creating New Learning Experiences on a Global Scale. Second European Conference on Technology Enhanced Learning, EC-TEL 2007 Crete, Greece*, Seiten 487–492. Springer Berlin, 2007.
- [SMR] Benedikt Schmidt, Christian Mletzko und Wolfgang Reinhardt. MetaXsA: Modulares Framework zur Metadatenextraktion durch semantische Dokumentanalyse auf unstrukturierten Textdokumenten. In *eingereicht bei den Informatiktagen 2008*.

# Automatische Extraktion von Metadaten aus verschiedenen Dateiformaten

**Christian Mletzko**

Universität Paderborn  
Institut für Informatik  
letris@uni-paderborn.de

**Benedikt Schmidt**

Universität Paderborn  
Institut für Informatik  
schmidt@uni-paderborn.de

**Abstract:** Metadaten dienen der Beschreibung von Dateien und als Grundlage für Suchanfragen in informationshaltenden Systemen. Aufbauend auf den Erkenntnissen einer Bachelorarbeit entstand im Rahmen einer Projektgruppe eine Softwarekomponente zur automatischen Extraktion von Metadaten aus verschiedenen Dateiformaten. Unter Zuhilfenahme mehrerer externer Metadatenextraktionswerkzeuge werden die zu einem Dokument zugehörigen Metainformationen bestimmt, verglichen und bewertet. Zusätzlich zu der Erstellung eines Gesamtergebnisses werden die erhaltenen Informationen mit einem Qualitätsattribut versehen und entsprechend der Ergebnissicherheit bewertet.

## 1 Problemstellung

Der Erfolg eines Unternehmens hängt signifikant von der internen Informationsbereitstellung und Möglichkeiten ihrer Verbreitung ab. Gerade unter Betrachtung der zunehmenden Anzahl von verwendeten Softwaresystemen und der rasch anwachsenden Informationsflut wird es immer schwieriger, dem Informationssuchenden angeforderte Wissensressourcen leicht zugänglich zu machen. Gleiches gilt für die Ablage oder Speicherung von Daten. Ein zentraler Zugangspunkt zu den informationshaltenden Systemen aller Bereiche einer Organisationseinheit ermöglicht einen einheitlichen Zugriff auf die verfügbaren Informationen. Dieses Konzept des “Single Point of Information” entstand aus der international angelegten Projektserie “MoKEx” (Mobile Knowledge Experience) [HHM<sup>+</sup>06]. Eine Verbesserung der Auffindbarkeit von Dokumenten innerhalb des Systems wird durch die Verwendung von qualitativ hochwertigen Metainformationen unterstützt. Zur Steigerung der Interoperabilität zwischen verschiedenen Systemen bietet sich eine strukturierte und standardisierte Speicherung der Metadaten, beispielsweise im LOM-Standard [IEEE02] an. Im Rahmen der Projektgruppe “MoKEx 3” entsteht eine Softwarekomponente, die den Benutzer bei der Vergabe von Metainformationen zu Dokumenten unterstützen soll. Hierzu müssen automatisiert Metadaten aus den in das System eingestellten Dokumenten extrahiert werden. Zur Analyse dieser Dokumente werden verschiedene externe Metadatenextraktionswerkzeuge benutzt, deren Einzelergebnisse bestimmt, verglichen und bewertet werden. In einem weiterführenden Schritt wird aus den Teilergebnissen ein Gesamtergebnis erstellt und dieses mit Qualitätsattributen versehen. Nach dem Ablauf der

Metadatenextraktion erfolgt, sofern es sich bei der zu analysierenden Datei um ein textuelles Dokument handelt, eine semantische Analyse [SMR08] des Inhalts. Die hieraus erschlossenen Metainformationen werden mit den Ergebnissen der Metadatenextraktion in ein XML-Dokument im LOM-Standard überführt.

## **2 MetaXsA - Metadatenextraktion**

Als Teilmodul der Softwarekomponente MetaXsA greift die Metadatenextraktion auf externe Programme zu. Es folgen Hinweise zur Auswahl geeigneter Werkzeuge und zur Umsetzung des Vorhabens.

### **2.1 Auswahlkriterien und Werkzeuge**

Die von dem MetaXsA-Modul verwendeten Extraktionswerkzeuge müssen einer Reihe von Anforderungen genügen, um sinnvoll eingesetzt werden zu können. Hierzu gehören:

1. Lizenzbestimmungen - Die verwendeten Extraktoren müssen unter der GPL oder der LGPL stehen, sodass die freie Verwendbarkeit und gegebenenfalls eine Modifikation gewährleistet werden kann.
2. Anschaffungskosten - Um die Kosten für etwaige anzuschaffende Software möglichst gering zu halten, sollten die Extraktionswerkzeuge möglichst kostenfrei sein.
3. Quantität der extrahierten Daten - Es sollen möglichst viele Informationen extrahiert werden können, um auf einer großen Datenbasis aufbauend, die durch hohe Relevanz ausgezeichneten Informationen extrahieren zu können.
4. Spektrum hinsichtlich lesbarer Dateiformate - Optimalerweise ist jedes Extraktionswerkzeug in der Lage Metainformationen aus verschiedenen Dateitypen zu extrahieren. Eine Kombination von mehreren Werkzeugen führt zwar zu dem gleichen Ziel, ist aber weniger performant.
5. Qualität der Informationen - Neben der Korrektheit der extrahierten Daten ist die Wiederverwendbarkeit von großer Bedeutung. Eine ungeeignete Aufbereitung und Darstellung der Informationen könnte die weitere Verwendung der Ergebnisse erschweren.
6. Integrierbarkeit - Die Werkzeuge zur Extraktion von Metadaten müssen in ein eigens entwickeltes Modul eingebettet oder über einen Kommandozeilenaufruf aus einem Programmteil heraus aufgerufen werden können.

Unter Berücksichtigung der an sie gestellten Anforderungen bilden die drei Werkzeuge – Libextractor [GS07], Hachoir-Metadata [SPT<sup>+</sup>08] und Exiftool [Har08] – die Menge, die sich aus den verfügbaren, frei erhältlichen Werkzeugen für die Verwendung im Rahmen des MetaXsA-Moduls eignet. Bei der Auswahl von Extraktionswerkzeugen, basierend auf Untersuchungen hinsichtlich der Extraktionsergebnisse, wurden Dokumente im Office- und PDF-Format höher priorisiert betrachtet, da eine stichprobenartige Untersuchung zeig-

te, dass diese Dateiformate im praktischen Einsatz die am häufigsten auftretenden Form digitalen Informationsaustauschs in der Bürokommunikation bilden<sup>1</sup>.

## 2.2 Umsetzung und Problembehandlung

Das Metadatenextraktionsmodul erhält eine textuelle oder binäre Datei, deren Metainformationen mit Hilfe der Metadatenextraktoren bestimmt und über den Standardausgabestrom des Betriebssystems zurückgeliefert werden. Realisiert wird dieser Ablauf durch ein Teilmodul, welches die zu analysierende Datei nacheinander an die angeschlossenen Extraktoren übergibt. Die ermittelten Ergebnisse werden eingelesen und durch ein Parsing-Modul zur Wiederverwendung aufbereitet. Es erfolgt die Übermittlung der Daten an ein drittes Teilmodul, welches die Einzelergebnisse interpretiert, aus diesen ein Gesamtergebnis erstellt und die Ergebnissicherheiten bestimmt. Ein viertes Teilmodul ermöglicht den Zugriff auf eine Konfigurationsdatei, in der die Aufrufparameter der Extraktoren und die Abbildungen der Elemente des LOM-Standards auf die Bezeichner der extrahierten Werte hinterlegt sind. Das Gesamtergebnis der Metadatenextraktion wird in ein XML-Dokument im LOM-Standard überführt und mittels Web-Service-Technologie an eine für die Datenerhaltung zuständige Komponente übermittelt. Bei einer simultanen Verwendung mehrerer Extraktionswerkzeuge zur Erstellung eines Gesamtergebnisses kann, wie eine Evaluation zeigte, nicht davon ausgegangen werden, dass alle Werkzeuge ein identisches Ergebnis erzielen. Zur Bestimmung eines Gesamtergebnisses sind Fallunterscheidungen notwendig, mit Hilfe derer das Ergebnis an sich und die entsprechende Ergebnissicherheit bestimmt werden können. Es gilt folgende Fälle zu überprüfen:

- Identische Ergebnisse - Die Ergebnisse von Hachoir-Metadate, Libextractor und Exiftool sind für das betrachtete Feld identisch. In diesem unproblematischen Fall kann ein beliebiger der drei extrahierten Werte verwendet werden.
- Zwei gleiche Ergebnisse - Es müssen die drei möglichen Kombinationen aus genau zwei gleichen Ergebnissen auf Sinnhaftigkeit oder leere Ergebnisse überprüft werden. Im zweiten Fall wäre das von den zwei identischen Ergebnissen abweichende Ergebnis das einzig sinnvolle und sollte zur Erstellung des Gesamtergebnisses verwendet werden. Andernfalls stellt das mehrfache Auftreten eines Ergebnisses im Regelfall eine Bestätigung dar und sollte als Ergebnis übernommen werden.
- Paarweise verschiedene Ergebnisse - Die Ergebnisse der drei Extraktionswerkzeuge sind für das betrachtete Feld paarweise verschieden. Basierend auf Evaluationsergebnissen sollte in diesem Fall das Ergebnis von Exiftool verwendet werden. Für den Fall, dass Exiftool das einzige Werkzeug war, welches außer Stande gewesen ist, ein Ergebnis zu bestimmen, sollte das Ergebnis von Hachoir-Metadate verwendet werden.

Aufgrund der Fallunterscheidungen kann während der Ermittlung des Gesamtergebnisses eine Aussage über die Ergebnissicherheit getroffen werden. Dazu wird von festen Basiswerten für jeden Extraktor ausgehend in Abhängigkeit von dem vorliegenden Fall und

---

<sup>1</sup>Für nähere Informationen sei an dieser Stelle auf die Studienarbeit [Mle07] verwiesen.

dem verwendeten Extraktionsergebnis mit Hilfe einer Art Bonussystem ein Wert zwischen 0 und 100 vergeben.

### 3 Zusammenfassung und Ausblick

Die simultane Verwendung von drei Metadatenextraktionswerkzeugen in der Softwarekomponente MetaXsA sichert eine hohe Ergebnisqualität und -vielfalt, die durch den ergänzenden Einsatz einer Komponente zur semantischen Analyse von textuellen Dokumenten bereichert wird. Die Kombination der beiden Teilkomponenten liefert umfangreiche Metainformationen zu Dokumenten, die im Einstellungsprozess in ein informationshaltendes System erschlossen werden und im Anschluss daran eine fundierte Basis für Suchanfragen an das System bilden. Ein modularer Aufbau ermöglicht den Austausch oder die Ergänzung von Extraktionswerkzeugen für einen Einsatz in heterogenen Umgebungen.

### Literatur

- [GS07] Christian Grothoff und Vids Samanta. Libextractor Projektseite. <http://gnunet.org/libextractor/index.php?xlang=German>, zuletzt eingesehen am 16. November 2007, 2007.
- [Har08] Phil Harvey. ExifTool Projektseite. <http://www.sno.phy.queensu.ca/~phil/exiftool/>, zuletzt eingesehen am 10. Januar 2008, January 2008.
- [HHM<sup>+</sup>06] Knut Hinkelmann, Kai Holzweißig, Johannes Magenheimer, Fabian Probst und Wolfgang Reinhardt. Linking Communities of Practice with Learning Communities in Computer Science Education. In *Education for the 21st Century — Impact of ICT and Digital Resources*, IFIP Computer Science, Seiten 83–92. Springer Berlin / Heidelberg, 2006.
- [Mle07] Christian Mletzko. Evaluierung und Leistungsvergleich automatischer Extraktionstools zur Gewinnung von Metadaten aus verschiedenen Dateiformaten. Studienarbeit, Universität Paderborn, 2007.
- [IEEE02] Learning Technology Standards Committee Learning Standards Committee of the IEEE. Draft Standard for Learning Object Metadata. [http://ltsc.ieee.org/wg12/files/LOM\\_1484\\_12\\_1\\_v1\\_Final\\_Draft.pdf](http://ltsc.ieee.org/wg12/files/LOM_1484_12_1_v1_Final_Draft.pdf), zuletzt eingesehen am 11. Januar 2008, 2002.
- [SMR08] Benedikt Schmidt, Christian Mletzko und Wolfgang Reinhardt. MetaXsA - Modulares Framework zur semantischen Dokumentanalyse für unstrukturierte Textdokumente. In *eingereicht bei den Informatiktagen 2008*, 2008.
- [SPT<sup>+</sup>08] Victor Stinner, Thomas Pabst, Pierre Thierry, Jean-Marc Libs und Elie Roudninski. Hachoir-Metadata Projektseite. <http://hachoir.org/>, zuletzt eingesehen am 10. Januar 2008, January 2008.

# Inhalts- und Personennetze im E-Mail-Verkehr

Tobias Nelkner

Universität Paderborn  
Institut für Informatik  
tobin@uni-paderborn.de

**Abstract:** E-Mails bieten durch ihre Sender und Empfänger eine Verknüpfung zwischen Inhalten und Personen. Es wird ein System vorgestellt, welches im Rahmen einer Diplomarbeit entstanden ist und eine semantische Analyse von E-Mails ermöglicht, um darauf aufbauend zum einen den Ähnlichkeitszusammenhang zwischen verschiedenen E-Mails zu ermitteln und zum anderen das dazugehörige Personennetzwerk zu extrahieren. Die Visualisierung wird durch ungerichtete Graphen realisiert und ermöglicht ein Erfassen des Zusammenhangs zwischen E-Mail-Inhalten und der korrespondierenden Personen. Darüber hinaus wird erkennbar, wieviel Kontakt verschiedene Personen bezüglich eines bestimmten Themas hatten.

## 1 Problemstellung

Die Suche nach Wissens- und Informationsträgern ist ein zentrales Problem bei der Generierung von Inhalten im Arbeitsumfeld. So ist es oftmals von Nöten Experten zu Rate zu ziehen, welche ein spezielles Thema über den eigenen Kenntnisstand hinaus tiefergehend erläutern und weiteres Wissen vermitteln können. Problematisch dabei ist allerdings das Auffinden entsprechender Experten, die zur Lösung beitragen können. Betrachtet man dabei ein betriebliches Umfeld, so sind oftmals Mitarbeiter solche Experten, wobei das Bewusstsein darüber entweder aus Erfahrungswerten gewonnen wurde, die im Laufe der Verweildauer in der Firma entstanden sind oder weil ein Hinweis aus dem Kollegium kam, wer eventuell mehr zu einem bestimmten Thema wissen könnte. In beiden Fällen kann nicht grundsätzlich davon ausgegangen werden, dass die eigene Fragestellung zufriedenstellend beantwortet werden könnte. Die letzte und zeitintensivste Alternative ist dann nur, sofern überhaupt möglich, die autodidaktische Aneignung eines Themas aus anderen Schriftquellen, was aber vermutlich durch die Komplexität eines Themas und den Schwerpunkt der eigenen Ausbildung und des eigenen Wissensstandes limitiert ist.

E-Mails sind im heutigen Geschäftsumfeld nicht mehr wegzudenken; ihre asynchrone Übermittlung von Sender zu Empfänger und ihre mögliche Persistenz machen sie zu einem der flexibelsten Wissensträger und sie stellen oftmals das Hauptkommunikationsmedium innerhalb von Geschäftsprozessen dar [TV05]. Prinzipbedingt können E-Mails nicht nur Inhalte entnommen werden, sondern auch eine Verbindung der Inhalte zu Personen geschaffen werden. Mit anderen Worten: Es kann die Aussage getroffen werden, dass zwei oder mehr Menschen einen gewissen Kenntnisstand über ein Thema teilen.

Im Rahmen der Diplomarbeit wurden zwei Teilaufgaben bearbeitet: Zum einen sollte nach einer (noch genauer zu erläuternden) Analyse von E-Mails bestimmt werden, welche der untersuchten E-Mails inhaltlich entsprechend einer gewissen Metrik, Ähnlichkeiten zueinander haben. Die zweite Aufgabe bestand in der Extraktion von Personen aus einer Menge selektierter E-Mails. Basierend auf einer Gewichtung der Header-Informationen wurde das soziale Netz dieser Personen bestimmt.

## 2 Inhalts- und Personennetzwerke

Im Folgenden wird erläutert, auf welcher Basis die Inhaltsnetze und die Personennetze aus diesen E-Mails extrahiert wurden und wie deren Visualisierung realisiert wurde.

### 2.1 Inhaltsnetze

Als Testdatensatz wurden E-Mails aus meinem persönlichen Bereich herangezogen, welche sich inhaltlich hauptsächlich im Rahmen meiner persönlichen Interessen und meiner Arbeit an der Universität bewegen und sich insgesamt auf eine Menge von über 14000 E-Mails belaufen. Als Grundlage für die Vergleichbarkeit von E-Mails wurde der Inhalt dieser betrachtet. Mit Hilfe einer modifizierten Version von *MetaXsA* (Metadatenextraktion und semantische Analyse), entstanden in der dritten Generation der *MoKE* Projektserie [HHM<sup>+</sup>06], werden sämtliche Nomen und sachbeschreibenden Wörter auf ihre linguistisch korrekte Stammform zurückgeführt und abgespeichert. Darüber hinaus wird eine Klassenanalyse betrieben, welche versucht, eine E-Mail einer gewissen Klasse zuzuordnen, was eine gewisse Gruppierung über gewisse E-Mails zulässt. Die Inhalte dieser E-Mails, als auch die Informationen über ihre Bindungsstärke zueinander, werden der semantischen Analyse anschließend in eine Datenbank gespeichert. Das wesentliche Ergebnis ist pro E-Mail eine Liste von Keywords und die Angabe der Bindungsstärke zu allen anderen E-Mails, sofern diese über einem gewissen Schwellwert liegt. Die Angabe der Bindungsstärke basiert dabei im Wesentlichen auf der Anzahl der gleichen Keywords zweier E-Mails. Obwohl es mit dem Vektorraum Modell [WZW85, Rag97] äußerst leistungsfähige und flexiblere Methoden im Information Retrieval gibt, wurde hier der boolesche Wortvergleich genutzt, da insbesondere die Ähnlichkeitsgewichtung sehr effizient durchgeführt werden kann.

### 2.2 Personennetze

Personennetze werden zur Laufzeit auf Basis gefundener E-Mails ermittelt. Grundlage dafür ist ein Maß für Bindungsstärke zwischen zwei Personen. Dies wurde ermittelt, indem unterschiedlich bewertet wurde, ob ein Empfänger einer E-Mail im *TO*-, *CC*- oder *BCC*-Header der E-Mail steht. Die Erkenntnis, welcher Headereintrag wie zu bewerten ist,

wurde aus einer Umfrage mit etwa 300 Personen aus dem universitären als auch betrieblichen Umfeld abgeleitet. Sollten mehrere E-Mails gefunden werden, welche gleiche Sender und Empfänger besitzen, so wird deren konkrete Bindungsstärke akkumuliert. Durch den Zusammenhang zwischen E-Mails und der damit in Verbindung stehenden Personen, als auch der Gewichtung zwischen diesen E-Mails und zwischen diesen Personen, wurde versucht abzuleiten, in welchem Umfang eine Person mit einer Thematik vertraut ist.

## 2.3 Visualisierung

Diese Inhalts- und Personennetze wurden als ungerichtete Graphen mit Hilfe der *Prefuse* API [HCL05] realisiert, wie in Abbildung 1 zu sehen ist. Für dünne Netze und Netzen mit einer geringen Knotenzahl ist dies ein gangbarer Weg. Dabei ist die Benutzeroberfläche zweigeteilt, wobei im linken Teil das Inhaltsnetz dargestellt wird und im rechten Teil das Personennetz. Zusätzlich wird noch eine Suchmaske angeboten und eine Liste, in welcher eine Rangordnung der gefundenen E-Mails nach dem *tf-idf*<sup>1</sup> Prinzip sortiert eingeblendet wird. Da diese Sortierung nur für die direkten Treffer möglich ist, also all diejenigen, welche das Suchwort selbst enthalten, werden die Treffer, die das Suchwort nicht enthalten aber eine hohe Bindungsstärke zu einer der gefundenen E-Mails haben, der Bindungsstärke nach sortiert. E-Mails als auch Personen werden ihrer Bindungsstärke nach jeweils mit unterschiedlicher Linienstärke und Farbhelligkeit dargestellt; stärkere Bindungen sind breiter und heller. E-Mails die das Suchwort nicht enthalten, werden zunächst nicht dargestellt, allerdings ist es möglich sich diese ebenfalls anzeigen zu lassen. Wird ein E-Mail Knoten mit der Maus berührt, so werden die dazugehörigen Personen entsprechend markiert und herausgehoben, analog geschieht dies mit E-Mails, wenn ein Personen-Knoten berührt wird. Zum Zwecke der Übersichtlichkeit kann auch nur ein einzelnes Teilnetz dargestellt werden, was alle Knoten zeigt, die eine direkte Bindung zum angeklickten Knoten haben. Markiert man in der nach Relevanz sortierten Liste der gefundenen E-Mails einen Eintrag, so wird die entsprechende E-Mail ebenfalls markiert.

---

<sup>1</sup>*tf-idf* steht für *text frequency-inverse document frequency*, wobei die relative Häufigkeit eines Wortes in einem Text in Relation zur Anzahl der Dokumente in einem Korpus gestellt wird, welche dieses Wort enthalten.

Abbildung 1: Visualisierung der E-Mail- und Personenverknüpfungen



### 3 Zusammenfassung und Ausblick

Es wurde ein System vorgestellt, mit dem der Inhalt von E-Mails analysiert werden kann und zum einen die Bindungsstärke zwischen E-Mails, als auch die Bindungsstärke zwischen den Personen, die diese E-Mails geschrieben bzw. empfangen haben, visualisiert werden kann. Damit ist es möglich, Menschen zu identifizieren, welche möglicherweise bezüglich eines bestimmten Themas Expertenwissen vorweisen können.

Dieses System ist in seiner prototypischen Fassung eine Standalone-Software, die konstruiert wurde, um den persönlichen E-Mail Verkehr zu analysieren. Ein deutlicher Mehrwert würde entstehen, wenn auf Basis einer Client-Server Architektur die E-Mails mehrerer Nutzer analysiert würden, zum einen zur Verbesserung der Qualität der Zusammenhangsdarstellungen, zum anderen für den tatsächlichen Nutzen bei der Verwendung der Software durch mehrere Benutzer. Diesem Schritt stehen allerdings rechtliche Bedenken gegenüber, da in jedem Falle das Bundesdatenschutzgesetz beachtet werden muss. Darüber hinaus sollte die Geschwindigkeit der semantischen Analyse verbessert und außerdem die Qualität der Ergebnisse gefördert werden, in dem z.B. deutlich mehr Dokumentklassenidentifizierung implementiert wird.

Im Ergebnis bleibt festzuhalten, dass eine explorative Identifizierung von Experten ermöglicht wird, was in gängigen E-Mail Clients auf Grund der fehlenden Übersichtlichkeit so nicht möglich ist.

### Literatur

- [HCL05] Jeffrey Heer, Stuart K. Card und James A. Landay. *prefuse: a toolkit for interactive information visualization*. In *CHI '05: Proceedings of the SIGCHI conference on Human factors in computing systems*, Seiten 421–430, New York, NY, USA, 2005. ACM.
- [HHM<sup>+</sup>06] Knut Hinkelmann, Kai Holzweißig, Johannes Magenheimer, Fabian Probst und Wolfgang Reinhardt. *Linking Communities of Practice with Learning Communities in Computer Science Education*. In *Education for the 21st Century — Impact of ICT and Digital Resources*, IFIP Computer Science, Seiten 83–92. Springer Berlin / Heidelberg, 2006.
- [Rag97] Prabhakar Raghavan. *Information retrieval algorithms: a survey*. In *SODA '97: Proceedings of the eighth annual ACM-SIAM symposium on Discrete algorithms*, Seiten 11–18, Philadelphia, PA, USA, 1997. Society for Industrial and Applied Mathematics.
- [TV05] Rana Tassabehji und Maria Vakola. *Business email: the killer impact*. *Commun. ACM*, 48(11):64–70, 2005.
- [WZW85] S. K. M. Wong, Wojciech Ziarko und Patrick C. N. Wong. *Generalized vector spaces model in information retrieval*. In *SIGIR '85: Proceedings of the 8th annual international ACM SIGIR conference on Research and development in information retrieval*, Seiten 18–25, New York, NY, USA, 1985. ACM.

# Performance Analysis and Optimization of an XML Database System

Yi Ou

Technische Universität Kaiserslautern  
Fachbereich Informatik  
o\_yi@informatik.uni-kl.de

Art der Arbeit: Diplomarbeit

Betreuer der Arbeit: Prof. Dr.-Ing. Dr. h. c. Theo Härder, Dipl. Inf. Karsten Schmidt

## 1 Introduction

Efficient transactional XML processing can only be supported by native XML database management systems (XDBMSs) that are tailored to the XML data model [HH06]. XTC (XML Transaction Coordinator) is an XDBMS prototype developed by the research group DBIS at the University of Kaiserslautern. So far, research and development in the context of XTC are focused on how to solve individual problems in the scope of native transactional XML processing, e.g., how to store XML data efficiently, how to support fine-grained transactional processing, how to efficiently evaluate queries on XML data, etc. Solutions to these individual problems do not necessarily improve the overall system performance, which is vital for a DBMS. Even if they do, it is difficult to quantify their contribution to the overall performance. With XTC becoming more and more a full-fledged XDBMS, there is a strong need for methods and frameworks that can support the analysis and optimization of its overall performance. Evaluation and development of such methods and frameworks is the *main objective* of this thesis.

## 2 Problem Analysis

The performance of XDBMSs is influenced by many factors simultaneously. These factors fall into four categories: *implementation quality*, *workload*, *database characteristics*, and *hardware resources*. The last three categories are the topics of *database tuning*, which is the typical task of database administrators (DBAs) in a production environment. It is also a difficult task, because too many factors and parameters have to be considered, as well as their inter-dependencies and interactions.

Moreover, a lot of performance-influencing factors *change over time* (workload, database, and even hardware resources). Dealing with these changes manually, e.g., by the DBAs, is not acceptable or even not possible. Research topics, such as *self-tuning* and *self-adaptive* DBMSs ([CN07][WM02][SH07]), try to find solutions to this kind of problems. Most of the former researches are based on the *monitor-diagnose-tune* [CN07] or the *feedback-control-loop* [WM02] paradigms. Based on their experience, we can conclude that a monitoring infrastructure is essential for developing a self-tuning XDBMS.

It is not possible to optimize a DBMS for all application domains at once, because the performance of database systems is largely influenced by the workload and database characteristics, which are domain-specific. However, since the market for XDBMSs is not yet mature, it is not clear what will be the major application domains of XDBMSs. Therefore, our efforts on the performance analysis and optimization of XTC focus on *typical use cases* of an XDBMS, which fall into three categories: document management, declarative accesses, and navigational accesses.

Finding *performance bottlenecks* and removing them, is the most effective way of optimization. That is, the system components that have the largest negative influence on the overall performance should be optimized in the first place. Note that bottleneck removal is an iterative process, because, once a bottleneck is removed, the next bottleneck (potentially in another component) shows up. For this reason, our methods and frameworks facilitating the detection and removal of performance bottlenecks effectively support the performance analysis and optimization of XTC.

### 3 Our Approach

Towards our objective, we employed an approach consisting of three parts: *benchmarking*, *profiling*, and *monitoring*. Their relationship to the XTC server is schematically shown in the Figure 1, where the components developed in the course of this diploma work are represented by shaded boxes. Their fine-grained interplay with XTC provided the monitoring and analysis framework and functionality sketched in the following.

An infrastructure for performance evaluation is a precondition for performance analysis and optimization. Our goal with benchmarking is to build this infrastructure so that we can assess the overall performance of XTC. We evaluated various existing XML benchmarks and benchmark tools in terms of their applicability for transactional XML processing and specified several internal benchmarks for performance evaluation of XTC, based on its current stage of development and typical use cases of XDBMSs. To support the benchmarking and automatic performance testing of XTC, we implemented a test and measuring framework (XtcLap) which is easily extensible and plugin-aware. In form of the plugins, we enriched the framework with various test classes simulating typical XDBMS applications. Our test framework is capable of generating scalable and repeatable workloads. With its help, we stress-tested the XTC server systematically. Abnormal outcomes of these stress-tests are good starting points for further investigation of performance problems.

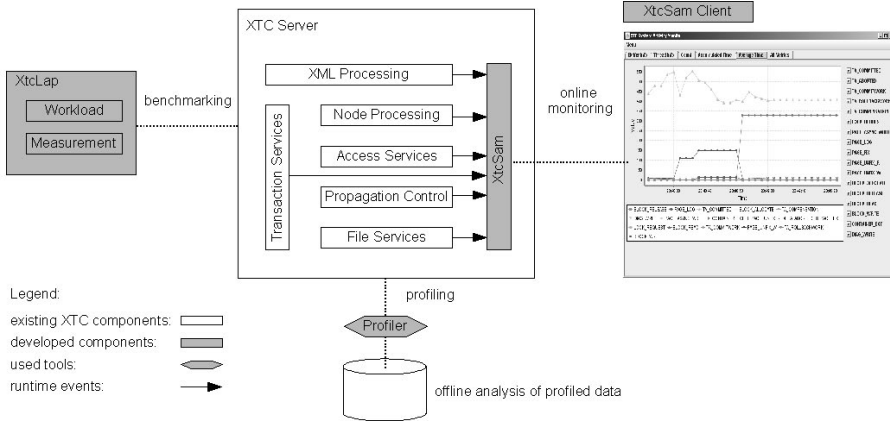


Figure 1 Schematic Diagram of Our Approach

Performance problems may result from flawed implementations. Trying to locate such implementations by static source-code analysis is unacceptable for systems such as XDBMSs. We compared various profiling tools with focus on functionality. We also evaluated the performance overhead of a chosen profiling tool using our internal benchmark. To identify the performance bottlenecks in the system, we collect its runtime data using profiling tools and analyze its behavior under stress.

Finally, we implemented an event-driven “system activity monitor” (XtcSam) for online monitoring of XTC. We analyzed selected system components of XTC and defined a set of performance-critical metrics representing the state of the system and its components at runtime. The future self-tuning components can access these metrics through a programming interface. For human users, we developed a GUI, in which these metrics are visualized with the help of an open-source chart library. Proven by our benchmarks, the system activity monitor has only an ignorable performance overhead. Therefore, it fits well into the monitor-diagnose-tune paradigm of self-tuning DBMSs.

## 4 Conclusions

Our systematic approach and our frameworks are effective for the performance analysis and optimization of XTC. With their help, we already identified some performance problems of XTC and provided hints for their solutions. Using our benchmarks, we are able to evaluate the performance improvements achieved by these solutions. Furthermore, the benchmarks facilitate the project management by enabling the quantitative assessment of the overall system performance. And the test framework assists the development, serving as a basis for a regression test suite. The system activity monitor, besides offering a user-friendly view of the system’s runtime behavior, made the initial step towards a self-tuning XDBMS.

## 5 References

- [CN07] S. Chaudhuri, V. Narasayya: Self-Tuning Database Systems: A Decade of Progress, in: Proc. Int. Conf. on Very Large Data Bases (VLDB), Vienna 2007, pp. 3-14.
- [HH06] Haustein, M. P., Härder, T.: An Efficient Infrastructure for Native Transactional XML, in: Proc. Data & Knowledge Engineering 61:3, pp. 500-523, Elsevier, 2007.
- [SH07] Schmidt, K., Härder, T.: An Adaptive Storage Manager for XML Documents, in: Proc. BTW-Workshop "Maßgeschneidertes Datenmanagement", Aachen, March 2007, pp. 317-328.
- [WM02] Weikum, G., Moenkeberg, A., Hasse, C., Zabback, P.: Self-Tuning Database Technology and Information Services: From Wishful Thinking to Viable Engineering, in: Proc. Int. Conf. on Very Large Data Bases (VLDB), Hongkong 2002, pp. 20-31.

# **Platform independent Data-Synchronization of mobile .NET and J2ME applications using a Service Oriented Architecture**

Marius Podwyszynski, Andreas Schwab

University of Passau, Germany  
Institute for Information Systems and Software Technology

podwyszy@fim.uni-passau.de  
schwab@fim.uni-passau.de

Type of thesis: Bachelor Thesis  
Supervisor: Dipl. Inf. Stefan Kurz

**Abstract:** The University of Passau consists of several information subsystems (LMS, HIS) that avail students to retrieve a variety of information regarding courses, grades or library services. We propose a solution to integrate these subsystems into a transactional Service Oriented Architecture (SOA) by deploying several web services which are transparently accessible by mobile clients running a stand-alone client-framework. To provide data-access in a not-always connected environment, the framework caches the data and synchronizes it with the SOA with millisecond precision. The presented work has been handed in as two Bachelor theses at the University of Passau in August 2007.

## **1 Web Service Framework**

The web service framework communicates via SOAP and supports flexible orchestration of multiple services. Figure 1 shows the service architecture. Within the architecture services play the role of a Base-Service or a Super-Service. While the first usually accesses plain and straight-forward business data, the latter is used to orchestrate and decide about service invocations and to trigger more intelligent actions.

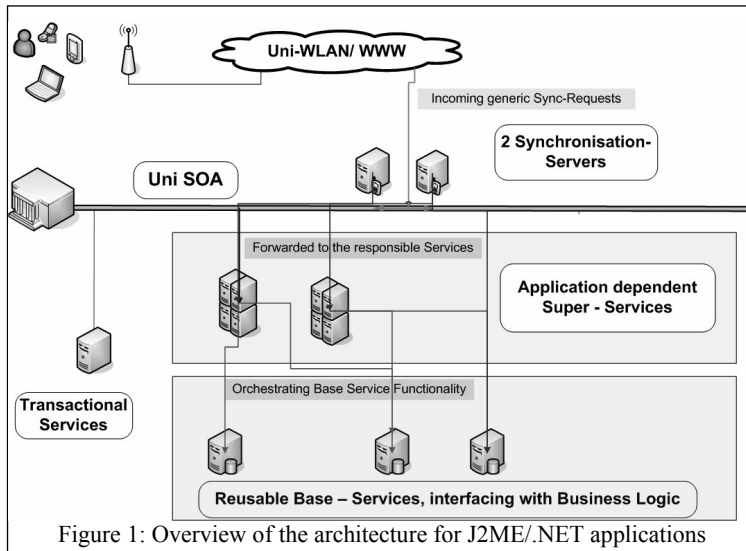


Figure 1: Overview of the architecture for J2ME/.NET applications

The typical usage scenario of Super-Services is faced by data-synchronization. The Super-Service is able to determine the delta data and the occurrence of a synchronization conflict by querying the corresponding Base-Services. Thus, Super-Services implement a synchronizing application logic by controlling access to public data and avoiding data inconsistencies in the various database backends. Moreover, a distributed atomic transaction component has been implemented by a Super-Service, respecting the 2PC protocol. Here, the Super-Service orchestrates the communication with various services which act as participants of a distributed atomic transaction. Due to the variety of use cases, where each use case provides a unique and specific service interface (via its WSDL document), a generic communication interface has been developed. This interface acts as the single communication endpoint for the mobile client. It receives requests from mobile clients and dispatches them to corresponding services within the architecture. All service calls and orchestrations – the simple retrieval of data, the consistency-aware data synchronization and the invocation of transactions – are transparent to the mobile application and handled within a single generic invocation from the mobile client. Each generic invocation is converted into several service specific invocations at the SOA-side and vice versa. Having performed sophisticated service orchestrations rather at the SOA-side than at the client-side, renders distributed atomic transactions insensible to unreliable wireless connections. By contrast, if the transaction protocol communication ran over an unstable wireless connection the probability of transaction aborts due to connection losses would increase.

## 2 Mobile Client Framework

The mobile client framework consists of a generic architecture that supports easy maintainability and expandability. So new use cases (e.g. a cafeteria schedule) can be integrated easily without recompilation of the framework. In general, the client framework provides several generic features. First of all, business data that is received from a service is cached locally. Like this, the user may queue data synchronizations and access the data which is organized in the memory of the mobile device while being offline. Furthermore, by defining a Meta model for the data binding, nominal and structural information about data from a database entity type is available at the mobile application to generate graphical representations according to the access rights of particular data. The mobile application supports an efficient on the fly GUI interpretation by reading the Meta data and the business data from the memory of the mobile device. Multidimensional relational data is displayed in such way that the relation is preserved graphically as well. Moreover, the mobile application is able to synchronize business data with a service in the SOA. Therefore, services which realize CRUD logic for data access are also supported by the framework. The fact that data is manipulated by different client interfaces may lead to inconsistencies due to out-of-date mobile data. To bypass this issue, a conflict resolution mechanism, which is controlled by the user, has been implemented at the mobile client.

## 3 Conclusion

Supporting mobile access to a distributed information system using a stand-alone application poses a challenge, both for the client application and for the service infrastructure. To enable future changes – especially the adoption to new use cases – the client and the communication interface provide generic architectures. Thus, we can benefit from the advantages of a stand-alone client (caching the data and thereby becoming more autonomous) and stay flexible for future maintainability and expandability. Finally, platform independent data-synchronization of .NET and J2ME applications using a SOA has been realized successfully, based on the interoperability of the SOA and the proposed generic architectures.

## Bibliography

- [AH02] Azim, O.; Hamid, A. K.: Caching soap services. [Online] <http://www.javaworld.com/javaworld/jw-03-2002/jw-0308-soap.html> [Last accessed: January 14th, 2008], 2002. Publisher: JavaWorld.
- [Cu02] Curbera F.,Khalaf R., Mukhi N., Tai S., and Weerawarana S.: The Next Step In Web Services. How three specifications support creating robust service compositions. Communications of the ACM, Volume 46, Issue 10, Service-oriented computing, October 2003.
- [Da07] Daum B., Rich-Client-Entwicklung mit Eclipse 3.2. Anwendungen entwickeln mit der Rich Client Platform. 2. Aktualisierte Auflage, dpunkt.verlag GmbH, Heidelberg, 2007.



- [Erl05] Erl, T.: Service Oriented Architecture. Concepts, Technology and Design. Prentice Hall, 2005.
- [LO03] Lange D. B., Oshima M.: Programming and deploying Java mobile agents with aglets. 3<sup>rd</sup> printing, September 2003.
- [Per06] Pernici B.: Mobile Information Systems, Infrastructure and Design for Adaptivity and Flexibility. Springer Berlin, 2006
- [Sin04] Singh I., Stearns B., Brydon S., Murray G., Ramachandran V.: Designing Web Services with the J2EE 1.4 Platform (The Java Series): JAX-RPC, SOAP, and XML Technologies. Book News, Inc., Portland, OR, 2004.

# MetaXsA: Modulares Framework zur Metadatenextraktion durch semantische Dokumentanalyse auf unstrukturierten Textdokumenten

**Benedikt Schmidt**

Universität Paderborn  
Institut für Informatik  
schmidtb@mail.upb.de

**Wolfgang Reinhardt**

Universität Paderborn  
Institut für Informatik  
wolle@uni-paderborn.de

**Christian Mletzko**

Universität Paderborn  
Institut für Informatik  
letris@uni-paderborn.de

**Abstract:** Die Zugriffsunterstützung auf Dokumente ist im Zuge großer Mengen unstrukturierter Dokumente zur Herausforderung geworden. Vorgestellt wird MetaXsA, ein Werkzeug zur Erzeugung hochwertiger, wohlstrukturierter Metadaten. Ein wesentlicher Bestandteil von MetaXsA ist die semantische Analyse. Diese stellt ein Framework zum Aufbau einer modularen Extraktionspipeline bereit, um semantisch relevante Metadaten automatisch aus Texten zu extrahieren.

## 1 Problemstellung

Informationsniederlegung erfolgt meist in unstrukturierten, natürlichsprachlichen Dokumenten. Im Unternehmensumfeld sind Nutzer durch das Intranet meist mit Dokumenten konfrontiert, die sie nicht selber erzeugt und nicht selbst im Dateisystem abgelegt haben. Eine bloße Abbildung des Dateisystems, unterstützt durch eine indizierte Suche, wie es die meisten Betriebssysteme mittlerweile integrieren, hat Schwächen. Die Nutzung eines hierarchischen Dateisystems bedeutet hohen kognitiven und zeitlichen Aufwand. Ergebnisse von Suchen auf indizierten Dokumenten werden nur Suchen gerecht, die vom Prinzip der Vagheit geleitet sind. Im Unternehmen steht jedoch Aufgabenangemessenheit und zeitliche Effizienz im Vordergrund. Eine spezifische Zugriffsunterstützung erscheint insofern sinnvoll. Grundlage von Zugriffsunterstützung bilden hochwertige, wohlstrukturierte Metadaten. Die Vorteile eines solchen Datensatzes sind weitreichend. Indizierte Suchen auf diesen Metadaten bieten beispielsweise semantisch präzisere Treffer. Kategoriebildung und Klassifizierung von großen Dokumentmengen wird möglich.

Grundlegende Metadaten, meist administrativer Natur, sind häufig bereits im Dokument abgelegt und können durch Nutzung von Metadatenextraktoren [MS08] gewonnen werden. Metadaten, die tiefergehende semantische Informationen über ein Dokument beinhalten, können nur erzeugt werden, wenn der Inhalt von Dokumenten mit einbezogen wird. Im Augenblick wird die Erzeugung derartiger Daten durch Nutzer unter dem Stichwort Tagging vielfach diskutiert. An dieser Stelle wird eine andere Richtung eingeschlagen. Ansätze des *Natural Language Processing* (NLP) bieten die Möglichkeit der automatischen Erzeugung von Metadaten nach kohärenten Mustern.

Im Zuge der Mokex Projektserie [HHM<sup>+</sup>06] wurde mit MetaXsA ein Werkzeug erstellt, das wohlstrukturierte Metadaten erzeugt und in einem LOM [IEE02] Dokument ablegt. MetaXsA arbeitet dabei in zwei Schritten.

Zunächst werden die in einem Dokument abgelegten Metadaten durch Metadatenextraktoren ausgelesen. Im zweiten Schritt wird eine semantische Analyse des Dokuments durchgeführt, die Techniken des NLP verwendet, um Metadaten zu gewinnen.

Dieses Papier beschreibt den Aufbau des Werkzeugs zur semantischen Analyse. Dieser Teil von MetaXsA zeichnet sich insbesondere durch einen modularen Aufbau aus, der eine einfache Erweiterung der semantischen Analyse ermöglicht.

## 2 MetaXsA - Semantische Dokumentanalyse

Dieser Abschnitt stellt das grundlegende Konzept der semantischen Dokumentanalyse vor. Dazu wird auch ein kurzer Überblick zu den aktuell umgesetzten Modulen gegeben. Dies umfasst unter anderem die Erkennung der Sprache in der ein Dokument verfasst ist, die Extraktion von Stopwörtern unter Berücksichtigung von vorgegebenen Stopwortlisten und Lemmatisierung relevanter Begriffe, eine LDAP synchronisierte Organisationseinheit-Erkennung, so wie eine administrierbare Dokumenttyperkennung.

### 2.1 Das Framework

Das Framework zur semantischen Dokumentanalyse erhält ein natürlichsprachliches Dokument, führt ein Parsing des Dokuments durch und erzeugt Annotationen für vorhandene Sätze und Worte. Anschließend gibt es das Dokument an verschiedene Module zur Informationsextraktion weiter, die in linearer Folge als Pipeline realisiert sind (s. Abbildung 1). MetaXsA implementiert die *Gate API* [CMBT07] als allgemeine Umgebung zur Text-Extraktion. Diese API stellt verschiedene grundlegende Konzepte zur Informationsextraktion bereit, die von den Modulen genutzt werden können.

Die Module stellen die Extraktionsfunktionalität von MetaXsA bereit und ermöglichen die komfortable Erweiterung der Funktionalität. Sie werden auf dem Dateisystem abgelegt und über eine Konfigurationsdatei in MetaXsA integriert. Die Module können Webservices bereitstellen, die automatisch in MetaXsA integriert werden und Administrierbarkeit und Aktualisierung der Modul-Datensätze in Verbindung mit externen Datenbanken ermöglichen.

Jedes Modul zielt auf die Extraktion eines Metadatensatzes ab. Dabei erzeugt es Annotationen am Dokument, das analysiert wird und kann auf Annotationen zurückgreifen, die früher ausgeführte Module abgelegt haben.

Einzelne Module errechnen die Qualität des von ihnen extrahierten Metadatum unter Berücksichtigung der von Modul getroffenen Annahmen. Diese Qualität wird zusammen mit dem Metadatum an das Framework zurückgegeben, das alle Ergebnisse in einer LOM-Datei zusammenfasst.

Sollten Module Ergebnissätze erzeugen, die sich nicht im klassischen LOM-Schema ab-

bilden lassen, wird das LOM-Schema und die LOM-Datei dynamisch um die Ergebnisdatensätze der Module erweitert.

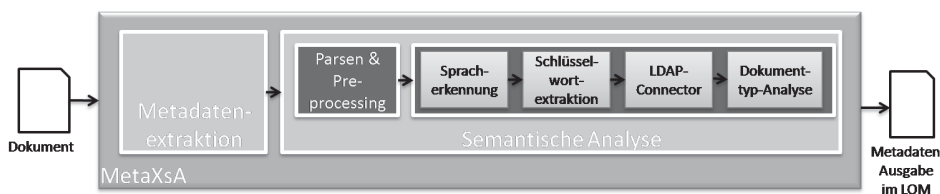


Abbildung 1: MetaXsA Pipeline

## 2.2 Die Module

Die semantische Analyse des im Rahmen des 3. Mokex Projekts entstandenen MetaXsA umfasst sechs Module, von denen eine Auswahl im Folgenden vorgestellt wird.

- **Spracherkennung/POS:** Unter Verwendung von *Aspell* [Atk04] wird eine von der Dokumentgröße abhängige Stichprobe von Wörtern auf Korrektheit für sechs unterschiedliche Sprachen hin untersucht. Die Sprache, bei der die größte Korrektheit vorliegt wird als Dokumentsprache identifiziert. Auf dieser Information basierend wird eine Wortarterkennung durch den Treetagger [ToSfRLtIfCS99] durchgeführt. Das Modul gibt nach außen die Sprache des Dokuments zurück.
- **Schlüsselwort:** Mit diesem Modul werden Schlüsselwörter aus einem Modul extrahiert. Dabei wird einerseits eine von außen administrierbare Liste von Stopwörtern und regulären Ausdrücken in dem Dokument gesucht und bei Zutreffen zurückgeliefert. Weiterhin wird der DURM Lemmatizer [PW05] verwendet, um auf mittels Lemmatisierung relevante Nomen und Nominalphrasen zu bestimmen. Als Auswahlkriterium wird dabei die Häufigkeit des Auftretens verwendet, die auch zurückgegeben wird und Grundlage zur Erzeugung von Schlüsselwortsammlungen sein kann.
- **LDAP Connector:** Mit diesem Modul wird eine Verbindung zu einem LDAP-System aufgebaut und die Organisationseinheiten und die zugehörigen Personen ausgelesen. Das Dokument wird auf diese Daten hin untersucht und es werden Personen und Organisationseinheiten des Unternehmens identifiziert, die mit dem Dokument verbunden sind.
- **Dokumenttyp-Analyse:** Die Dokumenttyp-Analyse ermittelt Dokumenttypen auf Grundlage von Stopwortlisten und regulären Ausdrücken, die jeweils einer Dokumentklasse zugeordnet sind. Das Dokument wird auf diese Begriffe hin untersucht. In Abhängigkeit von den Worttypen in den Listen wird für jede Dokumentklasse

ein Wahrscheinlichkeitswert berechnet, auf dessen Grundlage der Dokumenttyp bestimmt wird. Die zugrunde liegenden Listen sind von außen über Webservices voll administrierbar, können erzeugt und gelöscht werden.

### 3 Zusammenfassung und Ausblick

Die von MetaXsA erzeugten Metadaten besitzen bereits im Rahmen der umgesetzten Module eine hohe Qualität und ermöglichen effizienten Dokumentzugriff über Suchen und Klassifikationen über das im Rahmen des Mokex Projekts umgesetzte DyOgeneS [BS08]. Dabei werden die gewonnenen Schlüsselwörter verwendet, um die Dokumente untereinander über eine Linkstruktur zu verbinden. Durch die einfache Erweiterbarkeit von MetaXsA und die Möglichkeit in späteren Prozessschritten auf Annotationen früherer Module zurückzugreifen, sind umfassende Erweiterungen möglich.

### Literatur

- [Atk04] Kevin Atkinson. GNU Aspell. <http://aspell.net/>, 2004.
- [BS08] Thomas Bremes und Lars Scheuermann. DyOgeneS: Generierte Oberflächen in einem verteilten automatisierten System. In *eingereicht Informatiktage 08*, 2008.
- [CMBT07] Hamish Cunningham, Diana Maynard, Kalina Bontcheva und Valentin Tablan. Developing Language Processing Components with GATE Version 4. <http://www-odp.tamu.edu/isg/appsdev/docs/gate.pdf>, 2007.
- [HHM<sup>+</sup>06] Knut Hinkelmann, Kai Holzweißig, Johannes Magenheimer, Fabian Probst und Wolfgang Reinhardt. Linking Communities of Practice with Learning Communities in Computer Science Education. In *Education for the 21st Century — Impact of ICT and Digital Resources*. 2006.
- [IEE02] IEEE. Draft Standard for Learning Object Metadata. [http://ltsc.ieee.org/wg12/files/LOM\\_1484\\_12\\_1\\_v1\\_Final\\_Draft.pdf](http://ltsc.ieee.org/wg12/files/LOM_1484_12_1_v1_Final_Draft.pdf), 2002.
- [MS08] Christian Mletzko und Benedikt Schmidt. Automatische Extraktion von Metadaten aus verschiedenen Dateiformaten. In *eingereicht Informatiktage 08*, 2008.
- [PW05] Praharsana Perera und René Witte. A Self-Learning Context-Aware Lemmatizer for German. In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, 2005.
- [Tre99] Treetagger. Textcorpora und Erschließungswerkzeuge. and University of Stuttgart and Institute for Romance Linguistics and the Institute for Computer Science, <http://www.ims.uni-stuttgart.de/projekte/tc/>, 1999.

# Anfrageauswertung in nativen XML-Datenbankverwaltungssystemen

Andreas M. Weiner

AG Datenbanken und Informationssysteme  
Fachbereich Informatik  
Technische Universität Kaiserslautern  
D-67653 Kaiserslautern, Deutschland  
weiner@informatik.uni-kl.de

## 1 Motivation

Die *Extensible Markup Language* (XML) hat sich seit der ersten *W3C Recommendation* im Jahre 1998 zu einem De-facto-Standard für die Repräsentation und zum Austausch von strukturierten und semistrukturierten Daten entwickelt. Die Ursprünge dieser Sprache liegen bereits weit zurück. Einer ihrer Vorgänger – die *Standard Generalized Markup Language* (SGML) – wurde bereits in den 70er Jahren des vergangenen Jahrhunderts vorgestellt.

Heute existiert für Geschäftsobjekte nahezu jeder Branche eine standardisierte XML-Repräsentation, die den Datenaustausch zwischen und innerhalb von Unternehmen erleichtern soll. Mit der Einführung von *Web Services* und *serviceorientierten Architekturen*, bei denen XML-Nachrichten ausgetauscht werden, eröffnet sich ein weiterer Anwendungsbereich der Sprache XML.

Viele Anwendungsbereiche benötigen effiziente Verfahren zur Bearbeitung und Verwaltung sowie zur Durchsuchung von XML-Dokumenten. Obwohl XML ursprünglich als Datenaustauschformat gedacht war, wird aus diesem Grund seit einiger Zeit nach Möglichkeiten gesucht, um XML-Dokumente effizient zu speichern und zu verwalten.

*Relationale Datenbankverwaltungssysteme* (RDBVS) bilden seit Jahren eine zuverlässige Basis für die Speicherung und Verwaltung von großen Datenbeständen. Da XML-Dokumente eine Baumstruktur aufweisen, ist deren Speicherung in relationalen Datenbanksystemen nicht ohne Probleme möglich. Deshalb ist gegebenenfalls die Aufteilung des Dokuments auf eine Vielzahl von Tabellen erforderlich. Diese Aufteilung erschwert die Transaktionsisolation in RDBVS in erheblichem Maße [HH07]. Ebenso schwierig gestaltet sich der direkte Zugriff mit Hilfe einer XML-Anfragesprache (vgl. ebd.).

Um eine ständige und teure Übersetzung zwischen dem Relationenmodell auf der einen Seite und der XML-Repräsentation auf der anderen Seite zu vermeiden, werden so genannte *native XML-Datenbankverwaltungssysteme* (XDBVS) als vielversprechende Alternative vorgeschlagen [HH07]. Diese Systeme bieten den Vorteil, dass sie XML-Dokumente

derart speichern können, dass diese ihre hierarchische Struktur beibehalten. Durch die ordnungserhaltende Speicherung erwartet man, dass Anfragen mit einer XML-Anfragesprache wie zum Beispiel *XQuery* schneller beantwortet werden können, als dies mit Hilfe eines RDBVS möglich wäre, da hier der Umweg über das Relationenmodell nicht notwendig ist.

Im Gegensatz zur Anfrageverarbeitung in relationalen Datenbankverwaltungssystemen gestaltet sich die Anfrageauswertung auf XML-Dokumenten deutlich komplizierter. Die höhere Komplexität hat einerseits ihren Ursprung im XML-Datenmodell und wird andererseits durch die äußerst komplexe Anfragesprache *XQuery* bedingt. Die aktuelle Forschung auf diesem Gebiet lässt sich grob in zwei Klassen einteilen: Zum einen gibt es Forschungsvorhaben, die sich mit der Entwicklung von effizienten physischen *Pfadverarbeitungsoperatoren* (engl. *Path Processing Operators*) beschäftigen. Als prominente Vertreter dieser Ansätze lassen sich die *Structural-Join-Algorithmen* von Al-Khalifa et al. [AKJP<sup>+</sup>02] und die *Holistic-Twig-Join-Algorithmen* von Bruno et al. [BKS02] nennen. Auf der anderen Seite existieren viele Ansätze, die sich dem Problem der Anfrageoptimierung von der logischen Seite nähern, indem logische Algebren vorgestellt werden, mit deren Hilfe eine algebraische Optimierung der Anfrage durchgeführt werden kann. Dabei lassen sich baumbasierte Algebren wie zum Beispiel *TAX* von Jagadish et al. [JLST01] beziehungsweise *TLC* von Paparizos et al. [PWLJ04] und tupelbasierte Algebren wie zum Beispiel *NAL* von Brantner et al. [BHKM05] unterscheiden [Mat07].

Zwischen diesen beiden Forschungsbereichen gibt es derzeit nur sehr wenige Berührungspunkte. Für die erfolgreiche Entwicklung eines kostenbasierten Anfrageoptimierers müssen diese beiden Bereiche aber eng zusammenarbeiten. Erste Ansätze zur Zusammenführung dieser Bereiche wurden von Mathis mit *NAL<sup>STJ</sup>* [Mat07] und von Michiels et al. [MMS07] vorgestellt. Bei der tupelbasierten Algebra *NAL<sup>STJ</sup>* wird ein Structural-Join-Operator in die Algebra eingeführt. Dagegen wird beim Ansatz von Michiels et al. ein Operator zur Verarbeitung von *Baummustern* (engl. *tree patterns*) in die logische Algebra integriert. Mit diesen Ansätzen wurde eine Brücke zwischen den beiden Forschungsrichtungen geschlagen und der Weg für eine kostenbasierte Optimierung von *XQuery*/*XPath*-Anfragen bereitet. Im nächsten Schritt muss dieser Weg weiter gefestigt werden, so dass ein durchgängiger Prozess beschrieben werden kann, beginnend beim Empfang und der Übersetzung der Anfrage, über die kostenbasierte Anfrageoptimierung reichend, bis hin zu einem fertigen Anfrageausführungsplan.

## 2 Anfragerestrukturierung im XML-Anfragegraphmodell

Im Rahmen der Diplomarbeit [Wei07] hat der Autor Konzepte zur Anfrageauswertung in nativen XDBVS untersucht. Das Hauptaugenmerk lag dabei auf der Plangenerierung. Der Plangenerator bildet eine der wichtigsten Komponenten eines kostenbasierten Anfrageoptimierers. Dieser bewertet die ihm vorliegenden Anfragepläne anhand von statistischen Informationen und wählt den günstigsten Plan zur Ausführung aus.

Zunächst wurde ein *XML-Anfragegraphmodell* (XAGM) als neue Interndarstellung für *XQuery*/*XPath*-Anfragen vorgestellt (vgl. Abbildung 1), das die Ideen von Mathis [Mat07]

und Michiels et al. [MMS07] hinsichtlich der Einbettung neuer Operatoren in die logische Algebra aufgreift und erweitert. Dabei wird der (binäre) Structural-Join-Operator durch einen indexbasierten  $n$ -Wege-Join-Operator ergänzt, der mit Hilfe eines physischen Holistic-Twig-Join-Operators effizient ausgewertet werden kann. Das XAGM bildet die Grundlage für die Optimierungsphase der Anfrageverarbeitung in nativen XDBVS, bei der eine Anfragerestrukturierung und Anfragetransformation durchgeführt wird. Für das XAGM wurden eine Reihe von Restrukturierungsregeln [WMH08] definiert, die eine algebraische Optimierung der Anfragen unter Berücksichtigung von Implementierungsaspekten wie zum Beispiel der Verfügbarkeit von  $n$ -Wege-Verbund-Operatoren ermöglichen. Bei der Restrukturierung wird dabei das Ziel verfolgt, möglichst viele aufeinander folgende Structural-Join-Operatoren durch einen einzigen Holistic-Twig-Join-Operator zu ersetzen. Ferner wird versucht, möglichst viele verfügbare Indexstrukturen zur Anfrageauswertung heranzuziehen. Des Weiteren wurden Regeln zur nicht-algebraischen Optimierung von Anfragegraphen definiert, wie zum Beispiel zur Vertauschung der Verbundreihenfolgen und Verbundpartner. Abschließend wurde eine Abbildung der logischen Operatoren des XAGM auf physische Planoperatoren beschrieben. Aufgrund eines fehlenden Kostenmodells handelt es sich dabei derzeit um eine heuristikbasierte Abbildung.

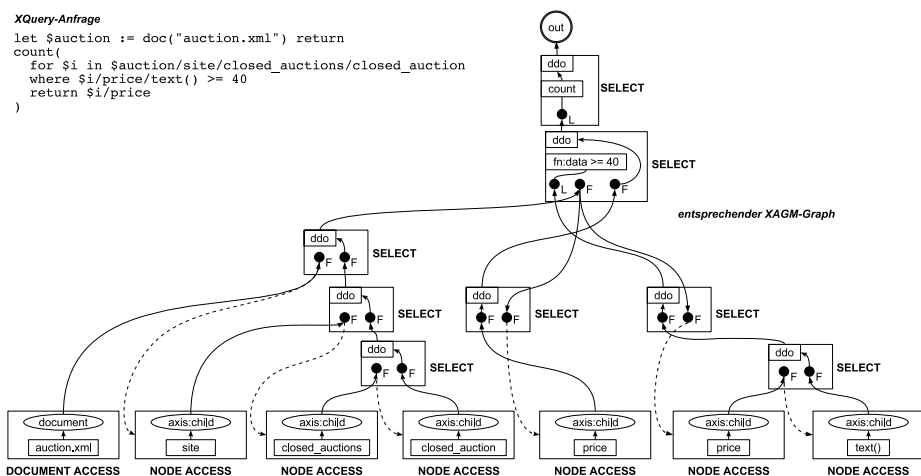


Abbildung 1: XQuery-Anfrage und entsprechender Anfragegraph vor der Anfragerestrukturierung

### 3 Ausblick auf zukünftige Arbeiten

Mit Hilfe des XAGM als neue Interndarstellung für XQuery-Anfragen und den beschriebenen Restrukturierungsregeln kann nun mit der Entwicklung eines kostenbasierten Anfrageoptimierers begonnen werden, der einen wichtigen Teil der Anfrageauswertungskomponente des *XML Transaction Coordinator* (XTC)<sup>1</sup> [HH07] bilden wird. Dabei handelt es sich um den Prototypen eines nativen XDBVS, das derzeit an der TU Kaiserslautern in

<sup>1</sup>Projekt-Webseite: <http://www.project-xtc.de>



der Arbeitsgruppe *Datenbanken und Informationssysteme* entwickelt wird. In einem ersten Schritt muss dabei eine geeignete Hauptspeicherrepräsentation entwickelt werden, die einen möglichst großen Teil des Suchraums der semantisch äquivalenten Anfragegraphen effizient verwalten kann. Nachdem mehrere alternative Ausführungspläne repräsentiert werden können, muss ein Kostenmodell definiert werden, das eine Bewertung der einzelnen Alternativen zulässt und die Auswahl des „optimalen“ Ausführungsplans gestattet. Abschließend ist dann die Güte der Anfrageausführungspläne zu vergleichen, die durch verschiedene Plangenerierungsstrategien erzeugt wurden.

## Literatur

- [AKJP<sup>+</sup>02] Shurug Al-Khalifa, H. V. Jagadish, Jignesh M. Patel, Yuqing Wu, Nick Koudas und Divesh Srivastava. Structural Joins: A Primitive for Efficient XML Query Pattern Matching. In *ICDE '02: Proceedings of the 18th International Conference on Data Engineering, 26 February – 1 March, 2002, San Jose, CA*, Seiten 141–154, Washington, DC, USA, 2002. IEEE Computer Society Press.
- [BHKM05] Matthias Brantner, Sven Helmer, Carl-Christian Kanne und Guido Moerkotte. Full-fledged Algebraic XPath Processing in Natix. In *ICDE '05: Proceedings of the 21st International Conference on Data Engineering, ICDE 2005, 5–8 April 2005, Tokyo, Japan*, Seiten 705–716, Washington, DC, USA, 2005. IEEE Computer Society Press.
- [BKS02] Nicolas Bruno, Nick Koudas und Divesh Srivastava. Holistic Twig Joins: Optimal XML Pattern Matching. In *SIGMOD '02: Proceedings of the 2002 ACM SIGMOD International Conference on Management of Data, Madison, Wisconsin, June 3–6, 2002*, Seiten 310–321, New York, NY, USA, 2002. ACM Press.
- [HH07] Michael Haustein und Theo Härder. An Efficient Infrastructure for Native Transactional XML Processing. *Data & Knowledge Engineering*, 61(3):500–523, 2007.
- [JLST01] H. V. Jagadish, Laks V. S. Lakshmanan, Divesh Srivastava und Keith Thompson. TAX: A Tree Algebra for XML. In *DBPL '01: Database Programming Languages, 8th International Workshop, Frascati, Italy, September 8-10, 2001, Revised Papers, Lecture Notes in Computer Science Bd. 2397*, Seiten 149–164, Berlin, Heidelberg, 2001. Springer Verlag.
- [Mat07] Christian Mathis. *Extending a Tuple-Based XPath Algebra to Enhance Evaluation Flexibility*. In *Informatik – Forschung und Entwicklung*, 21(3–4):147–164, 2007.
- [MMS07] Philippe Michiels, George A. Mihaila und Jérôme Siméon. Put a Tree Pattern in Your Algebra. In *ICDE '07: IEEE 23rd International Conference on Data Engineering, Istanbul, Turkey, April 15–20*, Seiten 246–255, Washington, DC, USA, 2007. IEEE Computer Society Press.
- [PWLJ04] Stelios Paparizos, Yuqing Wu, Laks V. S. Lakshmanan und H. V. Jagadish. Tree Logical Classes for Efficient Evaluation of XQuery. In *SIGMOD '04: Proceedings of the ACM SIGMOD International Conference on Management of Data, Paris, France, June 13–18, 2004*, Seiten 71–82, New York, NY, USA, 2004. ACM Press.
- [Wei07] Andreas M. Weiner. Plangenerierung zur Anfrageauswertung in nativen XML-Datenbankverwaltungssystemen. *Diplomarbeit, Juli 2007, Fachbereich Informatik, TU Kaiserslautern*, 2007.
- [WMH08] Andreas M. Weiner, Christian Mathis und Theo Härder. Rules for Query Rewrite in Native XML Databases. Erscheint in: *Proceedings of the EDBT Workshops, Third International Workshop on Database Technologies for Handling XML Information on the Web (DataX 2008), March 25, 2008, Nantes, France*, 2008.

# Optimierte Umsetzung eines Telekommunikations-Protokolls aus dem Bereich Internet-Telefonie auf Mikrocontrollern

Sebastian Blumenthal

Fachhochschule Bonn-Rhein-Sieg  
Fachbereich Informatik  
sebastian.blumenthal@smail.inf.fh-brs.de

Art der Arbeit: Bachelorarbeit  
Betreuer der Arbeit: Prof. Dr.-Ing. Norbert Jung  
Prof. Dr. Thomas Breuer

**Abstract:** Gegenstand der hier vorgestellten Bachelorarbeit ist die Erstellung eines Realisierungskonzepts und der Implementation des Kommunikation-Gateway-Protokolls MEGACO auf einer spezialisierten (embedded) Telekommunikations-Hardware-Plattform. Die angestrebte Anwendung ist die Verbindung von klassischen Telefonnetzen mit der Internet-Telefonie.

## 1. Einleitung

Neben der herkömmlichen, klassischen Telefonie hat sich Voice over IP (VoIP) als weitere Sprachkommunikationstechnologie entwickelt, deren Bedeutung immer stärker wächst. Gerade im Hinblick auf die zunehmende Konvergenz der Telekommunikationsnetze besteht eine besondere Anforderung an die Protokolle darin, verschiedenartige Netze, insbesondere paketvermittelnde IP- und leitungsvermittelnde Telefonnetze, zu koppeln.

Um die Kommunikation mit VoIP zu ermöglichen, werden bestimmte Protokolle und Architekturen benötigt. Es gibt bereits einige Protokolle die zwei Teilnehmer befähigen ein Telefonat mittels VoIP auch über heterogene Kommunikationsnetze zu führen. Dazu zählen unter anderem der H.323 Standard, ein proprietäres Protokoll von CISCO mit dem Namen „Skinny“, das Media Gateway Control Protocol (MGCP) und das aus MGCP entwickelte Nachfolgeprotokoll „MEGACO“ (**M**edia **G**ateway **C**ontrol).

Zwischen den unterschiedlichen Netzen residieren als Vermittlungsknoten spezielle Gateways. Um diese effizient zu implementieren, wird leistungsfähige, spezialisierte „embedded“ Hardware benötigt. Von den zuvor genannten Protokollen bietet sich MEGACO als am geeignetsten an, um als Kommunikationsprotokoll der Gateways zu dienen. Dieses Protokoll hat den Vorteil, dass es weniger komplex als zum Beispiel H.323 ist und es wird besonders auf das Ressourcenmanagement geachtet, welches die begrenzten Ressourcen einer embedded Plattform berücksichtigt. [GHP00, S. 5]

## 2. Das MEGACO Protokoll

MEGACO ist ein Protokoll, dass sich mit der Kommunikation zwischen Gateways befasst. Ziel ist es hierbei, die Intelligenz eines Netzwerkes an einem zentralen Punkt zu konzentrieren. An den Rändern des Netzes befinden sich Instanzen, die von einer zentralen Instanz gesteuert werden. Um dies zu realisieren, basiert MEGACO auf einer Client-Server Architektur.

### 2.1. Architektur

Die zentrale Komponente der Client-Server Architektur (vgl. Abbildung 1) bildet der **Media Gateway Controller** (MGC). Sein Aufgabenbereich umfasst die Kontrolle und Steuerung der Client-Elemente, der so genannten Media Gateways (MGs), und die Rufsignalisierung<sup>1</sup>. Als Signalisierung können beliebige Protokolle aus der leitungs- oder paketvermittelnden Domäne verwendet werden, wie zum Beispiel das *Session Initiation Protocol* (SIP) als bekannter Vertreter der Internet-Telefonie-Signalisierungsprotokolle.

Die eigentliche Sprachübertragung wird von den **Media Gateways** übernommen. Untereinander kommunizieren sie über das IP-basierte Sprachübertragungsprotokoll *Real Time Protocol* (RTP). Wenn ein MG an ein nicht auf IP-basierendes Netz angeschlossen ist, so muss der Sprachdatenstrom in das entsprechende Format des anderen Netzwerkes konvertiert werden.

Die Kommunikation zwischen einem MGC und „seinen“ zu kontrollierenden MGs wird über die in dem MEGACO Protokoll definierten Nachrichten abgehandelt. Ziel der Kommunikation ist es, eine Sprachverbindung von mehreren Endpunkten in dem logischen Verbindungsmodell von MEGACO aufzubauen. [Ba05, S. 300ff./Nö03, S. 114f.]

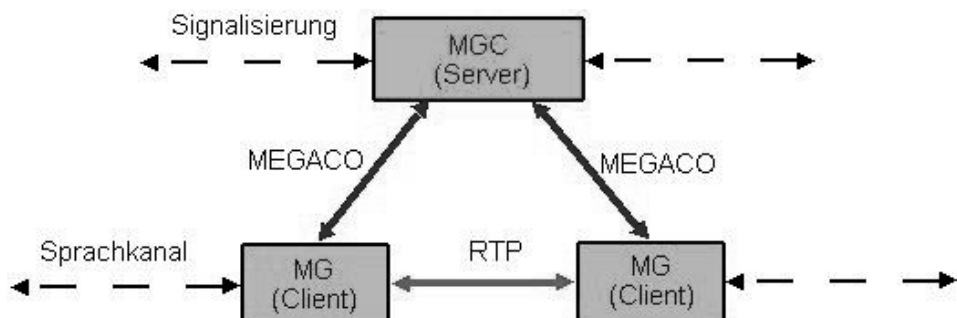


Abbildung 1: Architektur von MEGACO

<sup>1</sup> Rufsignalisierung umfasst den Verbindungsaufbau und -abbau.

## 2.2. Verbindungsmodell

MEGACO benutzt ein abstraktes Verbindungsmodell, das sich im wesentlichen aus zwei Kernelementen zusammensetzt: *Terminations* und *Contexts*. *Terminations* sind die logischen Quellen und Senken der Kommunikation, also Anschlüsse an den Gateways. Sie können Telefonendgeräte oder weitere Media Gateways repräsentieren.

Die Verbindung zwischen zwei oder mehreren Terminations in einem Media Gateway besteht genau dann, wenn sie in einem Kontext, der als *Context* bezeichnet wird, zu einander stehen. Gewissermaßen werden die Ports an einem MG miteinander verschaltet. [Gr03, S. 13ff.]

## 3. Realisierungskonzept und Implementierung

Das hier vorgestellte Konzept (vgl. Abbildung 2) umfasst die Funktionalität eines Vermittlungsgateways, das ein ISDN mit einem VoIP Netz (mit SIP als Signalisierungsprotokoll) verbinden kann. Als Hardwareplattform kommen Komponenten der Gesellschaft für Netzwerk- und Automatisierungstechnologie mbH (N.A.T GmbH) zum Einsatz.<sup>2</sup>

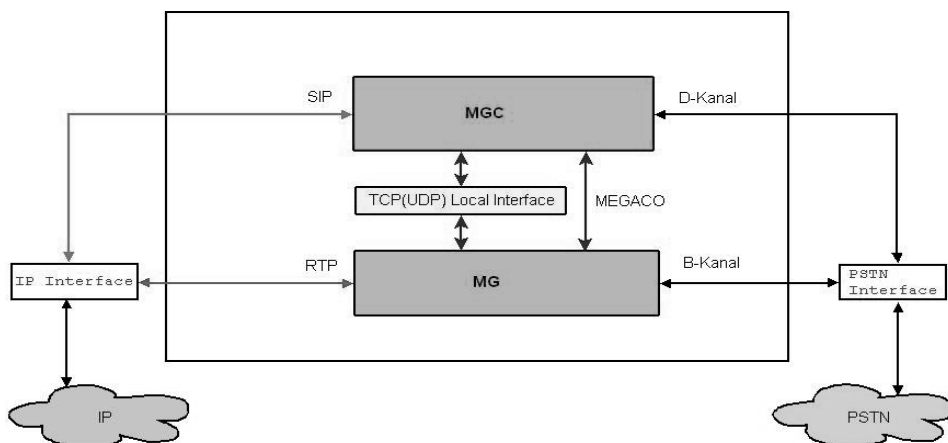


Abbildung 2: Konzept für einen Vermittlungsgateway

Es werden ein Media Gateway für die Sprachübertragung und ein Media Gateway Controller zur Steuerung des MGs und für die Signalisierung benötigt. Beide Instanzen sollen als unabhängige Prozesse auf einer einzigen Plattform laufen. Ein Echtzeit-Micro-Kernel<sup>3</sup> bietet hierfür die benötigte Funktionalität zur Prozessverwaltung. [Na06, S. 7ff.]

<sup>2</sup> Die Bachelorarbeit basiert auf Erfahrungen, die in Zusammenarbeit bei einer Kooperation mit der N.A.T. GmbH entwickelt wurden.

<sup>3</sup> Ein Betriebssystem mit minimaler Funktionalität wird als „Micro-Kernel“ bezeichnet.

Der MGC benötigt eine als *Signaling Gateway* bezeichnete Instanz, um die SIP-Nachrichten in entsprechende ISDN-Signalisierungskanal (D-Kanal) Nachrichten und umgekehrt umwandeln zu können.

Der MG muss die Sprachdaten vom ISDN Sprachkanal (B-Kanal) in die Nutzdaten einer RTP-Session überführen (und umgekehrt). Die Verwaltung der logischen Elemente *Termination* und *Context* ist ebenfalls in dem Prozess für den MG implementiert.

Die Kommunikation über das MEGACO Protokoll wird über ein lokales TCP/IP Loopback-Interface ermöglicht. Alternativ könnte die Kommunikation auch mit klassischer Interprozesskommunikation realisiert werden. Sie ist durch den geringeren Protokolloverhead performanter. Die auf TCP/IP basierende Variante bietet sich jedoch auf Grund ihrer größeren Flexibilität an, weil so leichter auch verteilte Szenarien möglich sind. Das bedeutet, der MGC läuft als einzige Instanz auf dem System und ein oder mehrere MGs residieren auf anderen Systemen.

## 4. Fazit und Ausblick

Das hier vorgestellte Konzept ermöglicht einen Kommunikationsgateway auf Basis des MEGACO Protokolls zu realisieren. Der Gateway kann ISDN mit VoIP Netzen miteinander verbinden. Die Implementierung der Kernelemente des MEGACO Protokolls konnten den Ablauf der Rufsignalisierung auf spezialisierter Telekommunikationshardware zeigen.

Diese Bachelorarbeit kann als Basis für eine Implementation eines Vermittlungsgateways nach MEGACO betrachtet werden. Darüber hinaus sind auch verteilte Szenarien mit mehreren Hardwareplattformen denkbar. In diesem Fall kann man die Flexibilität der Client-Server Architektur von MEGACO ausnutzen.

## 5. Literaturverzeichnis

- [Ba05] Badach, A.: Voice over IP: Die Technik, 2. Auflage, Carl Hanser Verlag, München/Wien, 2005
- [GHP00] Gurle, D.; Hersent, O.; Petit, J.: IP Telephony: Paket-based multimedia communications systems, Pearson Education Limited, Großbritannien, 2000
- [Gr03] Groves, C. et al.: Gateway Control Protocol Version 1, Network Working Group, RFC 3525, Internet Engineering Task Force, 6/2003
- [Na06] N.A.T. GmbH (Hrsg.): N.A.T. OK1 Reference Manual, Version 1.1, o. V., o. O., 2006
- [Nö03] Nölle, J.: Voice over IP, VDE Verlag GmbH, Berlin/Offenbach, 2003

# CSI: PC<sup>2</sup> - A High Performance Biometric System

Samir Brulic, Samira Brulic, Pascal Deppe, Dominic Eschweiler, Rosbeh Etemadi,  
Klaus Herbold, Christoph Konersmann, Robert Meiche, Christoph Scholz,  
Nils Timm, André Ückermann, Elmar Weber, René C. Zorn

{samir | sbrulic | bluesky | weiler | frost77 | klausher | c\_k | rmeiche |  
cscholz | timm84 | auecker | eweber | zorn }@uni-paderborn.de

## 1 Einführung

Biometrische Systeme gewinnen, vor allem im Bereich der Kriminalitätsverfolgung und Sicherheit, immer mehr an Bedeutung. Mit zunehmender Verbreitung steigen auch die Anforderungen an biometrische Systeme: Die Anzahl der biometrischen Templates erreicht zig Millionen und Anfragen müssen in sekundenschnelle beantwortet werden. Aktuelle Systeme sind hierfür nicht ausgelegt. So braucht z. B. das System des FBI's zwischen 2 (Polizeibehörden) und 24 (zivile Behörde) Stunden um eine Anfrage auf 50.000.000 Fingerabdrücke zu beantworten[FB05].

In der Projektgruppe CSI: PC<sup>2</sup>[CSI07] entwickeln seit April 2007 13 Studenten ein biometrisches System, welches durch Cluster Parallelisierung und unter Verwendung von FPGAs dieses Problem angeht. In dem Projekt werden Fingerabdrücke und Iriden als biometrische Merkmale betrachtet.

Die Ziele der Projektgruppe sind zwei: Hohe Performance und gute Erkennungsgenauigkeit. Eine hohe Performance ist nötig bei der Identifikation einer Person anhand von biometrischen Merkmalen, d. h. wenn z. B. ein am Tatort gefundener Fingerabdruck gegen eine Datenbank mit mehreren Millionen Abdrücken verglichen wird. Eine hohe Erkennungsgenauigkeit wiederum ist speziell dann erforderlich, wenn es um die Verifikation einer Person geht, dazu müssen vor allem „false positives“ verhindert werden.

## 2 Realisierung

Das komplette System wurde in C/C++ realisiert. Abbildung 1 zeigt eine grafische Darstellung der Architektur. Auf der Client Seite existieren Module um Scanner für Fingerabdrücke und Iriden anzusprechen, eine GUI zur Steuerung sowie ein Controller Modul welches den Ablauf koordiniert und die Anfragen an den Server übermittelt. Die Anfragen werden verschlüsselt an den Server gesendet wo der Master Cluster Control Daemon (CCD) die Anfrage entgegen nimmt und per MPI an alle verfügbaren Knoten verteilt.

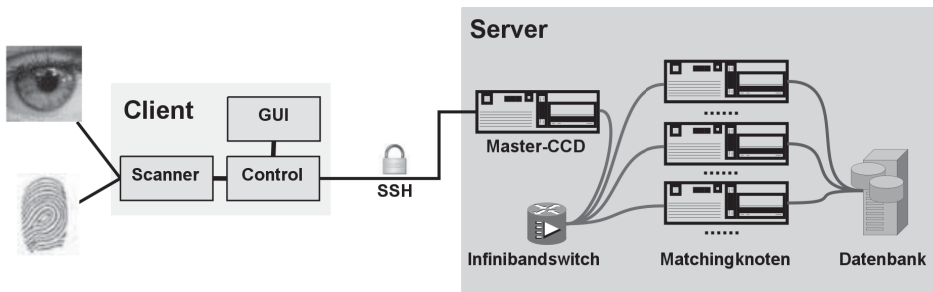


Abbildung 1: Übersicht über die gesamte Architektur des Systems

Anschließend vergleicht jeder Knoten das Template der Anfrage gegen seinen Teil der Templatedatenbank. Alle Ergebnisse über einem gewissen Schwellwert werden per MPI zurück an den Master-CCD geschickt, welcher die besten kombiniert und das Ergebnis an den Client zurückschickt.

Auf der Cluster Seite wird beim Start die komplette Datenbasis gelesen und per MPI im Round Robin Stil an alle Knoten verteilt. Da jeder Knoten nur einen Teil der Templatedatenbank erhält können die Templates i. d. R. problemlos im RAM gehalten werden.

Neben der Möglichkeit Anfragen zu stellen kann der Client auch die Datenbasis des Servers um neue Templates erweitern. Hierzu wird ein vom Scanner aufgenommenes Bild des biometrischen Merkmals (Fingerabdruck oder Iris) an den CCD weitergeleitet. Der CCD führt dann eine Feature Extraction durch, wandelt also das Bild in ein Template für den späteren Vergleich um und speichert das Template zusammen mit dem Bild in der Datenbasis.

Die Datenbasis besteht aus den aufgenommenen Bildern und den extrahierten Templates für Fingerabdrücke und Iriden. Die Templates werden binär kodiert hintereinander in eine Flatfile Datenbank geschrieben um ein effizientes Auslesen der Datenbank beim Start des Servers zu ermöglichen. Um die Startzeit des Servers weiterhin zu optimieren besteht die Möglichkeit die Datenbank auf mehrere Dateien zu verteilen, so dass ein paralleles Einlesen möglich ist.

Durch die Verwendung einer Client/Server Architektur ist es problemlos möglich den Client auszutauschen um z. B. sowohl ein Terminal als auch ein mobiles Gerät mit dem gleichen Server zu nutzen.

Wie anfangs erwähnt wollen wir eine Beschleunigung des Matchingvorgangs erreichen indem wir zusätzlich FPGAs verwenden. Hierzu sind einige Knoten des PC<sup>2</sup> Clusters mit FPGA Boards von Alpha Data [Dat08] ausgestattet welche über den PCI-X Bus mit dem System kommunizieren. Die Beschleunigung der FPGA Variante der Matchingalgorithmen gegenüber der Software Variante basiert im Prinzip darauf, dass wir auf einem FPGA die Möglichkeit haben mehrere Attribute parallel zu vergleichen. Hierdurch ist ein hoher Speedup gegenüber der Software Variante möglich (siehe Abschnitt 3).

Neben der Architektur sind die verwendeten Matchingalgorithmen mit der wichtigste Fak-

tor für die Performance und vor allem für die Erkennungsgenauigkeit. Für die Fingerabdrücke verfolgen wir drei verschiedene Ansätze. Der erste ist der in der Praxis weitverbreitete und erprobte Ansatz der die besonderen Punkte eines Fingerabdrucks (Minutien) miteinander vergleicht[RJR95, JFRQ06]. Als zweiten Ansatz setzen wir auf ein Clustering Verfahren, welches durch eine Baumstruktur die Suchmenge in wenigen Schritten von mehreren Millionen auf einige Tausend reduziert[LJK07]. Der dritte Ansatz ist ein relativ neuer Matching Ansatz welcher auf einem Bildfilter basiert, der die Grauwerte der Fingerabdrücke vergleicht[MJ02]. Durch die Verwendung von drei Ansätzen bauen wir auf eine Erhöhung der Erkennungsgenauigkeit durch einen Kreuzvergleich aller Verfahren. Weiterhin haben wir eine Auswahl und können den besten Algorithmus nutzen. Für das Matching der Iriden verwenden wir den Algorithmus von Daugman, dem de facto Standard in dem Bereich[Dau04].

### 3 Bisherige Ergebnisse und Ausblick

Die bisherigen Ergebnisse unserer Projektgruppe sind viel versprechend. Auf unserem Testcluster<sup>1</sup> erreichen wir pro CPU über 60.000 Vergleiche pro Sekunde bei den Fingerabdrücken. Konkret bedeutet dies, dass die unoptimierte Software Variante des auf Minutien basierenden Algorithmus (Minutienmatcher) es schafft, eine Datenbasis von 50.000.000 Fingerabdrücken auf 32 Knoten (64 CPUs) innerhalb von 13 Sekunden zu vergleichen. Die FPGA Variante des gleichen Algorithmus beschleunigt das ganze um den Faktor 27.

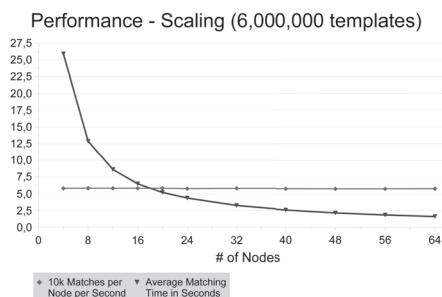


Abbildung 2: Skalierung und Durchsatz

Viel versprechende Ergebnisse erhalten wir auch bei dem Clustering Verfahren für Fingerabdrücke. Auf zwei Knoten (4 CPUs) und einer Datenbank mit 2.000.000 Fingerabdrücken dauert ein Vergleich nur 0,25 Sekunden. Da das Clustering unter Umständen mehrere Tausend Fingerabdrücke zurückliefert werden diese anschließend mit dem Minutienmatcher verglichen um eine bessere Erkennungsgenauigkeit zu erreichen.

Ein weiteres wichtiges Ergebnis ist, dass sich unser System proportional zur Anzahl der Clusterknoten skaliert, man also beliebig viele Knoten nehmen kann ohne Performance zu verlieren (vgl. Abbildung 2).

Der Abschluss der Projektgruppe CSI: PC<sup>2</sup> ist Mitte April 2008, bis dahin werden wir weiter an der Erkennungsgenauigkeit und der Performance arbeiten. Aktuelle Ergebnisse und Entwicklungen sowie weitere Dokumentation sind auf der Homepage der Projektgruppe CSI:PC<sup>2</sup>[CSI07] zu finden.

<sup>1</sup>Arminius Cluster des PC<sup>2</sup>, 200 Knoten mit jeweils zwei 64bit Xeon CPUs mit 3.2 GHz, weitere Details siehe [fPCP08]



## Literatur

- [CSI07] Homepage Projektgruppe CSI:PC<sup>2</sup>, 2007. <http://wwwcs.uni-paderborn.de/pc2/teaching/pg-csipc2.html>.
- [Dat08] Alpha Data, 2008, (accessed February 16, 2008). <http://www.alpha-data.com/>.
- [Dau04] John Daugman. How Iris Recognition Works. In *IEEE Transaction on Circuits and Systems for Video Technology*, Jgg. 14 (1), Seiten 21–30, 2004.
- [FBoIF05] Criminal Justice Information Services (CJIS) Division Federal Bureau of Investigation (FBI). Integrated Automated Fingerprint Identification System or IAFIS, 2005. <http://www.fbi.gov/hq/cjisd/iafis.htm>.
- [fPCP08] Paderborn Center for Parallel Computing (PC<sup>2</sup>). Architecture of the ARMINIUS cluster, 2006, (accessed February 16, 2008). <http://wwwcs.uni-paderborn.de/pc2/services/systems/arminius/system-architecture.html>.
- [JFRQ06] Ying Jie, Yuan Yi Fang, Zhang Renjie und Song Qifa. Rapid and brief communication: Fingerprint minutiae matching algorithm for real time system. *Pattern Recogn.*, 39(1):143–146, 2006.
- [LJK07] Manhua Liu, Xudong Jiang und Alex Chichung Kot. Efficient fingerprint search based on database clustering. *Pattern Recogn.*, 40(6):1793–1803, 2007.
- [MJ02] Muhammad Umar Munir und Dr. Muhammad Younus Javed. Fingerprint Matching Using Gabor Filters. *National Conference on Emerging Technologies NCET*, 2002.
- [RJR95] Nalini K. Ratha, Anil K. Jain und Diane T. Rover. An FPGA-based point pattern matching processor with application to fingerprint matching. In *CAMP '95: Proceedings of the Computer Architectures for Machine Perception*, Seite 394, Washington, DC, USA, 1995. IEEE Computer Society.

# Verteilte Ereigniserkennung in Sensornetzen

Norman Dziengel

Georg Wittenburg

Fachbereich Mathematik und Informatik  
Freie Universität Berlin  
Takustraße 9, 14195 Berlin  
{dziengel, wittenbu}@inf.fu-berlin.de

**Zusammenfassung:** Die Beobachtung und Bewertung eines Ereignisses in einem Sensornetz durch mehrere Sensorknoten wird als verteilte Ereigniserkennung bezeichnet. Inhärente, technische Grenzen der Sensorknoten sowie eine wünschenswerte Reduktion der Kommunikation zwischen Sensorknoten zur Energieeinsparung stellen die primäre Herausforderung in Sensornetzen dar. Somit verlangt der Einsatz klassischer Mustererkennungsalgorithmen in einem Sensornetz ohne Basisstation nach neuen Konzepten.

Das hier vorgestellte System zur verteilten Ereigniserkennung ist auf die spezifischen Eigenschaften von Sensornetzen zugeschnitten. Es nutzt die verteilt erfassenden Sensorknoten zur Optimierung der Ereigniserkennung. Innerhalb des Sensornetzes werden die Daten der einzelnen Sensorknoten zusammengeführt und gemäß einer von vier verteilten Fusionsmethoden klassifiziert. Das System wird durch Versuche bezüglich der Genauigkeit der lokalen und verteilten Ereigniserkennung quantitativ evaluiert und mit bestehenden Ansätzen verglichen.

## 1 Motivation und Zielsetzung

Sensornetze bestehen aus zumeist batteriebetriebenen Kleinstcomputern, die ihre Umgebung mit verschiedensten Sensoren überwachen und Informationen untereinander per Funk austauschen oder an eine Basisstation übermitteln. Ereigniserkennung ist das Erkennen von zuvor gelernten Ereignissen durch einen oder mehrere Sensorknoten. Mit dem hier vorgestellten System „Patrec“ könnten beispielsweise mit Beschleunigungssensoren die Vibrationen des Überkletterns eines Zaunes bei einem Einbruch verteilt erkannt werden. Die Idee ist hierbei, dass die Erkennungsrate des verteilten Systems besser ist, als es mit den jeweils einzelnen Sensorknoten möglich wäre. Dadurch würde die gegebene Redundanz der Sensorknoten im Sensornetz in der verteilten Ereigniserkennung wertschöpfend genutzt werden. Sensornetze können nach [Lli97] und [Rus07] zudem mit Datenfusionstechniken energiesparender agieren.

Für die verteilte Ereigniserkennung verwendet man in der aktuellen Forschung in der Regel sehr leistungsstarke und daher energieineffiziente Sensorknoten [Bro03]. Zusätzlich werden häufig zentrale Basisstationen [Wer06] oder mehrere zusätzliche Mikroserver [Bok06] eingesetzt. In Systemen, deren Hardware energiesparende Komponenten nutzt (wie der hier eingesetzte ScatterWeb MSB-430 Sensorknoten [Sch05]), kommen

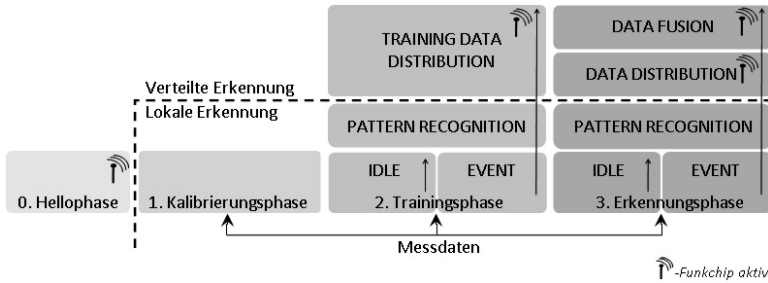


Abbildung 1: Phasen der lokalen und verteilten Ereigniserkennung in Patrec

bisher Schwellenwerte zum Einsatz, um einfache Ereignisse zu definieren [Wit07, Wäl07]. Verteilte Ereigniserkennung in Sensornetzen bedeutet nach dem bisherigen Stand der Technik folglich entweder den Einsatz von aufwendiger Hardware oder eine unzureichende Genauigkeit der Erkennung komplexer Ereignisse.

Mit Patrec wird in dieser Arbeit ein System zur autarken, verteilten Ereignisauswertung in drahtlosen Sensornetzen bei minimalem Hardwareaufwand und ohne Basisstation vorgestellt. Sensorknoten erkennen jeweils lokal Ereignisse und führen anschließend eine verteilte Datenfusion durch, wobei Datenfusionen auf Ebene der Merkmale sowie der Klassifizierung möglich sind. In der Implementierung von Patrec werden hierbei die Messdaten des im ScatterWeb MSB-430 verbauten Drei-Achsen-Beschleunigungssensors verwendet.

## 2 Konzept der lokalen und verteilten Ereigniserkennung

In der verteilten Ereigniserkennung müssen Daten zwischen Sensorknoten ausgetauscht werden. Dies erfordert eine Hellophase, in der sich alle Sensorknoten per Broadcast bekannt machen, siehe Abb. 1. Es folgt eine dynamische Kalibrierungsphase des Beschleunigungssensors, um die Sensoren auf ihre Ruheposition und ggf. einen Zitterbereich einzustellen. Die Musterbearbeitung der Trainings- und Erkennungsphase führt die lokale Datenauswertung nach einem klassischen Mustererkennungsmodell nach [Nie03] durch. Dabei kommt eine Datenvorverarbeitung, Mustersegmentierung, Normierung und eine histogrammbasierte Merkmalsextraktion zum Einsatz.

In der Trainingsphase ist das differenzierte Training von Bewegungsmustern auf den einzelnen Sensorknoten möglich. Abhängig vom Anwendungsfall können somit passende Muster trainiert werden [Blo06, Dud01]. Für die spätere Erkennung werden in dem überwachten Training für jeden zu erkennenden Ereignistyp (Klasse) typische, verdichtete Messwerte (Merkmale) erzeugt. Die Sensorknoten versenden die Merkmale der jeweiligen Klasse an die beteiligten Nachbarn. Die gesammelten Merkmale einer Klasse werden zusammengefasst und gespeichert. Damit entsteht ein verteilt trainierter Merkmalsvektor aus den vereinten Merkmalen aller beteiligten Sensorknoten als Referenz für jede Ereignisklasse. In der anschließenden Erkennungsphase wird äquivalent zum Training für ein erfasstes Ereignis ein Merkmalsvektor erstellt und mittels des „Prototype Modeler“ [Kal01] über den Euklidischen Abstand zu den Referenzvektoren klassifiziert.

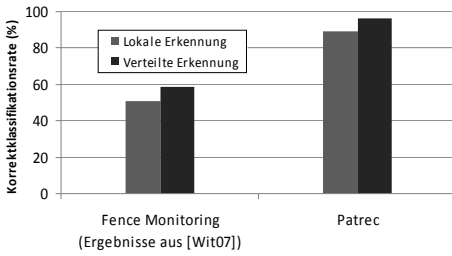


Abbildung 2: Vergleich zwischen lokaler und verteilter Ereigniserkennung

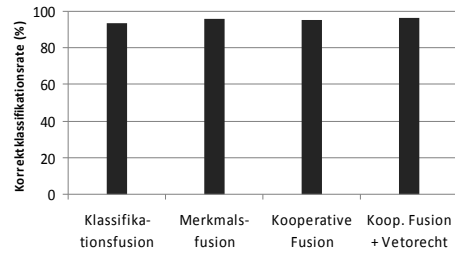


Abbildung 3: Vergleich der Fusionsmethoden

Die Datenfusion von komprimierten Daten wird auf Basis des Omnibusmodells nach [Rus07] umgesetzt. In Patrec können sowohl Merkmalsfusion als auch Klassifikationsfusion zum Einsatz kommen. Die Klassifikationsfusion fusioniert die lokal ermittelten Klassen der Sensorknoten, indem ein Mehrheitsentscheid über alle Einzelklassifikationen berechnet wird, und entspricht damit einem vereinfachten gewichteten Evaluationsansatz nach [Liu06]. Die Merkmalsfusion vergrößert den Merkmalsvektor um Merkmale der benachbarten Sensorknoten in der Trainings- und der Erkennungsphase. Die Klassifikation erfolgt nach dem gleichen Prinzip der lokalen Erkennung. Aus der Kombination der Klassifikationsfusion mit der Merkmalsfusion werden in Patrec zwei weitere Methoden abgeleitet. Die kooperative Fusion ist der Klassifikationsfusion ähnlich, greift jedoch auf die Merkmalsfusion zurück, falls die Klassifikationsdaten keine Mehrheit ergeben. Die kooperative Fusion mit Vetorecht fordert bereits zusätzlich Merkmale an, wenn nur ein Sensorknoten eine abweichende Klassifikation vorgenommen hat.

### 3 Ergebnisse

Für den Vergleich der lokalen und verteilten Erkennungsqualität wird die Korrektklassifikationsrate (KKR) gewählt, die als Anteil der korrekt klassifizierten Ereignisse im Verhältnis zu allen zu klassifizierenden Ereignissen definiert ist. Zur Evaluation wurden in 160 Versuchen jeweils mit drei Sensorknoten gleichzeitig eine von vier verschiedenen zweidimensionalen Figuren nach [Dzi07] erzeugt. Dies entspricht 480 lokalen Versuchen mit den einzelnen Sensorknoten. Die Vergleichsarbeit [Wit07], welche ebenfalls den MSB-430 einsetzt, hat mit 90 Versuchen das Ereignis der Zaunüberwindung untersucht. Im Verhältnis zur Vergleichsarbeit ist eine Steigerung der KKR sowohl in der lokalen als auch in der verteilten Erkennung von ca. 37 Prozentpunkten zu verzeichnen, siehe Abb. 2. Dieser Qualitätszuwachs lässt sich auf die Einführung des „Prototype Modeler“ zurückführen, der hier in ein klassisches Mustererkennungsmodell eingebettet ist. Die KKR der lokalen Versuche von 89% bildet die Grundlage für die verteilte Erkennung. Die Methoden der verteilten Erkennung liefern KKR zwischen 93,8% und 96,3%, siehe Abb. 3. Die Klassifikationsfusion erreicht hierbei die niedrigste KKR, da diese Fusion auf den komprimiertesten Sensordaten basiert. Die Merkmalsfusion steigert die KKR geringfügig, während die Kooperationsmethoden keinen nennenswerten Vorteil haben. Die KKR lässt sich durch die Nutzung der Merkmalsfusion optimieren.

## Literaturverzeichnis

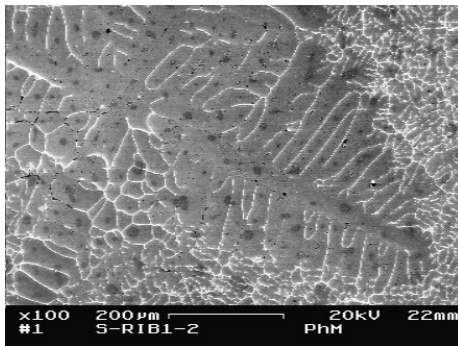
- [Blo06] Blotny, A.: Deployment, Calibration and Data Validation Techniques for Wireless Sensor Networks. Seminararbeit. Freie Universität Berlin, 2006.
- [Bok06] Bokareva, T., Hu, W., Kanhere, S., Ristic, B., Gordon, N., Bessel, T., Rutten, M., Jha, S.: Wireless Sensor Networks for Battlefield Surveillance. In: Proceedings of the Land Warfare Conference (LWC '06), Brisbane, Australia, Oktober 2006.
- [Bro03] Brooks, R. R., Ramanathan, P., Sayeed, A. M.: Distributed Target Classification and Tracking in Sensor Networks. In: Proceedings of the IEEE, 91(8):1163-1171, August 2003.
- [Dud01] Duda, R. O., Hart, P. E., Stork, D. G.: Pattern Classification. Wiley-Interscience, Chichester, England, Oktober 2001.
- [Dzi07] Dziengel, N.: Verteilte Ereigniserkennung in Sensornetzen, Diplomarbeit. Freie Universität Berlin, Oktober 2007.
- [Ian02] Akyildiz, I. F., Su, S., Sankarasubramaniam, Y., Cayirci, E.: A Survey on Sensor Networks. In: IEEE Communications Magazine, 40(8):102-116, August 2002.
- [Kal01] Kalton, A., Langley, P., Wagstaff, K., Yoo, J.: Generalized Clustering, Supervised Learning, and Data Assignment. In: Proceedings of the Seventh International Conference on Knowledge Discovery and Data Mining, New York, USA, 2001.
- [Liu06] Liu Chung-Ting, Huo Hong, Fang Tao, Shen Xiao: Classification Fusion in Wireless Sensor Networks. In: Acta Automatica Sinica, Shanghai, 2006.
- [Lli97] Llinas, J., Hall, D. L.: An Introduction to Multi-sensor Data Fusion. In: Proceedings of the IEEE, 85(1):6-23, Januar 1997.
- [Nie03] Niemann, H.: Klassifikation von Mustern. Springer-Verlag, Berlin, 1983.
- [Rus07] Ruser, H., Puente León, F. [Hrsg.]: Informationsfusion - Eine Übersicht. Oldenburg Verlag, 2007.
- [Sch05] Schiller, J., Liers, A., Ritter H.: ScatterWeb: A Wireless Sensornet Platform for Research and Teaching. In: Computer Communications, 28:1545-1551, April 2005.
- [Wäl07] Wälchli, M., Skoczylas, P., Meer, M., Braun, T.: Distributed Event Localization and Tracking with Wireless Sensors. In: Fifth International Conference on Wired/Wireless Internet Communications (WWIC '07), Bern, Schweiz, 2007.
- [Wer06] Werner-Allen, G., Lorincz, K., Johnson, J., Lees, J., Welsh, M.: Fidelity and Yield in a Volcano Monitoring Sensor Network. In: Proceedings of the Seventh USENIX Symposium on Operating System Design and Implementation, Seattle, USA, November 2006.
- [Wit07] Wittenburg G., Terfloth K., Villafuerte, F. L., Naumowicz, T., Ritter, H., Schiller, J.: Fence Monitoring - Experimental Evaluation of a Use Case for Wireless Sensor Networks. In: Proceedings of the Fourth European Conference on Wireless Sensor Networks (EWSN '07), Delft, Niederlande, 2007.

# A Profitability Heuristic that Reduces the Parameter Dependence of Dynamic Load Balancing

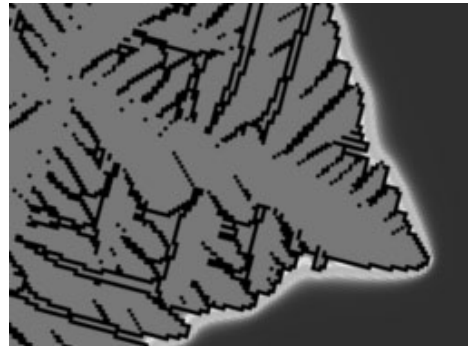
Jakob Erdmann, Andreas Schäfer, Christian Kauhaus, Dietmar Fey

## Summary

During the development of a metallurgic simulation using a continuous automaton model, we observed that individual cells cause considerably varying computational loads per simulation step depending on their states. The simulation is a parallel MPI code intended to run on cluster computers, done in collaboration with the Chair of Metallic Materials at our university. It models dendritic solidification during the cooling process of molten aluminium alloys to optimise material characteristics for use in technical components. Figure 1 shows a microscopic image of an Al-Cu dendrite and the corresponding simulation result.



(a) Microscope photo.



(b) Simulation snap-shot.

Figure 1: Microscopic dendrite photo versus simulation result. The colours denote the temperature gradient in the molten alloy, while the solidified areas are outlined in black.

The dynamics of our simulation lead to oscillating load behaviour on some cluster nodes. We employ a DLB algorithm to distribute loads evenly. Unfortunately, DLB can be inefficient when faced with oscillating loads: Too infrequent balancing may underachieve short-term savings, while too frequent balancing may cause harm by needless migrations. Thus, the performance of most DLB algorithms depends on parameter choices which need to be made in a time-consuming experimental process, which is impractical.

In response to this problem we have developed an event based DLB algorithm that learns some of its parameters at runtime and uses a prediction heuristic to reduce the influence of parameters which cannot easily be deduced.

## Algorithm Structure

Our DLB algorithm solves the problem of finding a reasonable balancing frequency by dynamically building models of load distribution, migration costs and load fluctuations. These models are used to weigh the costs of a load balancing operation against the expected long term performance gains. At the heart of this decision process lies our heuristic for predicting the long term performance impact of work migration. We call our heuristic *Predicted Long-Term Savings (PLS)*.

During the course of our simulation we periodically check whether the potential time savings from migrating work outweigh the associated costs  $C$ . The assessment of potential savings is always based on an assumption of continuity: Future loads are similar to loads that were measured since the previous balance check. Combined with our load distribution model this assumption allows us to select an optimal decomposition  $d_{\text{new}} = d_{\text{opt}}(i)$  for  $i$ -th check. Our distribution model allows us to predict how much time will be saved in the next iteration by using  $d_{\text{new}}$  instead of the current decomposition  $d_{\text{cur}}$ . It commonly occurs that those immediate savings are lower than  $C$  and do not justify balancing. However, if the trend of the changing load values would continue, migrating work would pay off in the long run. On the other hand, load values might reverse in the near future which would make balancing counterproductive.

Our prediction heuristic assumes that any trend of increasing savings will extend into the future as far as it has extended into the past.

$$\text{pred}(i_{\text{cur}}) = \sum_{i=e}^{i_{\text{cur}}} \text{savings}(i, d_{\text{cur}}, d_{\text{new}}) \quad (1)$$

with  $e < i_{\text{cur}}$  being the minimum index so that the following conditions hold:

1. all addends are positive and
2. the addends form a monotonously increasing sequence.

where  $\text{savings}(i, d_{\text{cur}}, d_{\text{new}})$  is the expected difference in running times for one simulation period when using domain decomposition  $d_{\text{new}}$  instead of  $d_{\text{cur}}$  with load values as measured at check  $i$ .

## Experimental Results

Figure 2 shows total running times for an exemplary instance of our dendrite simulation, computed on 9 nodes. As can be seen from the diagram, running times vary depending

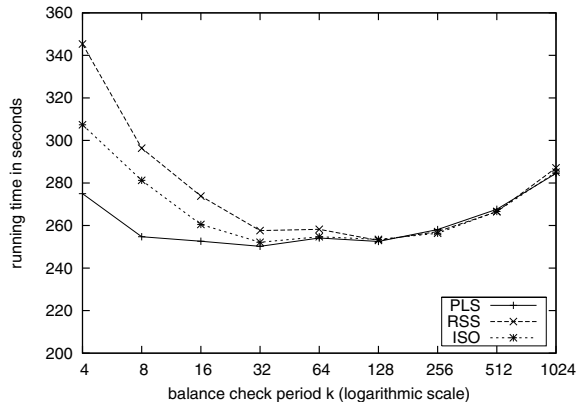


Figure 2: Running times of *PLS* compared to two reference algorithms for different check periods.

on parameter choice. Our heuristic *PLS* exhibits superior performance over a wide range of parameter choices when compared with other candidate DLB-algorithms. This greatly reduces the parameter tuning efforts and thus speeds up application deployment. The underlying concept is general enough to allow adaption for a broad range of DLB dependent problems.

## References

- [Ber87] S.H. Berger, M.J. Bokhari. A Partitioning Strategy for Nonuniform Problems on Multiprocessors. *IEEE Transactions on Computers*, C-36(5):570–580, 1987.
- [KLHP99] Ka-Yeung Kwok, Fu-Man Lam, Mounir Hamdi, and Yi Pan. *Application-Specific Load Balancing on Heterogeneous Systems*, volume 2. Prentice Hall, New Jersey, 1999.
- [MSM04] Timothy G. Mattson, Beverly A. Sanders, and Berna L. Massingil. *Patterns for Parallel Programming*. Addison Wesley Professional, 2004.
- [PRR99] A. Plastino, C. Ribeiro, and N. Rodriguez. Load balancing algorithms for SPMD applications, 1999.
- [SLKS99] Arjen Schoneveld, Martin Lees, Erwan Karyadi, and Peter M. A. Sloot. Dynamic load balancing in parallel finite element simulations. In P. Sloot, M. Bubak, A. Hoekstra, and B. Hertzberger, editors, *High-Performance Computing and Networking. Proc 7th HPNC Conference, Amsterdam*, volume 1593 of *Lecture Notes in Computer Science*, pages 409–419, Berlin/Heidelberg, 1999. Springer.
- [SW03] Stefan Schamberger and Jens-Michael Wierum. Graph Partitioning in Scientific Simulations: Multilevel Schemes versus Space-Filling Curves, 2003.
- [WL93] A.P. Willebeek-LeMair, M.H.; Reeves. Strategies for dynamic load balancing on highly parallel computers, 1993.





# Introducing Measurable Path Selection Metrics to Anonymizing Overlay Networks

Johannes Renner

RWTH Aachen University - Chair of Computer Science IV  
renner@i4.informatik.rwth-aachen.de

**Abstract:** Providing anonymity for users of the Internet is a difficult task. Currently, there are only a few systems that are of practical relevance for the provision of low-latency anonymity. One of the most important to mention is Tor which is based on onion routing. Practical client usage of the Tor network though, often leads to delays that are not tolerated by the average end-user. This, in return, discourages many of them from the use of such systems. The latter indirectly lowers the protection for remaining users due to a smaller anonymity set. This work proposes new methods of path selection for the Tor network that are based on measurable performance metrics, instead of self-advertised values. These can be used to improve the average achieved performance, as well as the security of such systems by preventing certain attacks.

## 1 Introduction

With the growth of the digitized world, privacy issues get more and more importance. Anonymous communication deals with hiding relationships between communicating parties and is a fundamental building block for e.g. privacy-friendly web browsing or identity management systems. Without such protection, attackers are able to deduce information about the network addresses of senders and recipients involved in communications, which is often enough to uniquely identify persons.

Many approaches have already been proposed in order to provide privacy-protection on the network layer, though only some of them have been implemented in praxis (e.g. [DMS04, BFK00]). This work is focused on the most popular and widespread anonymizing system there is today, which is Tor [DMS04], but the proposals can also serve as input for improved designs of future anonymity systems. The Tor network is a circuit switched, low-latency anonymizing overlay network. It is an implementation of the so-called *onion routing* technology that is based on routing TCP *streams* through randomly chosen paths in a network of *onion routers*, while using layered encryption and decryption of the content.

Currently, the publicly accessible Tor network consists of about two thousand servers that are run by volunteers, whereas the number of users is estimated to be hundreds of thousands<sup>1</sup>. To learn about the current *network status*, Tor clients (*onion proxies*) download *descriptors* of all currently active routers from a *directory* mirror. A router's descriptor contains all of the necessary information, like its IP address and the used port, as well as

---

<sup>1</sup>in January 2008

self-advertised information regarding its capacities. After choosing a *path* of, per default, 3 routers (*entry*, *middle* and *exit node*), a *circuit* through the network is created using layered encryption in a hop-by-hop manner. Once the circuit is established, it can be used as an anonymizing tunnel for arbitrary TCP connections (for a more detailed description of the Tor design see [DMS04]).

## 2 Problem Description

The Tor anonymizing network is very dynamic: Everybody can join it by running a server and thus offer available resources for other users. Client usage of Tor though, often leads to significant additional delays caused by the network layers. These delays are mostly due to the inhomogeneity of the network that is caused by its volunteer-based structure, and are often perceived as unnecessary by the users, who then most likely choose to continue surfing the Internet without using Tor. This leads to an overall smaller user base (*anonymity set*), since many users are not willing to sacrifice usability in order to achieve anonymity.

Further, currently used methods of path selection in Tor choose nodes weighted by bandwidth values that are self-advertised by the routers and distributed using the centralized directory service. This fact, however, offers multiple options for attacks, e.g. it is possible to advertise very high bandwidth values to gather as many client paths as possible [BMG<sup>+</sup>07, Dou02]. This work therefore proposes new methods of path selection that can provide an improved performance to the end-users than currently used methods. Further, the proposals will be based on measurable metrics instead of self-advertised values, for being able to prevent certain attacks. The Tor network might then attract more users and hence gain an improved security and anonymity for the single participants [Köp06, DM06].

## 3 Proposed Improvements

**Actively Measuring RTTs:** The Tor protocol does currently not provide any mechanisms to measure round-trip times (RTTs) of anonymous communication channels. For sticking with a low latency, it would be helpful to be able to actively measure latencies of Tor circuits, as well as of virtual links between single routers. This proposal is based on an existing prototypical implementation of a method for measuring RTTs in the Tor network that is described in [PPR08]. By making use of Tor's *leaky-pipe* circuit topology, it is possible to extend this technique for measuring RTTs of partial circuits by specifically addressing every single hop of a circuit as the target node once. RTTs of partial circuits can then be used to calculate link-wise RTTs between the single Tor routers in a path containing nodes 0 to  $n$  using the following equation:

$$RTT_{n-1,n} = RTT_{0,n} - RTT_{0,n-1}$$

For making use of the measured results, it is proposed to model the explored subnet of the Tor network in a graph structure. A so-called *network model* can contain nodes, links between these nodes and arbitrary *node-* or *link-wise* performance metrics. All of the metrics supplied to the model, as well as additional self-advertised information taken from the descriptors, can be used to calculate combined *ranking indices* for either nodes, links or even complete paths. A suitable path that is found in the model is called a *path proposal*. Path selection can then be done probabilistically from the set of all path proposals regarding the ranking indices. This keeps the random aspect in path selection for preserving anonymity, while further being able to control the specific influence of a certain metric on the selection by introducing additional weights. Besides that, the possibility to measure RTTs on completely established circuits enables us to optimize load-balancing in the Tor network on the circuit layer. This can be done by ensuring in the clients that user streams are always attached to circuits that are currently providing low latencies, using initial or frequent measurements of RTTs on a number of maintained circuits.

**Estimating Available Bandwidth:** Depending on the specific application, a constantly high throughput rate might be more important to a user than latency. Therefore, throughput also needs to be considered as a possible metric for path selection. Instead of weighting the routers by their self-advertised, node-wise bandwidth values from the directory, it is proposed to measure throughput link-wise, for being able to more precisely predict the capacities of specific nodes and links. Measuring throughput actively by transferring streams over certain nodes for probing their capacities is definitely too much overhead that would have negative impacts on the overall network performance. Therefore, the proposal is to measure throughput passively from within the routers, counting amounts of user-traffic on the single TLS links a node maintains to other nodes on the network, as well as the total amount of traffic it is relaying within a specific time interval. Based on these measurements, routers can give estimations about the currently available capacities on every link to another router. These estimations can regularly be reported to a centralized directory service that averages over all of the information from different sources. This makes it possible to diminish the influence of possible outliers with the aim of detecting, and thus weakening collusions or sybil attacks.

## 4 Conclusions

The methods that were outlined in Section 3 can be used to improve the performance that is perceived by end-users of Tor regarding the average setup duration, latency and throughput of anonymous channels. Using the existing implementation it is further possible to document any possible loss or gain of anonymity that is induced when varying the method of path selection. For a practical adoption of the proposed methods though, several important design decisions need to be made first. While the use of any link-wise path selection metrics, for example, would allow to more precisely predict the performance of certain paths, the amount of data that needs to be measured, stored in a model and somehow transferred to the clients is quadratically increasing with the number of routers in the network. This will most likely lead to scalability issues, considering the current size of the Tor network.

## References

- [BFK00] Oliver Berthold, Hannes Federrath, and Stefan Köpsell. Web MIXes: A system for anonymous and unobservable Internet access. In H. Federrath, editor, *Proceedings of Designing Privacy Enhancing Technologies: Workshop on Design Issues in Anonymity and Unobservability*, pages 115–129. Springer-Verlag, LNCS 2009, July 2000.
- [BMG<sup>+</sup>07] K. Bauer, D. McCoy, D. Grunwald, T. Kohno, and D. Sicker. Low-Resource Routing Attacks Against Anonymous Systems. Technical Report CU-CS-1025-07, University of Colorado at Boulder, February 2007.
- [DM06] Roger Dingledine and Nick Mathewson. Anonymity Loves Company: Usability and the Network Effect. In *Proceedings of the Fifth Workshop on the Economics of Information Security (WEIS 2006)*, Cambridge, UK, June 2006.
- [DMS04] Roger Dingledine, Nick Mathewson, and Paul Syverson. Tor: The Second-Generation Onion Router. In *Proceedings of the 13th USENIX Security Symposium*, 2004.
- [Dou02] John Douceur. The Sybil Attack. In *Proceedings of the 1st International Peer To Peer Systems Workshop (IPTPS 2002)*, March 2002.
- [Köp06] Stefan Köpsell. Low Latency Anonymous Communication - How Long Are Users Willing to Wait? In Günter Müller, editor, *ETRICS*, volume 3995 of *Lecture Notes in Computer Science*, pages 221–237. Springer, 2006.
- [PPR08] Andriy Panchenko, Lexi Pimenidis, and Johannes Renner. Performance Analysis of Anonymous Communication Channels Provided by Tor. In *Proceedings of the Third International Conference on Availability, Reliability and Security (ARES 2008)*, Barcelona, Spain, March 2008.

# Kooperation heterogener Internet-Anwendungen

Sebastian Wenzky

swenzky@krypthonas.de

Fachbereich Informatik, University of Applied Science Schmalkalden

Art der Arbeit: Projektarbeit

Betreuer: Prof. Dr. Harm Knolle

**Abstract:** Die Integration autonomer Internet-Anwendungen in eine bestehende Architektur stellt eine Herausforderung dar. Nicht nur die Endanwender erhoffen sich hierdurch qualitative Neuerungen wie z.B. die Zusammenführung bisher dezentral angebotener Funktionalitäten. Es muss auch die hierfür notwendige Technik überschaubar und das Gesamtsystem wartbar bleiben. Je nach Form der Kooperation und die damit verbundenen Anwendungszwecke kann die Integration mehr oder weniger Effizient auf unterschiedlichen Systemebenen erfolgen. Wie bei vielen verteilten Systemen steht auch hier die Effizienz im Spannungsfeld zwischen Autonomie und Kooperationsbereitschaft der beteiligten Systeme.

## 1 Einführung

Bei der Kooperation müssen mehrere Teilaspekte berücksichtigt werden [CDK02]. Zum einen soll der Endanwender durch die Zusammenfassung verschiedenartiger Inhalte in einem System einen klaren inhaltlichen Vorteil (den erhöhten Informationsgehalt) erlangen. Zum anderen muss eine einfach sowie leicht wartbare technische Integrationsmöglichkeit gefunden werden. Dennoch kann es je nach Kooperationsform erforderlich sein, das beide Systeme weiterhin unabhängig voneinander betrieben werden können.

Soll eine bestehende Internet-Anwendung (bspw. ein Portal) um eine bestimmte Funktionalität einer anderen Internet-Anwendung erweitert werden, stehen Ingenieure vor vielen Fragestellungen. Grundsätzliche Punkte wie Wartbarkeit, laufende Kosten, Sicherheit, Aktualität, Performance, Erweiterbarkeit sind nur einige von vielen Faktoren, die im Vorfeld sorgfältig evaluiert werden sollten. Ist die inhaltliche Entscheidung hinsichtlich der Erweiterung gefallen, bedarf es u.a. das abwägen folgender Punkte.

- Sind Struktur und Verhalten der ausgewählten Routine ausreichend bekannt, um sie effizient in die bestehende Infrastruktur zu integrieren?
- Welche Tiefe erreicht die Integration von neuen Routinen im bestehenden System-Umfeld?
- Kann der unabhängige Betrieb der beteiligten Portale weiterhin gewährleistet werden (ist es überhaupt erforderlich)?

In Folgenden soll der Prozess der Benutzer-Anmeldung als Beispiel herangezogen werden (siehe Abbildung 1). Das Ziel hierbei ist, dass der Benutzer sich nur einmal anmelden muss

um die Funktionalitäten beider Portale nutzen zu können. Auf welche Art und Weise eine Methode letztendlich implementiert werden kann, damit die Integration für die Endanwender Transparent bleibt und auf welcher Systemebene sie integriert wird, ist vielfach von der zugrunde liegenden Kooperation und den sich ergebenden Freiheitsgraden abhängig.

## **2 Arten der Kooperation**

Zunächst muss eine geeignete Form der Zusammenarbeit gefunden werden. Dabei spielen vor allem auch psychologische Aspekte eine große Rolle. So können Bedenken bestehen, dass Alleinstellungsmerkmale im Rahmen einer Kooperation verloren gehen.

### **2.1 Starke Kooperation**

Eine Verschmelzung, sowohl im technischen als auch im semantischen Sinn, kann zum Ziel haben, dass ein System ein anderes in sich aufnimmt und somit aufgewertet wird (siehe Abbildung 2). Diese Möglichkeit bietet sich vor allem für Systeme an, die sich ähnlichen Thematiken widmen. So könnte beispielsweise eine vollständige Fusionierung zweier Job-Portale die Problematik der gemeinsamen Benutzer-Anmeldung obsolet werden lassen, da diese nur noch an dem vereinigten Portal erfolgen muss. Eine solche Verschmelzung hat häufig einen hohen Implementierungsaufwand zur Folge.

### **2.2 Beständige Kooperation**

Im Gegensatz zur Verschmelzung behalten die Kooperationspartner die Technologien sowie ihre Datenbasen bei. Eine einfache Kooperationsmöglichkeit besteht darin, dass die Portalbetreiber Schnittstellen zur Verfügung stellen, die einen Zugriff auf die jeweiligen Funktionalitäten der beteiligten Portale ermöglichen. Dieses Konzept besitzt den Vorteil, dass der eigentliche Implementationsaufwand bei der Durchführung der Integration, trotz mächtiger Kooperationsmöglichkeiten, vergleichsweise gering ausfällt. Jedoch muss die Verteilung der Verwaltung auf mehreren Systemen kritisch betrachtet werden [SCF97]:

- Unter Umständen müssen Konzepte zur Transaktionsverarbeitung in verteilten Systemen angewendet werden.
- Ist eine replizierte Verwaltung bestimmter Daten in den beteiligten Systemen erforderlich, so muss die Konsistenz dieser Daten gewahrt bleiben.

## 2.3 Interessensgemeinschaft

Liegt die Kooperationsbereitschaft nur in einem stark beschränkten Maße vor, existieren nur wenige Möglichkeiten Zugriff auf den Kontext des Kooperationspartners zu erhalten. In vielen Fällen besteht diese Art von Zusammenarbeit im bloßen Bereitstellen von Informationen. Eine gemeinsame Haltung der Daten ist vielfach auch aus Gründen des Datenschutzes, nicht gewünscht. Eine gemeinsame Benutzer-Anmeldung gestaltet sich hinsichtlich dieser Einschränkungen als hinderlich und ist deswegen für diese Art der Kooperation ungeeignet.

## 3 Integrationsformen

Die Umsetzung einer ausgewählten Kooperationsart lässt sich auf verschiedenen Wegen und in verschiedenen Ebenen realisieren. In diesem Abschnitt werden Möglichkeiten der Systemebenen Server, Client und Netz bzw. Middleware beleuchtet und bewertet.

### 3.1 Server-orientierte Methode

Steigen die Ansprüche an die Art der Integration, sind in der Regel technische Einschnitte in die jeweiligen Systeme notwendig. Bei einer Fusionierung mehrerer Portale muss eine sehr starke Kooperationsbereitschaft vorausgesetzt werden, die eine explizite (Re-) Implementierung von Funktionalitäten des einen in das andere Portal ermöglicht (siehe Abschnitt 2.1). Das Ergebnis verkörpert vielfach die Implementierung eines (völlig) neuen Anwendungs-Servers (siehe Abbildung 2), das die Funktionalitäten der ursprünglich getrennten Portale sehr effizient realisiert.

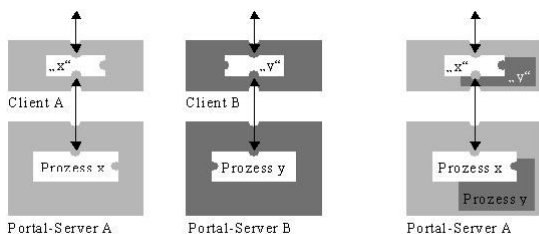


Abb. 1: Autonome Portale

Abb. 2: Fusion

Bei einer beständigen Kooperation kann die Datenhaltung der unterschiedlichen Portale über die Datenkommunikation erfolgen. Über bereitgestellte Schnittstellen in Form von Protokollen (bspw. SOAP<sup>1</sup>) lassen sich zahlreiche Funktionalitäten auf Server-Ebene sehr einfach austauschen (siehe Abbildung 3).

<sup>1</sup>Simple Object Access Protocol [IMS07]



### 3.2 Client-seitige Methode

Die Kommunikation über Schnittstellen kann auch client-seitig erfolgen. Eine relativ einfache Möglichkeit ist der Einsatz von „Spionen“ innerhalb einer Web-Seite (siehe Abbildung 4). Hierbei werden mit Hilfe spezieller Software-Komponenten sukzessive zusätzliche Informationen vom Partnerportal nachhaltig geladen und auf der darzustellenden Seite platziert. Ein größerer Einschnitt in die Programmlogik ist hierfür nicht notwendig. Solche Spione lassen sich beispielsweise mit JavaScript in der Web-Seite realisieren.

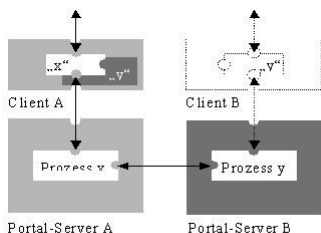


Abb. 3: Server-seitige Kooperation

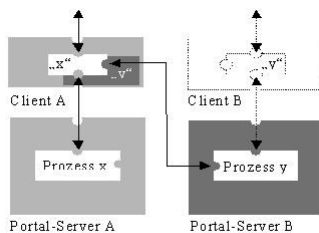


Abb. 4: Client-seitige Kooperation

### 3.3 Netzbasierende Methode

Um die technische Vermaschung innerhalb einer Systemarchitektur möglichst gering zu halten<sup>2</sup>, bietet sich der Einsatz netzbasierender Middleware an. Dabei wirken alle Zugriffe nicht direkt auf das dafür zuständige Portal, stattdessen auf einen zentralen Punkt im Netz. Dieser ist dafür zuständig, die Ein- und Ausgangsdaten zu überwachen. Erklärt sich ein Kooperationspartner hierzu bereit, können die Router-Einstellungen des Ziels so konfiguriert werden, dass alle Anfragen zunächst an eine Art Vermittler gesendet werden. Dieser interpretiert die Anfrage und schickt diese an den eigentlichen Empfänger weiter (siehe Abbildung 5).

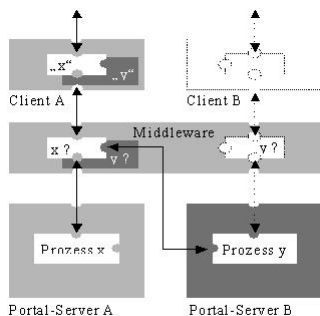


Abb. 5: Middleware

<sup>2</sup>Stichwort: Cross Cutting Concerns

## 4 Ausblick

Vor dem Zusammenspiel mehrerer Portale bzw. der Akquirierung neuerer Komponenten, sollte die Art der Kooperation und die damit verbundenen Integrationsmöglichkeiten ermittelt werden. Geschieht dies nicht in ausreichenden Maße, kann es schnell zu Problemen bei der Umsetzung und der anschließenden Wartung aber auch zu Laufzeitproblemen kommen. Um die vielfältigen Auswirkungen der unterschiedlichen Integrationsmöglichkeiten auf Kosten, Performance und Nutzen besser einschätzen zu können, wurde eine Testumgebung geschaffen [WFI08]. Diese Testumgebung verknüpft das Job-Portal „FuThuer“<sup>3</sup> mit einem Forumssystem, welches eine direkte Interaktion zwischen Jobsucher und Jobanbieter ermöglicht.

## Literatur

- [CDK02] George Coulouris, Jean Dollimore, Tim Kinberg: „Verteilte Systeme. Konzepte und Design, Pearson Studium“, München 2002
- [IMS07] Ingo Melzer: „Service-Orientierte Architekturen mit Web Services“, Springer, 2007
- [SCF97] Stefan Conrad: „Föderierte Datenbanksysteme. Konzepte der Datenintegration“, Springer, 1997
- [WFI08] Toni Wolke: „Fusionierung Heterogener Internetanwendungen“, Diplomarbeit Fachhochschule Schmalkalden, 2008

---

<sup>3</sup>[www.FuThuer.de](http://www.FuThuer.de) - Eine Initiative zur mittel- bis langfristigen Sicherung des Fachkräftebedarfs in Thüringen



# Ein einfaches Anonymisierungsverfahren basierend auf offenen Standards

Benedikt Westermann  
RWTH-Aachen - Lehrstuhl für Informatik IV  
westermann@i4.informatik.rwth-aachen.de

**Abstract:** Alle praktisch existierenden Verfahren zur Anonymisierung im Internet sind wegen ihrer hohen Komplexität nur aufwendig zu realisieren. Weitere Probleme treten aufgrund der Verwendung proprietärer Protokolle auf, da diese als Folge von fehlender Standardisierung öfters geändert werden. All dies erschwert eine parallele Entwicklung alternativer Clients, das Verständnis der Anonymisierung und die mathematische Analyse des Anonymisierungsverfahrens. Das im folgenden beschriebene Verfahren bietet eine Anonymisierung, die auf standardisierten Protokollen basiert und durch eine geringe Komplexität und eine gute Verständlichkeit einfach und effizient ist.

## 1 Einführung in die Thematik

Die Anonymität ist in [PH00] wie folgt definiert:

*„Anonymität ist der Zustand, in dem ein Subjekt nicht aus einer Menge von Subjekten, der Anonymitätsmenge, identifiziert werden kann.“*

Die Anonymisierung ist der Prozess, der die Anonymität für ein Subjekt schafft. Dies geschieht, indem die Zuordnung zwischen einer Aktion und dem Subjekt aufgehoben wird. Folglich kann nicht mehr unterschieden werden, welches der möglichen Subjekte die Aktion ausgeführt hat.

In unserer Gesellschaft nimmt die Anonymisierung eine bedeutende Rolle ein. Durch die zunehmende Digitalisierung werden Abläufe des Alltags vermehrt über Computer und das Internet abgewickelt, wodurch entsprechende Anonymisierungsverfahren auch im Internet benötigt werden. Ziel dieser Verfahren ist die Anonymisierung auf Verbindungsebene, also das Verbergen der IP-Adressen.

## 2 Probleme bekannter Anonymisierungsverfahren

Es existieren bisher nur zwei praktisch eingesetzte Verfahren: *Tor* [DMS04] und *JonDonym*<sup>1</sup> [BFK00]. Die anderen in der Literatur vorgestellten Verfahren sind indes theoretischer Natur, zwei Beispiele sind in [RR98, FM02] zu finden.

---

<sup>1</sup>Auch unter dem Namen AN.ON (Anonymität.Online) bekannt

*Tor* und *JonDonym*, die beiden praktisch eingesetzten Verfahren, sind aufwendig zu implementieren, sodass es meistens nur eine einzige Implementierung gibt. Die bei diesem Verfahren eingesetzten Protokolle sind einerseits recht komplex, wodurch eine formale Analyse schwierig ist. Andererseits gibt es ständig Änderungen an den Protokollen. Dies behindert zusammen mit der hohen Komplexität die Entwicklung alternativer Clients sowie das Verständnis des Anonymisierungsprozesses.

Durch diese Problematik motiviert, wird im Folgenden ein Verfahren vorgestellt, dass mit Hilfe des *HTTP*-Protokolls [FGM<sup>+</sup>99] eine Anonymisierung bietet, die ähnliche Eigenschaften aufweist wie *Tor* und *JonDonym*. Der Name des hier vorgestellten Verfahrens ist *Shalon* und steht für: *Scalable HTTP-based Anonymisation Lightweight Overlay Network*.

### 3 Aufbau einer anonymen Verbindung

Die grundlegende Idee des Verfahrens ist, eine TCP-Verbindung über eine Kette mehrerer HTTP-Proxyserver aufzubauen. Diese Kette wird im Folgenden als *Proxykette* bezeichnet. Ziel der Proxykette ist es, die Verknüpfung zu dem Initiator, also dem Urheber einer Verbindung, und dem Empfänger zu verschleiern. Jedoch reicht eine einfache Verkettung der Proxyserver allein nicht aus, da der Betreiber des ersten Proxyservers ohne eine Verschlüsselung in der Lage ist, den Initiator einer Aktion zuzuordnen.

Eine TCP-Verbindung wird in *Shalon* wie folgt anonymisiert: Der Initiator baut eine TCP-Verbindung zu dem Proxyserver  $P_1$  auf. Anschließend wird die Verbindung mit Hilfe von TLS [DA99] verschlüsselt. Im nächsten Schritt wird über die verschlüsselte Verbindung der Connect-Befehl des HTTP-1.1-Protokolls zusammen mit der Adresse des Proxyservers  $P_2$  geschickt. Dadurch wird die Verbindung von  $P_1$  zu  $P_2$  verlängert. Alle Daten, die nun in die geöffnete Verbindung geschrieben werden, werden von  $P_1$  an  $P_2$  weitergeleitet. Als nächstes wird die Verbindung zu  $P_2$  verschlüsselt, in dem der Initiator über die aufgebaute Verbindung einen TLS-Handshake initiiert. Eine weitere Verlängerung der Verbindung kann erneut, wie oben beschrieben, über den Connect-Befehl erreicht werden. Eine auf diese Weise aufgebaute Verbindung ist schematisch in Abbildung 1 dargestellt.

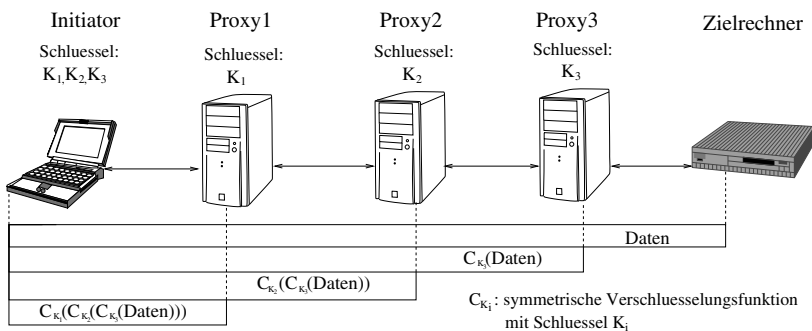


Abbildung 1: Struktur der verschlüsselten Kommunikation

Ist die Verbindung wie in Abbildung 1 aufgebaut, hat diese die gleichen Eigenschaften wie eine aufgebaute TCP-Verbindung. Durch die schichtweise Verschlüsselung ist es einem einzigen Proxybetreiber, der die Leitung abhört, nicht möglich, eine Verknüpfung zwischen dem Initiator und der Aktion zu erstellen. Für den Zielrechner ist es auch nicht möglich, den wahren Initiator anhand der IP-Adresse zu herauszufinden. Die Pfadlänge drei der anonymen Verbindung stellt einen Kompromiss zwischen Sicherheit und Geschwindigkeit dar.

## 4 Zusammenfassung

Der hier beschriebene Aufbau einer anonymen Verbindung ist im Vergleich zu Tor und Jondonym einfacher zu verstehen und zu implementieren, sodass das hier vorgestellte Verfahren durch einen erfahrenden Programmierer innerhalb von einer Woche programmiert werden kann. Durch die geringere Komplexität ist es zudem leichter Fehler zu finden. Messungen der Leistung zeigten außerdem, dass Shalon um einen Faktor sechs schneller ist als Tor. Die Verwendung des HTTP-1.1-Protokolls ermöglicht weiterhin die Wiederverwendung von bereits existierender Software, sodass ausschließlich der Client programmiert werden muss. Als Server kann zum Beispiel der Proxyserver Squid verwendet werden.

## Literatur

- [BFK00] Oliver Berthold, Hannes Federrath und Stefan Köpsell. Web MIXes: A system for anonymous and unobservable Internet access. In *Proceedings of Designing Privacy Enhancing Technologies: Workshop on Design Issues in Anonymity and Unobservability*, Jgg. 2009/2001, Seiten 115–129. Springer Berlin / Heidelberg, July 2000.
- [DA99] Tim Dierks und Christopher Allen. The TLS Protocol Version 1.0. Internet Engineering Task Force: RFC 2246, Januar 1999.
- [DMS04] Roger Dingledine, Nick Mathewson und Paul Syverson. Tor: The Second-Generation Onion Router. In *Proceedings of the 13th USENIX Security Symposium*, August 2004.
- [FGM<sup>+</sup>99] Robert Fielding, Jim Gettys, Jeff Mogul, Henrik Frystyk, Larry Masinter, Paul Leach und Tim Berners-Lee. Hypertext Transfer Protocol: HTTP/1.1. Internet Engineering Task Force: RFC 2616, June 1999.
- [FM02] Michael J. Freedman und Robert Morris. Tarzan: A Peer-to-Peer Anonymizing Network Layer. In *Proceedings of the 9th ACM Conference on Computer and Communications Security (CCS 2002)*, Washington, DC, November 2002.
- [PH00] Andreas Pfitzmann und Marit Hansen. Anonymity, Unobservability, and Pseudonymity: A Consolidated Proposal for Terminology. Draft, July 2000.
- [RR98] Michael K. Reiter und Aviel D. Rubin. Crowds: Anonymity for Web Transactions. *ACM Transactions on Information and System Security*, Seiten 66 – 92, April 1998.



# Integration von Routing und Service Discovery in drahtlosen Mesh Netzwerken

Sumedha Widyadharma  
Lehrstuhl für Informatik 4  
RWTH Aachen  
sumedha.widyadharma@rwth-aachen.de

**Abstract:** Die Vernetzung von Alltagsgeräten schreitet immer weiter voran. Schon heute besitzt fast jeder Mensch einer Industrienation technische Geräte die in der Lage sind über Funk Daten zu übertragen. Die Infrastruktur, die dieser Entwicklung in der Zukunft folgen wird sind vermaschte multi-hop Netzwerke oder Wireless Mesh Networks (WMNs). Eine weitere Technologie die in diesem Zusammenhang immer mehr an Bedeutung gewinnt ist die der Dienstsuche (Service Discovery). Meist sind Benutzer nicht an Verbindungen zu expliziten Netzadressen interessiert, sondern an der Nutzung eines Dienstes, der auf einem Gerät angeboten wird. Die Integration dieser beiden Technologien wird somit in Zukunft eine wichtige Rolle spielen. Die zur Zeit präsentesten Beispiele sind Drucker, die sich automatisch in ein lokales Netzwerk einbinden lassen und Multimedia Streaming Appliances, die automatisch Musiksammlungen auf Geräten in lokalen Netzen finden und abspielen können. Im Folgenden wird ein Lösungsansatz basierend auf reaktivem Routing und DNS basierter Dienstsuche präsentiert.

## 1 Einführung

In Zukunft wird sich das Problem ergeben, dass auch in WMNs[AWW05] Dienste zur Verfügung gestellt werden sollen. Die Liste der gesuchten Dienste soll dabei sowohl möglichst schnell zur Verfügung stehen als auch möglichst vollständig sein.

Grundsätzlich sind WMNs den klassischen Mobilen Ad-Hoc Netzwerken (MANETs) ähnlich. Man geht allerdings von einer gewissen Stabilität der zu Grunde liegenden Infrastruktur aus, so dass einige der Probleme von MANETs nicht oder nur in geringerem Maße berücksichtigt werden müssen. Dienstsuche ist in diesem Umgebungen kaum mit den klassischen Mitteln möglich, da diese meist eine zentralisierte Infrastruktur mit Dienstverzeichnissen benötigen oder Multi-/Broadcast Domänen.

## 2 Grundlagen

Das Problem lässt sich in zwei Teile gliedern:

- *Dienste im Netzwerk finden*



- *Routen zu Diensten finden*

Zentralisierte Ansätze können nicht verwendet werden, da WMNs meist nicht durch eine zentrale Instanz verwaltet werden. Dienstsuche beschreibt also hier den Prozess, Dienste ohne a-priori Wissen über die Dienste oder das Netzwerk anzubieten und zu finden. Aufgrund der Nutzbarkeit und Verbreitung wurde als Komponente zur Dienstsuche für mDNS-SD (Multicast-DNS based Service Discovery) entschieden. Dieses Verfahren basiert wiederum auf DNS-SD[CK06a], einem zentralisierten, DNS basierten Dienstverzeichnis und mDNS[CK06b], einer dezentralen DNS Variante, die auf Multicast basiert und für spezielle, lokale Netzbereiche definiert ist.

Bei Routing Algorithmen existieren grundsätzlich zwei Varianten:

- **Reaktiv:** Routen werden nur dann gesucht, wenn Daten für ein Ziel vorliegen, für das keine Einträge in den Routingtabellen vorhanden sind.
- **Proaktiv:** Die Knoten des Netzwerkes tauschen untereinander Informationen über die bekannte Topologie aus und berechnen Routen im Voraus.

Die Vorteile von proaktivem Routing (ständige Kenntnis aller Routen) werden allerdings in mobilen Netzwerken von den Nachteilen überwogen:

- *Batterielaufzeit: Häufige Funkverbindungen und ständige Berechnungen verbrauchen viel Strom*
- *Prozessorzeit: Mobile Computer haben meist nur eine beschränkte Rechenkapazität und die Algorithmen zur Berechnung der aktuellen Routen würden die dem Benutzer zur Verfügung stehende Rechenzeit signifikant einschränken.*

Dies ist gerade für mobile Geräte nicht tragbar und gerade diese erfreuen sich einer immer größeren Verbreitung. Reaktives Routing hat diese Probleme nicht, Datenverbindungen werden, abgesehen von einer initialen Routensuche, nur zum tatsächlichen Übertragen von Daten verwendet. Dies führt natürlich dazu, dass vor jeder Datenübertragung eine Verzögerung auftritt in der schnellstmöglich eine Route zum Zielknoten gesucht werden muss. Besteht die Route einmal und fließen, auch nur wenige, Daten darüber bleibt sie erhalten. Aus diesem Grunde fiel die Wahl der Routing Software auf dymoum, einer DYMO Implementation für Linux und NS2.

### 3 Technische Umsetzung

Für die Entwicklung stand eine realistische Test- und Simulationsumgebung mit 40 Hardwareknoten bereit. Dieses wird im Rahmen des UMIC(Ultra High Speed Mobile Information and Communication)[IU] Exzellenzclusters zur Entwicklung bereitgestellt. Auf den Testbettknoten läuft ein auf Ubuntu Linux basierendes GNU/Linux Betriebssystem auf x86 Hardware.

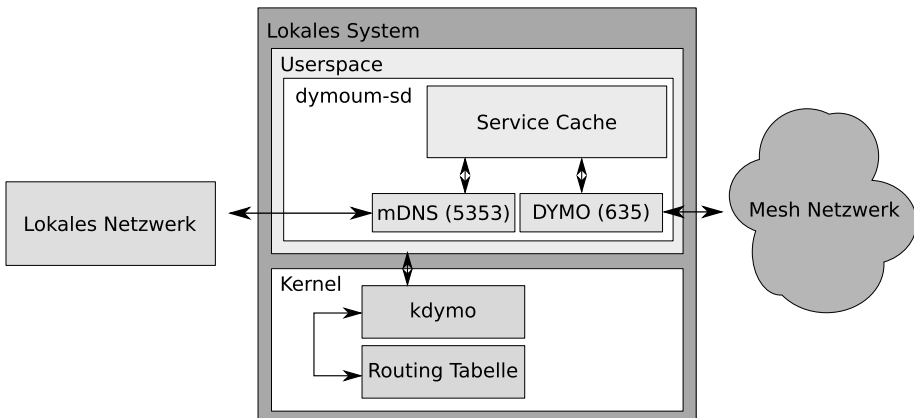


Abbildung 1: Architektur

Die Implementation eines reaktiven integrierten Routing/Dienstsuche Programms soll auch den Vergleich zur Dienstsuche in proaktiven Routing Umgebungen bieten, zum Beispiel das auf mDNS-SD (multicast DNS based service discovery) und OLSR[CJ03] basierende OLSR-mDNS[KKK08].

Als Basis für das Routing wurde die dymoum[Ros] DYMO Implementation gewählt und als Basis für die Dienstsuche die Avahi[Ava] DNS-SD Implementation, da ein Prototyp geschaffen werden sollte, welcher sich in einem realistischen Szenario verwenden lässt. DYMO wird als Nachfolgerprotokoll zu AODV betrachtet. Die Architektur des Systems ist in Abbildung 1 zu sehen.

Normalerweise läuft eine mDNS-SD Dienstsuche so ab, dass sowohl Frage als auch Antwort über Multicast geschehen. Dies unterdrückt z.B. zeitnahe Anfragen anderer Knoten nach gleichen Diensten oder redundante Antworten. Auch mDNS-SD stellt allerdings die Möglichkeit bereit, Antworten nur per Unicast zu erhalten. Dies ist grundsätzlich die Idee der hier beschriebenen Lösung.

Applikationen, die auf den Knoten oder in darüber gerouteten Netzen laufen, können Standard mDNS-SD Anfragen verwenden um Services zu suchen. Der in die Routing Applikation eingebettete mDNS Stack nimmt diese Anfrage entgegen und startet gegebenenfalls eine Dienstsuche im Mesh-Netzwerk. Dazu sendet der Knoten einen sogenannten Service Request (SR) durch das Netzwerk. Diese SR Pakete sind denen der Routensuche (RE, Routing Element) ähnlich. Sie sind um die Beschreibungen der gesuchten Dienste erweitert wurde und enthalten keinen definierten Zielknoten mehr. Empfangende Knoten, die diese Dienste zur Verfügung stellen, oder Knoten kennen die diese Dienste zur Verfügung stellen senden ihrerseits Antwortpakete zurück, die auf dem Rückweg zum anfragenden Knoten Routeninformationen zu den betreffenden Knoten auffrischen.

Der wichtigste Unterschied zur Routensuche ist das potentielle Mehraufkommen zurückfließender Daten. Um diesem Problem entgegenzuwirken werden verschiedene Mechanismen getestet:

- Generell wird davon ausgegangen, dass die Infrastruktur eines Mesh Netzwerkes relativ stabil bleibt. Service Daten werden deshalb länger vorgehalten.
- Knoten die nah am Ursprung sind und Antworten kennen warten erst auf Antworten weiter entfernter Knoten um diese dann gesammelt zu verschicken.
- Services, die Anforderungen an Verbindungen zwischen Knoten haben (geographische Nähe, erreichbare Datenrate, etc.) werden nicht weitergeleitet, wenn bekannt ist, dass die Verbindung die Anforderungen nicht erfüllen kann.

## 4 Ergebnis und Ausblick

Wie beschrieben gibt es sowohl Ansätze die versuchen Service Discovery sowohl mit proaktiven als auch mit reaktiven Routing Protokollen effektiv zu verbinden. In Umgebungen mit ausreichend Rechen- und Akkulaufzeit der Knoten sind proaktive Routing Protokolle die bessere Wahl. Die Infrastruktur der Zukunft setzt sich jedoch zum großen Teil aus mobilen Kleingeräten zusammen. Eine weiterer interessanter Ansatz sind somit sogenannte hybride Routing Protokolle die sowohl Knoten unterstützen, die proaktiv Routingtabellen zwischen sich und anderen proaktiven Knoten vorberechnen als auch reaktive Knoten die, um Energie oder Rechenzeit zu sparen, nur nach Routen oder Diensten fragen, wenn sie benötigt werden.

## Literatur

- [Ava] Avahi. Avahi mDNS implementation. <http://www.avahi.org/>.
- [AWW05] Ian F. Akyildiz, Xudong Wang und Weilin Wang. Wireless mesh networks: a survey. *Computer Networks*, 47(4):445–487, March 2005.
- [CJ03] Thomas Clausen und Philippe Jacquet. Optimized Link State Routing Protocol (OLSR). (3626), October 2003. <http://rfc.net/rfc3626.txt>.
- [CK06a] Stuart Cheshire und Marc Krochmal. DNS-Based Service Discovery. IETF Internet Draft, August 2006. work in progress.
- [CK06b] Stuart Cheshire und Marc Krochmal. Multicast DNS. IETF Internet Draft, August 2006. work in progress.
- [IU] Ultra High-Speed Mobile Information und Communication (UMIC). UMIC-Mesh Testbed. <http://www.unic-mesh.net/>.
- [KKK08] Martin Krebs, Karl-Heinz Krempels und Markus Kucay. Service Discovery in Wireless Mesh Networks. In *Proceedings IEEE Wireless Communications and Networking Conference 2008 (WCNC'08)*. IEEE Communications Society Press, March 2008.
- [Ros] Francisco J. Ros. DYMOUM, DYMO for Linux and NS2.

# An Energy-Efficient Routing Protocol for Linear Wireless Sensor Networks

Marco Zimmerling

marco.zimmerling@inf.tu-dresden.de

**Abstract:** Economical power use is essential to allow for long-lasting operation of wireless sensor networks. This applies equally to linear sensor networks as they emerge when sensors are deployed along bridges or pipelines. In this work, we present MERR (Minimum Energy Relay Routing), a novel approach to energy-efficient routing to a single control center in a linear sensor topology. Based on an optimal transmission distance, MERR finds paths that minimize total power consumption. Our analytic and simulation results show that MERR saves significant power compared to conventional approaches and has near-optimal performance.

## 1 Introduction

A sensor network can consist of thousands of nodes that are limited in power, computational capacities, and memory. Thus, the primary constraints for sensor network protocols and algorithms are *energy efficiency*, *scalability*, and *localization*. Scalability is ensured if each node exchanges information only with its neighbors [EGHK99]. In a localized routing protocol, each node decides on the next hop based only on the position of itself, of its neighbors, and possibly of the destination node.

Many routing protocols have been designed for wireless sensor networks over the past years [AKK04]. Most of them consider the most general case where sensors operate in a mesh topology. For many practical scenarios, however, a mesh topology may not be appropriate or simply not feasible. Consider, for example, structural health monitoring of bridges [KPC<sup>+</sup>07] and pipelines [SNM07], geolocation in underground mines [ND05], border surveillance, or lighting control along corridors. These applications feature, by and large, a linear sensor topology that is predetermined by physical structure, measuring point distribution, and application requirements.

In this work, we study the problem of energy-efficient data delivery from multiple source nodes to a single sink node (*base station*) in a linear sensor network. We first discuss two conventional approaches, direct transmission and MTE (minimum-transmission-energy) routing, and show how paths can be established that have minimum total power consumption. Then we present MERR [ZDR07], a novel routing protocol that approximates optimal paths by selecting suitable nodes for retransmission. We evaluate all algorithms analytically and by simulation using a stochastic model for the sensor distribution. As our results show, MERR achieves power savings of up to 80% compared to MTE routing and deviates less than 10% from the theoretical optimum in practicable sensor networks.

## 2 System Model

In this work, we use a radio model where a sensor dissipates  $P_{tx}(r, d) = r(\alpha_{tx} + \epsilon d^\gamma)$  of power when transmitting a bit-stream of rate  $r$  over a distance  $d$  and  $P_{rx}(r) = r\alpha_{rx}$  when receiving the same stream [BGC01]. Parameters  $\alpha_{tx}$  and  $\alpha_{rx}$  ( $\alpha_{tx} + \alpha_{rx} = \alpha$ ) are the energy per bit consumed in the transmitter and receiver circuits, respectively, and  $\epsilon$  accounts for the energy dissipated in the transmit amplifier. The *path loss exponent*  $\gamma$  typically ranges between 2 and 6 [Rap96]. It is closer to 2 if there is a perfect line-of-sight between transmitter and receiver and can go up to 6 in dense urban areas.

We also need to model the sensor arrangement. To make our results portable to a wide range of applications, we want haphazardly distributed sensors rather than a fixed arrangement. We therefore use a *one-dimensional homogeneous Poisson process* with constant *rate*  $\lambda$  to model the distribution of sensors. Its points represent a random sequence of sensors distributed on a straight line. Hence, the Poisson rate  $\lambda$  is a measure of node density, and  $1/\lambda$  corresponds to the mean distance between adjacent nodes.

## 3 Conventional Approaches and the Optimal Case

There are two basic approaches to deliver data from a source node to the base station. First, data can be transmitted directly. This strategy is not feasible if there are many obstacles in between or the base station is too far apart in order to guarantee proper reception. However, direct transmission can be the method of choice if distances are relatively short or the energy required for reception is large. Another approach is to use intermediate nodes (relays) that retransmit data upon reception. In MTE routing, these relays are chosen such that the transmit amplifier energy ( $\epsilon d^\gamma$ ) is minimized [HCB00]. Hence, in a linear topology, each sensor transmits to its direct downstream<sup>1</sup> neighbor. For long-distance transmissions, MTE routing can dramatically reduce transmission power compared to direct communication. The drawback is that immoderate receive energy is dissipated if nodes are close to each other or the energy required for reception is high.

Consider now the case where data is to be delivered from a sensor, located at distance  $D$ , to the base station with minimum total energy consumption. Direct transmission to the base station is optimal if  $D \leq \alpha/(\epsilon(1 - 2^{1-\gamma}))$ . Otherwise, it is best to select  $(K_{opt} - 1)$  equally spaced nodes for retransmission [SL01].  $K_{opt}$  is the optimal number of hops which is either  $\lfloor D/d_{char} \rfloor$  or  $\lceil D/d_{char} \rceil$ , where  $d_{char}$  is the *characteristic distance* [BGC01] given by  $(\alpha/(\epsilon(\gamma - 1)))^{1/\gamma}$ . The characteristic distance is a constant, provided that all nodes are equipped with the same radio and the propagation environment is stable.

The problem we looked at is which of the two alternatives for  $K_{opt}$  is in fact the best choice. Let  $m = \lfloor D/d_{char} \rfloor$  and  $\delta = D/d_{char} - m$ . The choice  $\lfloor D/d_{char} \rfloor$  is optimal if  $\delta \leq (m^2 + m)^{1/2} - m$ , for  $\gamma = 2$ , and  $\delta \leq (3m^3(m + 1)^3/(3m^2 + 3m + 1))^{1/4} - m$ , for  $\gamma = 4$ , respectively. Otherwise,  $\lceil D/d_{char} \rceil$  is preferable in each case.

---

<sup>1</sup>Downstream means toward the base station.

## 4 MERR: Minimum Energy Relay Routing

In the previous section, we discussed the theoretical case that a certain optimal number of relays can be placed at desired positions to set up a minimum energy path. However, in a real linear network with sensors at arbitrary positions, such an optimal path is very unlikely to exist. The best we can do is to select appropriate nodes for retransmission in order to approximate the optimal case.

To this end, consider the characteristic distance, which can be seen as the globally optimal forwarding distance, that is, the distance a node should transmit its data onward in order to minimize the energy consumed on the entire path from source to base station. This observation leads to the basic idea of MERR. *Each sensor seeks locally for that downstream node within its maximum transmission range whose distance is closest to the characteristic distance.* Once a sensor has decided on the next hop, it adjusts its transmission power to the lowest possible level such that the radio signal can just be received by the respective node. In operation, a sensor transmits any data to its preassigned next hop node.

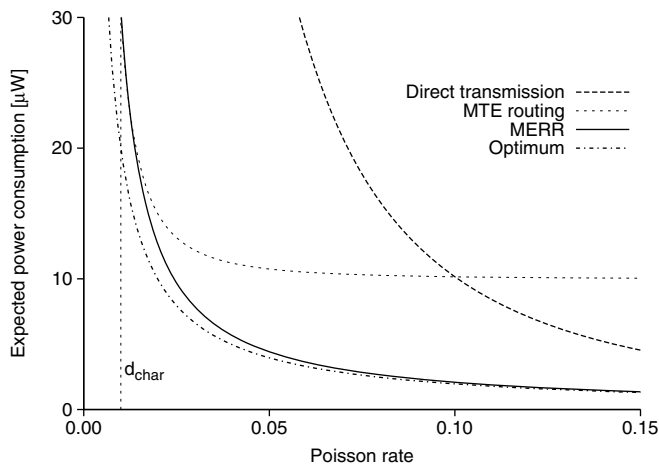


Figure 1: Expected power consumption against Poisson rate (node density) for a linear network of 100 sensors and path loss exponent 2. The radio parameters are adopted from [Hei00], yielding a characteristic distance of 100 m ( $\equiv 0.01$ ). MERR and MTE routing have equal power consumption when the Poisson rate corresponds to the characteristic distance (indicated by a dashed vertical line).

Figure 1 depicts the performance of MERR compared to the conventional approaches and the theoretical optimum. The graphs are generated using equations from our stochastic analysis and show expected power consumption against Poisson rate (node density). It can be seen that direct transmission falls far short for small node densities. MERR is bounded by MTE routing (upper bound) and the optimum (lower bound). Furthermore, MERR approaches the theoretical optimum as node density increases because the probability to find a near-optimal path increases as well. In fact, for  $\gamma = 2$ , MERR saves 80% of power compared to MTE routing and deviates less than 10% from the optimum if the distances between adjacent sensors are on average 10 m ( $\equiv 0.10$ ) and 50 m ( $\equiv 0.02$ ), respectively.

## 5 Conclusions

We propose MERR, a scalable and localized routing protocol that allows for energy-efficient data delivery in linear wireless sensor networks. Each sensor only needs to know the distances to its reachable downstream nodes and selects the next hop autonomously based upon the characteristic distance. Our analytic and simulation results show that MERR constitutes a significant improvement in terms of power compared to conventional techniques and performs close to the theoretical optimum.

## References

- [AKK04] J. N. Al-Karaki and A. E. Kamal. Routing Techniques in Wireless Sensor Networks: A Survey. *IEEE Wireless Communications Magazine*, 11(6):6–28, December 2004.
- [BGC01] M. Bhardwaj, T. Garnett, and A. Chandrakasan. Upper Bounds on the Lifetime of Sensor Networks. In *Proc. IEEE International Conference on Communications (ICC 2001)*, pages 785–790, June 2001.
- [EGHK99] D. Estrin, R. Govindan, J. Heidemann, and S. Kumar. Next Century Challenges: Scalable Coordination in Sensor Networks. In *Proc. IEEE/ACM International Conference on Mobile Computing and Networking (MobiCom 1999)*, pages 263–270, August 1999.
- [HCB00] W. Heinzelman, A. Chandrakasan, and H. Balakrishnan. Energy-Efficient Communication Protocol for Wireless Microsensor Networks. In *Proc. Hawaiian International Conference on Systems Science*, pages 1–10, January 2000.
- [Hei00] W. Heinzelman. *Application-Specific Protocol Architectures for Wireless Networks*. PhD thesis, Massachusetts Institute of Technology, 2000.
- [KPC<sup>+</sup>07] S. Kim, S. Pakzad, D. Culler, J. Demmel, G. Fenves, S. Glaser, and M. Turon. Health Monitoring of Civil Infrastructures Using Wireless Sensor Networks. In *Proc. IEEE/ACM International Conference on Information Processing in Sensor Networks (IPSN 2007)*, pages 254–263, April 2007.
- [ND05] M. Ndoj and G. Y. Delisle. Geolocation in Underground Mines Using Wireless Sensor Networks. In *Proc. IEEE Antennas and Propagation Society International Symposium*, pages 229–232, July 2005.
- [Rap96] T. S. Rappaport. *Wireless Communications: Principles & Practice*. Prentice-Hall, New Jersey, 1996.
- [SL01] I. Stojmenovic and X. Lin. Power-Aware Localized Routing in Wireless Networks. *IEEE Transactions on Parallel and Distributed Systems*, 12(11):1122–1133, November 2001.
- [SNM07] I. Stoianov, L. Nachman, and S. Madden. PIPENET: A Wireless Sensor Network for Pipeline Monitoring. In *Proc. IEEE/ACM International Conference on Information Processing in Sensor Networks (IPSN 2007)*, pages 264–273, April 2007.
- [ZDR07] M. Zimmerling, W. Dargie, and J. M. Reason. Energy-Efficient Routing in Linear Wireless Sensor Networks. In *Proc. IEEE International Conference on Mobile Ad-Hoc and Sensor Systems (MASS 2007)*, pages 1–3, October 2007.

# The Virtual Aachen Project

Gregor Fabritius  
Computer Graphics Group, RWTH Aachen University  
gregor.fabritius@rwth-aachen.de

## Abstract:

The goal of this student project was to create a realistic visualisation of the city of Aachen, based on 2D data available from the land registry office and photographs of the buildings. Therefor several applications were developed to automate the process, including tools for repairing defective 2D data and creating and texturing a 3D mesh, modeling landmark buildings for higher detail and a real-time viewer using advanced rendering effects.



Figure 1: Screenshots of the viewer showing the complete, textured mesh.

## 1 Introduction

Input to the project was data available from the land registry office in the form of vector-based 2D maps of buildings, streets and real estates and information about the terrain as height field. This data was inconsistent and partly incomplete, and in the original form not useable to create a suitable 3D mesh.

The overall task of virtualising a city based on this data has been split into four sub-projects in the context of a practical course at the Computer Graphics Group of RWTH Aachen University. At first the 2D data had to be repaired and transformed to a 3D mesh.



Second, to improve the visual realism of the mesh it had to be textured. Photographs have been taken from well-known parts of the city, the remaining buildings were textured using a procedural rule based algorithm. Third, semi-automatic modeling of landmark buildings was used to create models for those buildings that needed a more detailed and accurate model. The fourth and last task was to develop a viewer to get a realistic real-time visualisation using state-of-the-art rendering techniques.

## 2 2D Mesh Repair and Model Generation

The available 2D data had problems, e.g. the borders of the buildings (walls) did not belong to a specific building, the edges bordering two houses were present twice, the street data had gaps, and so on. It was necessary to repair these problems by snapping nearby vertices, creating polygon cycles that represent the outline of a building and separating outlines from inner courtyards. Afterwards it was possible to use the 2D data to create a 3D mesh (Fig. 2). At first the height of the buildings was estimated to expand the model to the third dimension. Then the buildings were extended to include roofs, depending on their outline. Finally the buildings were projected to a terrain mesh to get a complete mesh of the landscape. The 2D data contained information about different areas (street, green, building, ...) that was transferred to the 3D mesh.



Figure 2: Left: An example of erroneous street data with gaps. Middle: A close-up of the repaired 2D building floor plans. Right: The generated 3D mesh.

## 3 Texturing

To create textures from real pictures various distortions have to be removed. At first the pincushion distortion of the camera has to be corrected. The next step is to rectify [HZ04] the image (Fig. 3), because usually it is not possible to take a picture of a wall from an orthogonal view. Finally it is important to take care for different lighting conditions by correcting brightness, contrast and gamma values.

Since we only took pictures from the most prominent parts of the city, we additionally used a randomised rule based algorithm [MWH<sup>+</sup>06] to create textures for the rest of the buildings. This algorithm is based on a database of texture-templates from which wall textures are composed. These templates consist of a set of two-layered images: The ground layer is a wall base-texture (e.g. concrete or bricks) that allows to be seamlessly lined up. The second layer is a picture of a window-type with small variations (e.g. curtains open or closed).

To accomplish a random assignment of textures, the building is virtually subdivided into floors. Each floor is further subdivided into window-wide pieces. Then each piece is assigned a randomly selected window texture (with the adequate wall-element), the lowest floor gets a door-texture. The selection of the texture database is determining the variety of the final textures.

## 4 Modeling Landmark Buildings

Landmark buildings can not be modeled in sufficient detail from 2D outlines fully automatically. Hence we chose a semi-automatic [DTM96] approach. A coarse model of the building is created manually and it's edges are linked to the corresponding edges in a set of pictures (Fig. 3). After that, the automatic part of the algorithm calculates the real proportions of these edges in the model and creates corresponding textures as well. This way a detailed model of the building is created that can be embedded in the city-mesh.

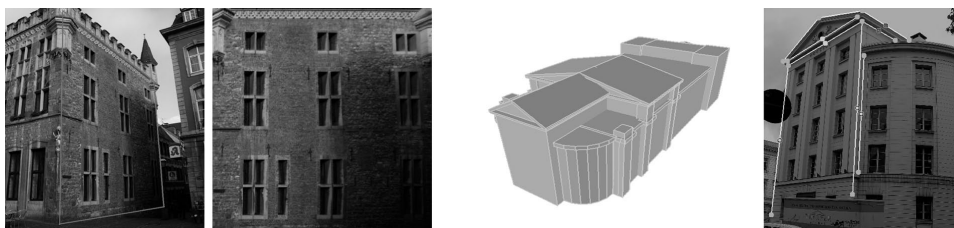


Figure 3: Left: To create a texture from a real picture it is necessary to remove perspective distortion by rectifying the image. Right: The edges of the hand-modeled building are connected to the corresponding edges on the picture. This semi-automatic process allows the creation of a detailed model with only moderate effort.

## 5 Visual Presentation

The final step is the visualisation of the textured model created by the above steps. To accomplish that, a sophisticated viewer was developed that can handle the huge amounts of data that were produced before: The final mesh consisted of 6.7M vertices, 2.2M triangles, more than 400 textures and virtually covered 41km<sup>2</sup> of area. An octree was used to keep

the memory usage reasonable and to optimise the GPU calls in order to achieve real-time performance.

The final viewer implements state-of-the-art features like procedural clouds, altitude of the sun with corresponding tone mapping, shadow maps, ambient occlusion, bloom and HDR effects (Fig. 1). Depending on the type of application it is possible to walk through the city like a tourist bound to the ground with collision detection for the buildings, or to smoothly fly over it on user-defined paths.

## 6 Conclusion

During this practical course a set of applications was created to handle the 2D maps and generate and display a 3D mesh. Together they represent a pipeline that can be used to visualise a city without major manual intervention. These tools are not limited to the city of Aachen, but may be applied to maps of other towns as well. By giving the user the opportunity to fine-tune the model manually, an arbitrary level of detail is possible. The viewer module gives a fast real-time user experience while presenting images of very high quality.

## 7 Acknowledgements

This project was a practical course at the Computer Graphics Group of RWTH Aachen University in the summer semester 2007. The members were (in alphabetical order): Gregor Fabritius, Lars Krecklau, Jan Krassnigg, Christopher Manthei, Christian Martelock, Matthias Passon, Dominik Rausch, Tillmann Vogt and Melanie Winkler. The supervising tutors were Martin Habbecke, Alexander Hornung and Leif Kobbelt.

## References

- [DTM96] Paul E. Debevec, Camillo J. Taylor, and Jitendra Malik. Modeling and Rendering Architecture from Photographs: A Hybrid Geometry- and Image-Based Approach. *Computer Graphics*, 30(Annual Conference Series):11–20, 1996.
- [HZ04] Richard Hartley and Andrew Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521540518, second edition, 2004.
- [MWH<sup>+</sup>06] Pascal Müller, Peter Wonka, Simon Haegler, Andreas Ulmer, and Luc Van Gool. Procedural modeling of buildings. *ACM Trans. Graph.*, 25(3):614–623, 2006.

# Qualitätssicherung von 3D Stadtmodellen

Stephan Gänß, Kay Kischlat, Jochen Kraushaar

Hochschule für Technik Stuttgart

Studiengang Informatik

Stephan-G@gmx.de

KayKis@web.de

jkraush@web.de

Art der Arbeit: Software-Projekt

Betreuer/in der Arbeit: Prof. Dr.-Ing. Volker Coors

**Abstract:** Es wird das Geometrie-Prüfprogramm QS-City 3D beschrieben.

## 1 Motivation

Ob Wetterdienst, Routenplaner oder andere Informationsmedien: 3D Landschaftsmodelle sind in aller Munde. Durch die Verbreitung des Internets steigt das Angebot an frei zugänglichen Geoinformationsdaten täglich an. Gleichzeitig erhöhen sich die Anforderungen an die Qualität der Daten. Bedingt durch Mess- oder Eingabefehler bei komplexen photogrammetrischen Verfahren der Luftbilderfassung gewinnt die Qualitätssicherung zunehmend an Bedeutung. So bieten inzwischen Standard Datenbanksysteme wie Oracle 11 die Möglichkeit, 3D Geometrien zu speichern und stellen Methoden zur Verfügung, welche die Gültigkeit dieser Geometrien prüfen [Oracle 07], [Kazar et al 08]. Dabei liegt der Definition gültiger 3D Geometrien die OGC / ISO GML 3 Spezifikation zugrunde [GML3].

Nun ist die Prüfung der Geometrie zwar durchaus sinnvoll, nur was geschieht, wenn in einem vorhandenen umfangreichen Modell zahlreiche fehlerhafte Geometrien vorkommen? Hier wird es interessant, Algorithmen zu entwickeln, die nicht nur die Geometrie prüfen, sondern auch in der Lage sind, nicht gültige Geometrien automatisiert zu korrigieren bzw. einen Korrekturvorschlag zu unterbreiten. Dies war die Ausgangslage eines Software-Projekts an der HfT Stuttgart. Ziel des Projekts war es, ein Programm zur Qualitätsprüfung und -verbesserung von CityGML-Modellen zu entwickeln. CityGML ist ein speziell für den Austausch von 3D-Stadtmodellen definiertes Anwendungsschema von GML3 [CityGML]. Dadurch sind die Ergebnisse des Projekts direkt übertragbar auf raumbezogene Datenbanken wie [Oracle 07], da dieselben geometrischen Elemente zugrunde liegen. Als Ausgangsdaten wurde das 3D Stadtmodell vom Stadtmessungsamts Stuttgart zur Verfügung gestellt, das insgesamt etwa 190.000 Gebäude umfasst.

## 2 Aufgabe des Programms

Die Aufgabe des Programms ist die halbautomatisierte Überprüfung der Gebäudegeometrie zur Unterstützung der Arbeitsprozesse. Wenn möglich sollen gefundene Fehler automatisch korrigiert und dem Benutzer visuell als Vorschläge präsentiert werden. Der Benutzer hat dann die Möglichkeit diese Vorschläge zu übernehmen oder zu verwerfen. Dadurch ist das Programm vielseitig einsetzbar: Es kann zur Endabnahme von Daten verwendet werden, die durch externe Firmen erstellt wurden.

Die bei einem CityGML Modell möglichen auftretenden Fehler sind vielfältig:

- Polygone enthalten doppelt definierte Punkte
- Polygone sind nicht geschlossen
- Polygone sind nicht planar
- Polygone schneiden sich
- Polygone sind falsch orientiert
- Polygone sind konkav
- Polygone enthalten Löcher
- Gebäude ist kein geschlossener Körper

## 3 Prüfalgorithmen

Für die Überprüfung der Modelle sind diverse geometrische und algebraische Prüfalgorithmen notwendig:

- **Überprüfung auf doppelte Punkte:** Die doppelten Punkte des Polygons werden unter Verwendung eines Hash-Sets entfernt.
- **Überprüfung auf Anzahl der Punkte:** Es müssen mindestens drei Punkte für ein Polygon gegeben sein, damit es als Fläche angesehen wird.
- **Überprüfung auf Planarität eines Polygons:** Bei dieser Überprüfung wird das Kreuzprodukt zweier unabhängiger Vektoren gebildet. Die Ebene wird durch das Kreuzprodukt und einen beliebigen Polygonpunkt eindeutig definiert. Wenn alle Punkte des Polygons in dieser Ebene liegen ist die Planarität erfüllt.

- **Überprüfung, ob ein Polygon sich selbst schneidet:** Zuerst wird geprüft, ob benachbarte Kanten übereinander liegen. Anschließend werden die nicht benachbarten Kanten betrachtet. Die Kanten werden als Geraden definiert. Für jedes Geradenpaar wird der eventuelle Schnittpunkt berechnet. Das geschieht durch Lösen eines Gleichungssystems mit Hilfe des Gauß Algorithmus. Existiert bei keinen Geradenpaaren ein Schnittpunkt ist das Polygon schneidungsfrei. Liegt ein Schnittpunkt vor, muss geprüft werden, ob dieser nur auf den Geraden, nur auf einer Kante oder auf beiden Kanten liegt. Im letzten Fall schneidet sich das Polygon.
- **Überprüfung auf Erfüllung des Polyedersatzes:** Der Eulersche Polyedersatz besagt, dass bei konvexen Polyedern folgende Gleichung erfüllt sein muss: Ecken + Flächen - Kanten = 2. Diese Formel wird vom Programm geprüft.
- **Überprüfung auf Anzahl der Polygone:** Da ein Gebäude mindestens durch vier Flächen gegeben sein muss, wird die Anzahl der Polygone geprüft.
- **Überprüfung auf richtige Orientierung der Polygone:** Die sichtbare Vorderseite eines Polygons ist gegeben durch die Anordnung der Punkte im mathematisch positiven Sinn. Im Kontext eines Gebäudes bedeutet richtige Orientierung also, dass der Normalenvektor des Polygons nach außen zeigt. Mit Hilfe der Winged Edge Datenstruktur werden alle Durchlaufrichtungen der Kanten auf ihre Korrektheit geprüft. Wenn Durchlaufrichtungen falsch sind, werden diese durch einen raffinierten Algorithmus korrigiert.

Zusätzlich bietet das Programm eine Möglichkeit zur Validierung der Modelle anhand des XML Schemas. Damit wird das Modell sowohl semantisch als auch syntaktisch auf ihre Korrektheit überprüft.

## 4 Fazit

Seit Juli 2007 wird das Programm QS-City 3D beim Stadtmessungsamt Stuttgart erfolgreich eingesetzt. *„Besonders hervorzuheben ist die semiautomatische Arbeitsweise des Tools, das auch die Überprüfung von 3D Beständen in CityGML erlaubt und wirtschaftlich unterstützt. Die aussagekräftige Identifikation von Fehlern erlaubt ein gezieltes Editieren der Modelle, falls eine automatische Korrektur nicht möglich ist. Insgesamt liegt uns hier ein unkompliziertes, wirkungsvolles Tool vor, das bislang so nicht auf dem Markt zu haben war. Tolle Leistung!“*, lobte Hans-Ulrich Mohl vom Stadtmessungsamt Stuttgart.

Weitere Informationen unter:

<http://wwwdev.hft-stuttgart.de/servlets/QSCity3D-WebDemo/>

## Literaturverzeichnis

- [GML3, CityGML] The Geographic Markup Language Specification und CityGML. Version 3.1.1, <http://www.opengeospatial.org>, <http://www.opengeospatial.org/standards/bp>
- [Oracle 07] Validation Rules and Algorithms for 3D Geometries in Oracle Spatial, 2007
- [Kazar et al 07] Kazar, B., R. Kothuri, P. van Oosterom, and S. Ravada: On Valid and Invalid Three-Dimensional Geometries, In P. Van Oosterom et al (Eds.) LNG&C Advances in 3D Geoinformation Systems, Springer 2008, pp 19-46.

# Massively Parallel Contact Simulation on Graphics Hardware using NVIDIA CUDA

Jens-Fabian Goetzmann

post@jefago.de

**Abstract:** This article describes an approach to use the massively parallel computing capabilities of modern graphics hardware to conduct a physical simulation in order to find good solutions to difficult packing problems. Using graphics hardware is shown to be able to provide significant speedups to some classes of problems while other classes of problems are not able to benefit from the massively parallel execution.

## 1 Introduction

The computing resources available on current generation graphics hardware are enormous. Driven especially by the game market's demand for ever more realistic 3-D graphics, current graphics adapters such as the NVIDIA GeForce 8 series feature theoretical computing performances that are a multiple higher than those of current CPUs ([Gre07], [NVI07]), with an even increasing gap.

NVIDIA CUDA<sup>1</sup> (Compute Unified Device Architecture) is an SDK (Software Development Kit) released by graphics hardware manufacturer NVIDIA with the purpose of making it possible for programmers to accelerate general-purpose computations by using the computational resources available to modern GPUs (Graphic Processing Units). The current 1.0 release was announced in June 2007. It features a specially augmented C language.

CUDA is the first SDK to expose the resources on the graphics hardware for general purposes. Before the advent of CUDA, it was only possible to use the GPU for non-graphics purposes by writing specially crafted shader programs, the programming language of which is specifically designed for graphics applications, and also has severe limitations regarding control flow and memory access.

### 1.1 Specialities of CUDA programs

There are numerous facts that have to be considered prior to writing CUDA programs. First and foremost is the fact that in order to fully take advantage of the processing capabilities

---

<sup>1</sup>Called "CUDA" in the following for the sake of simplicity.



of the GPU, the number of threads run parallelly has to be as high as possible. Furthermore, threads have to be *batched* into equally-sized *blocks*. Threads from the same blocks can be synchronized and have effective means of communication, while threads from different blocks may be run on different multiprocessors and thus can not communicate at all.

CUDA also offers multiple types of memory, all of which have specific advantages and disadvantages regarding speed, accessibility and available amounts. Program data may have to be copied between different memory areas to maximize program efficiency.

The control flow of CUDA programs has to be specially crafted: Because of the SIMD (single instruction, multiple data) design of the GPU multiprocessors, deviations within the control flow of threads executed on a single multiprocessor will lead to serialization of the threads' execution thus losing the advantage of parallel execution. Additionally, recursion is not allowed in CUDA functions; It is also impossible to start another batch of threads from within a parallel execution.

## 2 Packing Problems and Contact Simulation

The objective of the work done for the thesis was to use CUDA to accelerate contact simulation algorithms in order to find good solutions for packing problems. Packing problems are a well-studied class of problems in computer science. They are usually defined as either decision problems ("Given a set of items, do they fit into a specified container?") or as optimization problems ("How many items fit into a given container?" or "What is the minimum container for a given set of items?"). A large sub-class of packing problems are of geometrical type – whether two- or three-dimensional. In these packing problems, the objects to be packed and the container can be described as geometric objects. The objects and the container may have arbitrarily complex shapes and a different number of degrees of freedom.

As most geometric packing problems are NP-complete, an optimal solution is in most cases impossible to find ([BSW]). Thus all approaches for finding good packings compute some approximation to an optimal solution. One possible approach is a physically based *contact simulation*, i.e. a simulation of the objects to be packed as rigid bodies that have a certain mass and can move, collide with each other and the container and exchange impulses. A series of contact simulation steps will in almost all cases lead to a stable state, in which no object overlaps any other object anymore. Thus optimizations can be performed by e.g. putting one more object into the container (in a possibly illegal state) and legalizing the packing using contact simulation.

The problems considered in this case are the two-dimensional packing of circles of the radii 1 to  $n$  in the smallest possible enclosing circle ([Zim05], [ALS06]) and the packing of as many  $200 \times 100 \times 50$ -mm-sized boxes as possible into a trunk, which is modeled by a dense point cloud ([EFRS03], [EFK<sup>+</sup>05]).

### 3 Using CUDA for Contact Simulation

For the problem of circle packing, different parallelization strategies were evaluated.

The first one uses one packing at a time and parallelly executes several randomized improvement trials followed by a series of contact simulation steps. This strategy leads to a quick convergence of the packing to a *stable* packing that cannot be improved anymore by randomized improvement trials.

The second strategy uses multiple independent packings, that are parallelly tried to be improved. The improvement of the different instances is monitored centrally, and if a packing does not improve anymore, it is replaced by a newly generated, random starting packing. This strategy was conceived to be run for a long time, in the hope that eventually a very good packing would result.

The last strategy fully parallelizes the contact simulation of just one packing. However, it turned out that this is not efficient at all, mainly because there is a need for very much inter-thread communication, which is a severe bottleneck.

The problem of trunk packing is incomparably more difficult: The boxes have more degrees of freedom, and a point cloud is a very complicated container when compared to an enclosing circle.

To be able to parallelize the contact simulation for the trunk packing problem, several optimization steps have to be performed: The points are stored in a regular grid. For each box, the grid cells overlapped by the axis aligned bounding box of the box are then searched parallelly and the points are checked for collision with the box. For the collision test of the boxes among each other, the pairs are first filtered by the distance of their centers, before performing a separating axes test ([GLM96]) to check the eligible pairs for collision.

### 4 Results

The evaluation of using NVIDIA CUDA for massively parallel contact simulation has lead to ambivalent results: For the problem of circle packing, an efficient implementation was created that performs about 50 times as fast as a comparable serial implementation. For the complex problem of trunk packing, the conceived implementation performs only slightly better than a comparable serial implementation.

All in all, it could be proved that the enormous computing resources of modern graphics hardware can now be easily used for non-graphics tasks. On the other hand, it has become clear that no problem can be parallelized without thorough consideration of the specialities of CUDA and the GPU, and that some problems can not benefit from the hardware in spite of a highly optimized implementation.

## References

- [ALS06] Bernadetta Addis, Marco Locatelli, and Fabio Schoen. Efficiently packing unequal disks in a circle: A computational approach which exploits the continuous and combinatorial structure of the problem. *Optimization Online*, 1343, 2006.
- [BSW] Tobias Baumann, Elmar Schömer, and Kai Werth. Solving geometric packing problems based on physics simulation. Unpublished Manuscript.
- [EFK<sup>+</sup>05] Friedrich Eisenbrand, Stefan Funke, Andreas Karrenbauer, Joachim Reichel, and Elmar Schömer. Packing a Trunk - now with a Twist! In *ACM Symposium on Solid and Physical Modeling*, pages 197–206, 2005.
- [EFRS03] Friedrich Eisenbrand, Stefan Funke, Joachim Reichel, and Elmar Schömer. Packing a Trunk. In *11th European Symposium on Algorithms*, pages 617–629, 2003.
- [GLM96] S. Gottschalk, M. C. Lin, and D. Manocha. OBBTree: A Hierarchical Structure for Rapid Interference Detection. *Computer Graphics*, 30(Annual Conference Series):171–180, 1996.
- [Gre07] Simon Green. NVIDIA CUDA FAQ Version 1.0. <http://forums.nvidia.com/index.php?showtopic=36286>, 2007.
- [NVI07] NVIDIA Corporation. NVIDIA CUDA Compute Unified Device Architecture, Programming Guide. [http://developer.download.nvidia.com/compute/cuda/1\\_0/NVIDIA\\_CUDA\\_Programming\\_Guide\\_1.0.pdf](http://developer.download.nvidia.com/compute/cuda/1_0/NVIDIA_CUDA_Programming_Guide_1.0.pdf), 2007.
- [Zim05] Al Zimmerman. Al Zimmermann's Circle Packing Contest. <http://www.recmath.org/contest/CirclePacking/index.php>, 2005.

# **Anwendergestützte Erkennung und Auswertung ausgewählter Holzstrukturen für die Dendrochronologie mit Methoden der digitalen Bildverarbeitung**

Simon Pietsch

Fachhochschule Wiesbaden  
Fachbereich Design Informatik Medien  
[NomisP@gmx.de](mailto:NomisP@gmx.de)

Art der Arbeit: Diplomarbeit  
Betreuer der Arbeit: Prof. Dr. D. Richter  
N. Bleicher, M.A.

**Abstract:** Im Rahmen der vorliegenden Arbeit werden bildanalytische Möglichkeiten zur Auswertung von saisonalen Klimainformationen innerhalb eines Jahresrings geschaffen. Hierzu müssen wasserführende Gefäße und Grenzen von Jahresringen in Holzschnitten exakt erkannt und ausgewertet werden. Es kommen verschiedene Methoden der digitalen Bildverarbeitung, wie Bildfilter, morphologische Operatoren und Segmentierungsverfahren zum Einsatz. Die damit erzielten Ergebnisse werden zu einem neuen Verfahren zur Erkennung von Jahresringgrenzen kombiniert. Einen wichtigen Bestandteil bildet die Benutzeroberfläche für „Nichtinformatiker“, die alle Verarbeitungsschritte detailliert visualisiert und interaktiv steuerbar macht.

## **1. Einleitung**

Die Dendrochronologie<sup>1</sup> ist die Wissenschaft zur Altersbestimmung von Bäumen. Jeder Baum bildet während der Vegetationsperiode eines Jahres einen kompletten Jahresring aus. Das Wachstum korreliert unmittelbar mit den verfügbaren Nährstoffen, insbesondere dem vorhandenen Wasser. Die Breite eines Jahresrings lässt somit direkte Rückschlüsse auf das Ausmaß von Niederschlägen während eines speziellen Jahres zu [Sch83].

Bislang wurde für die Klimarekonstruktion hauptsächlich die Breite der Jahresringe herangezogen. Dabei blieben sämtliche saisonale Informationen innerhalb der Jahresringe unberücksichtigt.

Mit Methoden der Digitalen Bildverarbeitung wird untersucht, wie die wasserführenden Gefäße von zerstreutporigen Laubhölzern in digitalen Bildern detektiert und klassifiziert werden können. Es wird ein Verfahren dargestellt, um plausible Jahresringgrenzen zwischen einzelnen Jahresringen festlegen zu können. Aus dem Verlauf einer variablen Verteilung der Porengröße innerhalb eines Jahresrings wird eine saisonale Niederschlagsrekonstruktion möglich.

---

<sup>1</sup> dendron = Baum, chronos = Zeit und logos = Lehre

## 2. Das Bildmodell

Für die Auswertung der Bilder wird ein Format von 768\*576 Pixel und eine Grauwertauf Auflösung von 8 Bit vorausgesetzt. Zusätzlich müssen folgende Bedingungen erfüllt sein:

- In den Aufnahmen der Holzbilder verläuft die Wachstumsrichtung des Baumes vom rechten zum linken Bildrand.
- Eine Jahresringgrenze verläuft im Bild immer über die komplette Höhe, vom oberen zum unteren Rand.
- Das Bild muss in der Breite mindestens einen vollständigen Jahresring enthalten. Ein Jahresring liegt vollständig im Bild, wenn mindestens eine Vegetationsperiode vollständig durch zwei sichtbare Grenzen abgebildet ist.

## 3. Die Wasserscheidentransformation

Die wasserführenden Gefäße von zerstreutporigen Laubhölzern müssen in einem Bild möglichst exakt erkannt werden. Durch eine problemadaptierte Wasserscheidentransformation wird erreicht, dass für jedes einzelne Gefäß automatisch eine separate Pixelregion erzeugt wird.

Der Vincent- Soille Algorithmus [Vin91] beschreibt den Überflutungsprozess eines Grauwertgebirges. Die Täler des Gebirges werden als Auffangbecken für das von unten einströmende Wasser bezeichnet. Jede der 256 Höhenstufen wird bei 0 beginnend der Reihe nach durchlaufen. Jedes Pixel wird dabei einem eindeutigen Becken zugeordnet. Wenn das Wasser aus zwei oder mehreren verschiedenen Becken zusammenfließt, werden Wasserscheiden (Dämme) genau in der Mitte zwischen den einzelnen Becken gebildet [Soi98].

Da die Wasserscheidentransformation häufig stark übersegmentiert, muss das Bild mit einem geeigneten Glättungsfilter vorverarbeitet worden sein.

## 4. Die Objekterkennung

Jede erzeugte Pixelregion enthält im Idealfall genau ein wasserführendes Gefäß und einige Hintergrundpixel.

### 4.1. Die Binarisierung von Otsu

Die automatische Binarisierung von Otsu [Ric05] verwendet statistische Verfahren, um aus den Bilddaten einen guten Schwellwert zu berechnen. Dabei spielt vor allem die Varianz, ein Streuungsmaß für die Abweichung von Grauwerten, eine wichtige Rolle. Ausgangspunkt der Überlegung ist ein annähernd bimodales Histogramm mit zwei lokalen Häufungen. Unabhängig von der tatsächlichen Verteilung terminiert der Algorithmus immer mit einer Grauwertschwelle.

Durch ein anschließendes Opening, d.h. eine Aufeinanderfolge von Erosion und Dilata-

tion, werden in dem entstandenen Binärbild Ausreißer eliminiert und schwach zusammenhängende Flächen vereinzelt.

#### **4.2. Der “Connect Component Labeling“- Algorithmus**

Der Connect Component Labeling Algorithmus [Ric05] erkennt zusammenhängende Objekte in einem Binärbild. Es werden alle eigenständigen Objekte mit einem eindeutigen Label markiert. Im Anschluss können alle benötigten Informationen wie Breite, Höhe, Objektgröße und Schwerpunkt eines wasserführenden Gefäßes ermittelt werden.

### **5. Die Jahresringerkennung**

Jahresringe weisen folgende Eigenschaften auf:

- Zu Beginn eines Jahres bildet ein Baum unmittelbar die ersten wasserführenden Gefäße aus. Im so genannten Frühholz sind große und zahlreiche Gefäße vorhanden.
- Gegen Ende eines Jahres, im Spätholz, bildet ein Baum nur noch spärlich, relativ kleine Gefäße aus.

Eine Jahresringgrenze wird im Übergang von kaum und mit kleinen Gefäßen besetzten Regionen zu Gebieten mit vielen großen Gefäßen gesucht.

#### **5.1. Die Bildeinteilung**

Jahresringe sollen durch einen Polygonzug mit einigen Stützpunkten approximiert werden. In einem ersten Schritt wird das Bild in gleich große Streifen eingeteilt. Für die Streifenbreite wird wahlweise das flächenmäßig größte Wassergefäß oder das größte Auffangbecken herangezogen. Jeder Streifen stellt ein lokales Koordinatensystem dar und ist 768 Pixel lang.

#### **5.2. Aufstellung einer Analysefunktion**

Für jeden Streifen wird eine eigenständige Funktion  $f(x)$  mit  $x = 0, \dots, 767$  aufgestellt. Der Funktionswert an der Stelle  $x$  wird aus dem Mittelwert aller lokal erkannten Gefäßgrößen berechnet. Jahresringgrenzen, die mit Hilfe der Objektgrößenfunktion ermittelt werden, können nur durch Nullstellen der Funktion verlaufen.

#### **5.3. Die „Nullketten“ der Objektgrößenfunktion**

Eine Nullkette ist eine konstante Folge von Funktionswerten mit der Eigenschaft  $f(x_1) = f(x_2) = \dots = f(x_n) = 0$ ,  $n \in \mathbb{N}$ . Die Länge einer solchen Kette ist die Anzahl aufeinander folgender  $x$ -Werte. Der linke Rand einer Nullkette gibt die mögliche Position einer Jahresringgrenze an.

#### 5.4. Die Kostenfunktion

Die entwickelte Kostenfunktion bewertet die Länge jeder Nullkette und den Abstand zweier Ketten aus übereinander liegenden Streifen. Die Kosten werden minimal, wenn die Nullketten besonders lang und der Abstand zur nächsten Kette besonders gering ist. Die Bewertungsfunktion wählt automatisch die am besten bewerteten Polygonzüge entlang der Nullketten aus.

### 6. Oberflächengestaltung und Anwendung

Die Benutzeroberfläche wurde so entwickelt, dass sie jeden einzelnen Bearbeitungsschritt detailliert dokumentiert und visualisiert und viele interaktive Unterstützungsmöglichkeiten für den Programmablauf anbietet. Diese Unterstützung ist nötig, da nach der Wasserscheidentransformation trotz einer Vorfilterung des Originalbildes einzelne falsche Wasserscheiden verbleiben können. Während der Objekterkennung werden gelegentlich einzelne Gefäße über- oder untersegmentiert. Außerdem werden ab und zu einige Störstellen, z.B. Markstreifen, als Wassergefäße erkannt.

Die Qualität der Erkennung der Jahresringgrenzen hängt maßgeblich von der Güte der jeweils vorhergehenden Zwischenschritte ab. Wenn die Ergebnisse der Wasserscheidentransformation und der Objekterkennung ausreichend nachbearbeitet wurden, können sehr genau die Grenzen der Jahresringe ermittelt werden.

Zusammenfassend kann festgestellt werden, dass die untersuchten Verfahren eine weitgehend automatische Jahresringanalyse ermöglichen. Trotz einiger interaktiver Zwischenschritte wird die Bearbeitungszeit für die Klassifizierung von Jahresringen in Holzschnittbildern erheblich verkürzt.

Diese Arbeit wurde in Zusammenarbeit mit dem Römisch-Germanischen Schiffahrtsmuseum in Mainz erstellt.

### Literaturverzeichnis

- [Sch83] SCHWEINGRUBER, F.H.: Der Jahrring. Standort, Methodik, Zeit und Klima in : Dendrochronologie. Verlag Paul Haupt Bern und Stuttgart, 1983
- [Vin91] VINCENT/SOILLE: Watersheds in Digital Spaces: An Efficient Algorithm Based on Immersion Simulations, 1991  
<http://www.cse.msu.edu/~cse902/S03/watershed.pdf>
- [Soi98] SOILLE, Pierre: Morphologische Bildverarbeitung. Springer, 1998
- [Ric05] RICHTER, Detlef: Lehrveranstaltung Digitale Bildverarbeitung WS 2005.  
<http://www.informatik.fh-wiesbaden.de/~richter>

# Unterstützung der Prozessmodellierung im Notfallmanagement

Emilija Arsenova

RWTH Aachen  
Informatik V (Media Informatics)  
emilija.arsenova@gmail.com

Art der Arbeit: Master-Arbeit

Betreuer/in der Arbeit: Prof. Thomas Rose, Prof. Matthias Jarke

**Abstract:** Die Vorbereitung auf Notfallslagen, die sich durch Naturkatastrophen oder technische Unfälle ergeben können, nimmt eine zunehmende Bedeutung für Rettungsorganisationen und Behörden ein. Derartige Vorgehensweisen werden aber überwiegend in Textdateien dokumentiert, die nicht formal verarbeitbar sind. In diesem Beitrag wird ein prozessorientierter Ansatz für die Erfassung und Auswertung von Vorgehensweisen im Notfallmanagement vorgestellt. Als Beispiel wird das ÜMANV Einsatzkonzept (Überörtliche Unterstützung beim MassenANfall von Verletzten) [Sc07] verwendet. Die für diese Domäne speziell erarbeitete Modellierungsumgebung versetzt die Rettungsorganisationen in die Lage, ihre Prozesse selbst zu modellieren, zu visualisieren und zu analysieren. Die modellierten Vorgehensweisen in Form von formalen Referenzprozessen gewährleisten Einheitlichkeit, Wiederverwendbarkeit und Auswertbarkeit.

## 1 Hintergrund

Bei einem Großschadensfall liegt die Komplexität des Vorgehens in der Abstimmung der zu erbringenden Leistungen verschiedener Organisationen. Die Konsequenzen einer mangelnden Prozessorientierung, welche sich in dem System des Katastrophenschutzes immer wieder finden, sind eine Verteilung von Prozessen auf viele Aufgabenträger, zeitliche Verzögerungen, redundante Ausführungen von Aufgaben sowie ein hoher Abstimmungsaufwand [Bü05]. Aus diesen Gründen müssen Rettungsorganisationen ihr gegenseitiges Zusammenwirken während der Planung von Vorgehensweisen in Betracht ziehen. Demzufolge müssen Behandlungen, Dienstleistungen und logistisches Vorgehen in Form von Referenzprozessen vereinbart werden. Die Referenzprozesse ermöglichen einen Know-how-Austausch und eine Wiederverwendung von Kenntnissen zwischen regional unterschiedlichen Organisationen. Um diese Prozesse auswerten und analysieren zu können, müssen die Aktivitäten, die in der Praxis überwiegend auf Papier dokumentiert sind, formal modelliert werden. Dies hat wiederum eine Verringerung der Übungen zur Auswertungszwecken zur Folge. Außerdem liefert das Modellieren von Prozessen Informationen über die Gründlichkeit des Katastrophenplans [HJ05].



Prozesse bilden eine natürliche Metapher für die formale Darstellung von Vorgehensweisen im Notfallmanagement. Ein Prozessmodell kann den Ausgangspunkt zur Koordination von Vorgehensweisen zwischen den beteiligten Parteien bilden. Außerdem können Prozesse die Basis für die Simulation von Krisenszenarien und die Schulung von Rettungseinheiten errichten. In der Vergangenheit haben sich bei Großschadenslagen vor allem Schwächen in der Kooperation und Kommunikation der Beteiligten gezeigt [Bü05]. Eins der vielen Ziele der Prozessmodellierung ist das Ermöglichen von Verständigung und Kommunikation [CK92]. Prozesstransparenz hilft den Prozessbeteiligten zu verstehen, welche Rolle sie in dem Spiel haben [NF99].

Bisher konnte sich allerdings noch keine formale Prozessmodellierung in der Praxis etablieren; insbesondere keine Prozessmodellierung durch die Domänenexperten selbst. Es existieren lediglich einzelne Implementierungen von Vorgangsmanagementsystemen zur Steuerung und Absicherung des Informationsflusses [MM99].

Am Beispiel des ÜMANV Einsatzkonzepts werden die Möglichkeiten der Prozessmodellierung für das Notfallmanagement gezeigt. Ausgangspunkt der Modellierung ist eine textuelle Spezifikation des Vorgehens beim Massenansturm von 500 - 1200 Verletzten. Die vorliegende Spezifikation der Vorgehensweise [Sc07] umfasst ca. 45 Seiten Textdokument, das um Grafiken und Diagramme zur Visualisierung angereichert ist, die bereits eine prozessartige Sicht suggerieren aber nicht formalisieren.

## **2 Lösungsansatz der Prozessmodellierung im Notfallmanagement**

Gemeinsam mit Domänenexperten von der Berufsfeuerwehr Köln wurde im Rahmen dieser Arbeit eine Modellierungsmethodik für Notfallprozesse entwickelt. Mittels Interviews lernten wir, dass Feuerwehren sowie andere Rettungsorganisationen die Handhabung von Katastrophen als eine Wechselfolge von Meldungen und Maßnahmen wahrnehmen. Aus diesem Grund wurde auf die Methodik der EPK (Ereignisgesteuerten Prozessketten) [KN91] aufgebaut, die aber terminologisch und semantisch auf Seiten des Werkzeugs adaptiert wurde. Die resultierende Modellierungssprache spiegelt alle Elemente und Beziehungen von der Domäne wieder, welche für die Spezifizierung von Notfallprozessen erforderlich sind. Die semi-formale Modellierungsmethodik ermöglicht den Mitarbeiter die Prozessmodelle zu verstehen und sie selbst zu gestalten.

Um genauere Prozessbeschreibung aus mehreren Sichten zu ermöglichen, wurde ein domänenspezifischer Ordnungsrahmen entwickelt. Als eine passende und verwendbare Basis für die Entwicklung des Ordnungsrahmens wurde ARIS (Architektur integrierter Informationssysteme) [SN00] betrachtet, da die getrennten jedoch in Beziehung stehenden Ansichten eine Übersicht strategischer Ziele sowie eine ausführliche Sicht auf die Erbringung von Leistungen ermöglichen. Die Modellierungssprache für Notfallprozesse erforderte jedoch eine entsprechende Anpassung der bestehenden sowie Ergänzung durch neuen Modellierungssichten. Die hieraus entstandenen sechs Modellierungssichten fördern die Verdichtung der wichtigsten Aspekte eines Einsatzplans und erleichtern somit seine Abstimmung unter den Beteiligten. In der *Strategiesicht* werden sowohl strategische Ziele, als auch die Hauptmaßnahmen die sie

unterstützen, dargestellt. Die *Konzeptsicht* befasst sich mit der logisch und zeitlich eingeschränkten Darstellung der Hauptmaßnahmen eines Einsatzkonzepts, um den natürlichen Ereignisfluss aufrechtzuerhalten. Innerhalb der *Flusssicht* werden Leistungen bis zu der Ebene von einzelnen Maßnahmen dargestellt (Abb. 1). Die Maßnahmen können durch Maßnahmenträger, taktische und operative Ziele, Einsatzregeln sowie Leistungsanforderungen genauer beschrieben werden. Die *Aufbausicht* stellt die erforderlichen Kompetenzen der Maßnahmenträgern, die Kommunikationsstruktur zwischen den teilnehmenden Einsatzkräften sowie technischen Ressourcen dar. Die Vorstellung dessen was die Maßnahmen erreichen sollen, und warum, wird in der *Zielsicht* definiert. Die *Anforderungssicht* illustriert Leistungsanforderungen, die für die Ausführung von Maßnahmen erforderlich sind, oder aus deren Ausführung resultieren.

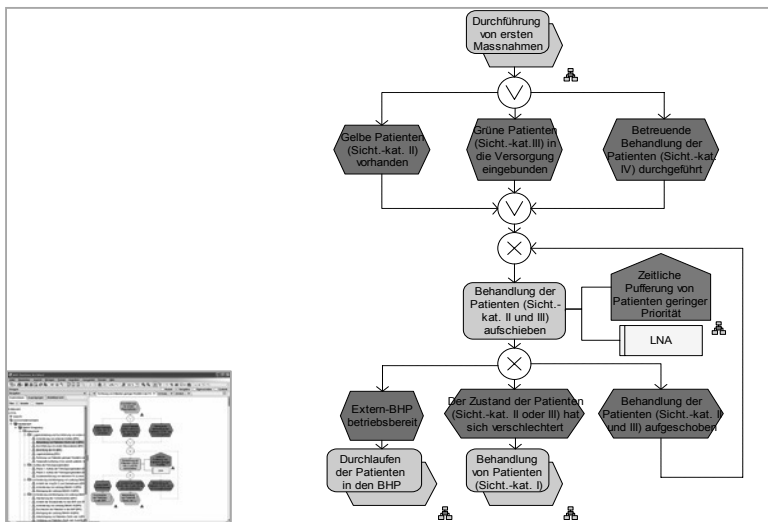


Abbildung 1: Prozessmodell der Pufferung von Patienten geringer Priorität

Die in Übereinstimmung mit ihrer derzeitigen Arbeitsweise entworfenen Leitlinien helfen den Prozessgestalter bei seinem Modellierungsansatz. Während die vorgeschriebenen Modellierungsschritte die Ansichten des Ordnungsrahmens einheitlich vollenden, stellen die vorgeschlagenen Auswertungsschritte sicher, dass sich keine Konzeptdiskrepanzen in das Prozessmodell einschleichen. Im Unterschied zu textuell erstellten Vorgehensweisen wird die frühe Entdeckung von Unstimmigkeiten unterstützt.

Um eine effiziente Handhabung der Katastrophenszenarien zu ermöglichen, wurde eine schrittweise Anleitung zur Auswertung von Prozessen erstellt. Durch eine Umwandlung der identifizierten kritischen Faktoren in Auswertungskriterien wurde das Prüfverfahren direkt von den Bedürfnissen der Domäne abgeleitet. So wird schnell klar, ob und wo es Koordinationsengstellen oder einen Mangel bestimmter Ressourcen gibt, die eine Umgestaltung des Prozesses verlangen. Somit haben die modellierten Referenzprozesse erstmals eine ganzheitliche, analytische Untersuchung der Vorgehensweisen ermöglicht und zu einer fachlichen Präzision geführt.

## Literaturverzeichnis

- [Bü05] Bürmann, R.: Geschäftsprozessmanagement im Bevölkerungsschutz. Nationale Sicherheit und Bevölkerungsschutz, Homeland Security, 2005.
- [CK92] Curtis, B.; Kellner, M.I.; Over, J.: Process Modeling. In: Communications of the ACM 35, New York, 1992; S. 75-90.
- [HJ05] Hoogendoorn, M.; Jonker, C.M.; Popova, V.; Sharpanskykh, A.; Xu, L.: Formal Modeling and Comparing of Disaster Plans. In (B. Van de Walle and B. Carlé, eds.): Proceedings of the 2nd International ISCRAM Conference, Belgium, 2005.
- [KN91] Keller, G.; Nüttgens, M.; Scheer, A.-W.: Semantische Prozessmodellierung auf der Grundlage „Ereignisgesteuerte Prozessketten (EPK)“. Veröffentlichung des Institut für Wirtschaftsinformatik, Paper 089, 1991.
- [MM99] Mak, H. Y.; Mallard, A. P.; Bui, T.; Au, G.: Building online crisis management support using workflow systems. In: Decision Support Systems, 25, 1999; S. 209-224.
- [NF99] Negele, H.; Fricke, E.; Schrepfer, L.; Härtlein, N.: Modeling of integrated product development processes. In: Proceedings of the 9th Annual Symposium of INCOSE, UK, 1999.
- [Sc07] Schmidt, J.: Einsatzkonzept ÜMANV - "MANV überörtlich", Rheinische Projektgruppe „MANV Überörtlich“, Köln, 2007.
- [SN00] Scheer, A.-W.; Nüttgens, M.: ARIS Architecture and Reference Models for Business Process Management. In (van der Aalst, W.M.P.; Desel, J.; Oberweis, A.): Business Process Management - Models, Techniques, and Empirical Studies, LNCS 1806, Springer-Verlag, Berlin et al., 2000; S. 366-399.

# Integration of Content Into Enterprise Applications - Analysis, Conceptualization and Prototypical Implementation Using Java EE and JCR

Hendrik Beck

Fachbereich Informatik  
Hochschule Darmstadt  
hendrik.beck@camunda.com

**Abstract:** Der Anteil von unstrukturierten und semi-strukturierten Daten, sog. „Content“, in Unternehmen wächst kontinuierlich. Unternehmensanwendungen, die wichtigsten Systeme in durchschnittlichen Software- und Systemlandschaften, basieren auf dem Paradigma vollstrukturierter Daten und bieten keine gute Unterstützung zur Verarbeitung von Content an. Diese Arbeit untersucht Realisierungsvarianten, sammelt Einflussfaktoren für eine fundierte Entscheidung und analysiert das Kostenverhalten. Außerdem wird eine prototypische Implementierung zur Integration von Hybrid-Lösungen vorgestellt.

## 1 Problemstellung

Der Begriff „Content“, außerhalb theoretischer Forschung vor einigen Jahren im Umfeld Daten-zentrierter Web-Applikationen und mit Content-Management-Systemen populär geworden, erlangt immer größeren Stellenwert in Organisationen. Beim Gesamtblick auf Unternehmen lassen sich insbesondere mit der Content-Eigenschaft schwach strukturierter bzw. unstrukturierter Daten große Mengen an vorhandenen Informationen identifizieren. Nach Gartner lassen sich rund 80% aller Unternehmensdaten dieser Kategorie zuordnen. Die Disziplin des Enterprise Content Managements unternimmt hierbei Definitionen und Abgrenzungen und liefert Lösungen zur effektiven Verarbeitung dieses Typs von Daten. Das Enterprise Content Integration erarbeitet Strategien und Konzepte zur Schaffung einheitlicher und integrierter Daten-Repositories auf Unternehmensebene.

Im Kontext von Unternehmensanwendungen dagegen spielt Content traditionell eine untergeordnete Rolle. Diese Anwendungen arbeiten grundsätzlich mit dem Ansatz vollstrukturierter Daten, bei denen meist objekt-orientiert modellierte Entitäten auf relationale Datenbanken abgebildet werden. Unstrukturierte und semi-strukturierte Daten lassen sich mit diesem Ansatz nur unzureichend verarbeiten. Meist beschränkt es sich darauf, diese Daten und Dokumente als Binary Large Objects oder als Text anzusehen und jegliche Struktur und Semantik zu ignorieren. Spezielle Anwendungsfälle im Zusammenhang mit Content wie beispielsweise Versionierung, Internationalisierung, mächtige Volltextsuchen und effektive Verarbeitung von Binärdateien finden in Unternehmensanwendungen mangelhafte

Unterstützung bzw. müssen im Einzelfall teuer und aufwändig implementiert werden. In der Praxis resultiert daraus oft ein paralleler Einsatz von Unternehmensanwendung und Content-Management-System, um die Anforderungen beider Welten abdecken zu können.

## 2 Architekturvarianten

Bei dieser sog. „Hybrid“-Lösung werden die Speicherorte für Daten, also für einzelne Attribute von fachlichen Entitäten, aufgeteilt je nach Zugehörigkeit zu einer der Typen. Dadurch geht eine einheitliche Sicht auf diese Daten verloren, obwohl sie zu fachlich gleichen Entitäten identifiziert wurden. Darüber hinaus werden die Entscheidungen über den konkreten Speicherort oft mit unzureichendem Bewusstsein für den Problembereich und falschen Argumentationen getroffen. Schließlich enden diese Realisierungsszenarien oft mit erhöhtem Aufwand bei der Re-Integration beider Datenquellen, um die ursprünglich einheitliche Sicht wieder her zu stellen. Dies wird immer dann notwendig, wenn Daten am gleichen Ort benötigt werden, z.B. bei der Darstellung auf Webseiten oder zur Pflege in Backend-Anwendungen.

Die Arbeit gibt Architekten und Entwickler einen konkreten Entscheidungsprozess an die Hand, um auf Grundlage von Daten, die vom Zielsystem verarbeitet werden müssen, auf eine Architekturentscheidung zu schließen. Die Daten werden dazu mittels konkreten Methoden untersucht und individuell in eine der beiden Kategorien „Enterprise-Daten“ oder „Content-Daten“ eingeordnet. In einem ersten Schritt führt die Summe dieser Ergebnisse zur Entscheidung zwischen folgenden Architekturvarianten:

1. Umsetzung als reine Enterprise-Anwendung
2. Umsetzung als reine Content-Anwendung
3. Umsetzung als hybride Anwendung, bei der die Enterprise-Anwendung die Geschäftslogik implementiert und die Content-Anwendung den Content-Anteil der Daten verarbeitet.

Die Begründung für den Einsatz einer Hybrid-Lösung besteht in den Kosten, die auf beiden Seiten entstehen, um sowohl geschäftliche Anforderungen als auch Anforderungen an Content abzubilden. In Enterprise-Anwendungen würde dies zu Implementierungen verschiedener Use-Cases wie einer Volltextsuche über Binärdateien führen, die eigentlich von Content-Anwendungen im Standard erfüllt werden. Im Falle einer Hybridlösung entstehen per se keine Kosten durch Implementierung von Content-Use-Cases. Allerdings entstehen initialer Overhead sowie im Anschluss Kosten zur Re-Integration der Daten, um z.B. Daten aus beiden Systemen innerhalb der Enterprise-Anwendung behandeln zu können.

Abbildung 1 vergleicht die entstehenden Kosten bei beiden Lösungen. Dabei wird angenommen, dass tendenziell die Kosten zur Implementierung von Content-Use-Cases höher sind als Integrationskosten bei Hybridsystemen. Es zeigt ebenfalls, dass es einen, projektspezifischen, Break-Even-Point gibt, bei dem der Anteil von Anwendungsdaten und Content-

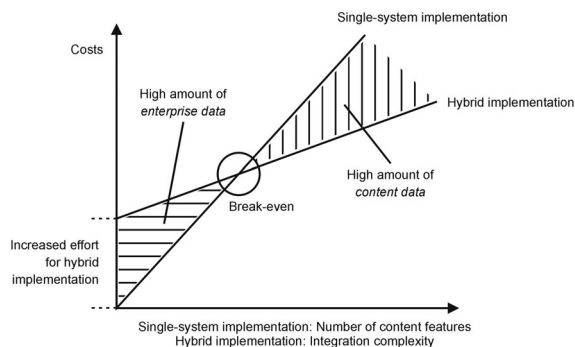


Abbildung 1: Kostenverhalten Unternehmensanwendung, Hybrid-Lösung

Daten, die erwartete Integrationskomplexität sowie die Anzahl verschiedener Content-Features einfließt.

Bei genauerer Betrachtung der beiden entstehenden Kosten wird klar, dass sich Teile der initialen Kosten einsparen lassen. Durch eine Reduzierung der technischen Komplexität lassen sich zum Beispiel Kosten reduzieren, die in direktem Zusammenhang mit der Implementierung stehen. Dagegen lassen sich Kosten nicht senken, die in direktem Zusammenhang mit der zweiten Umgebung stehen, z.B. Aufwand für Evaluierung und Einarbeitung oder auch Lizenzkosten. Auch die Folgekosten aufgrund hoher Integrationskomplexität lassen sich durch geeignete Integrationsmaßnahmen reduzieren. Und schließlich lässt sich mit einer Entkopplung des Konzepts der hybriden Persistenz die Komplexität des eigentlichen Softwareprojektes reduzieren. Dies wirkt sich vor allem positiv auf die Wartungskosten aus, die über den Lebenszyklus gemessen einen großen Teil der Gesamtkosten darstellen.

### 3 Implementierung einer Integrationslösung

Im zweiten Teil der Arbeit wird ein Framework implementiert, das eine generische Integration beider Datenmengen auf Anwendungsebene der Enterprise-Anwendung durchführt, das sog. „Total Integration“. Dadurch werden die Folgekosten der Hybrid-Lösung unabhängig von der Integrationskomplexität.

Die entwickelte Lösung baut auf Java EE 5 und dem Standard Java Content Repository auf. Content-Daten werden dabei in einer spezialisierten Content-Applikation abgelegt, wodurch Content-Use-Cases ohne weitere Kosten verwendet werden können. Allerdings werden sie über ein Mapping in ein POJO-Modell transformiert und steht so innerhalb der Enterprise-Anwendung zur Verfügung. Verknüpfungen mit geschäftlichen Entitäten kann damit auf Anwendungsebene realisiert werden. Schließlich werden alle Persistenzopera-

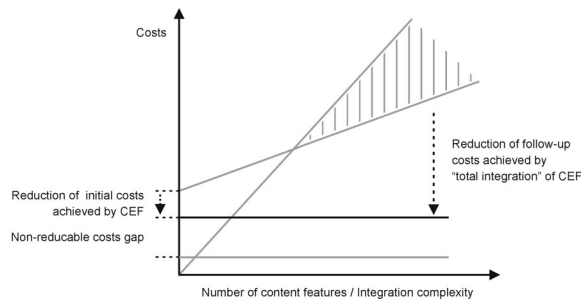


Abbildung 2: Erreichte Kostensenkung mit Hilfe der implementierten Integrationslösung

tionen auf den geschäftlichen Entitäten durch aspekt-orientiert eingewebten Aktivierungs-Code dupliziert und synchron auf den Content-Objekten ausgeführt. Daraus folgt, dass der Content stets gemeinsam mit den geschäftlichen Entitäten erzeugt, geladen, gespeichert und gelöscht wird, der entscheidende Punkt der „Total Integration“.

Diese Implementierung macht weitere manuelle Integrationsaufgaben obsolet, da stets eine integrierte Repräsentation von Anwendungsdaten und Content-Daten zur Verfügung steht. Damit wird die ursprünglich einheitliche Sicht auf fachliche Geschäftsobjekte wieder hergestellt, nachdem sie durch die Hybridlösung erst einmal getrennt war. Abbildung 2 veranschaulicht die Kostenreduzierung, die damit erreicht werden konnte.

## 4 Fazit

Die Arbeit soll in Zusammenhang mit dem Entscheidungsprozess sowie dem implementierten Integrationsframework das Bewusstsein steigern für die Rolle, die Daten und Datentypen in Anwendungen spielen. Entscheidungen, die bisher oft nach einem Bauchgefühl getroffen wurden, werden nun geleitet und sollen in einem möglichst frühen Projektstadium zu fundierten und dokumentierten Ergebnissen führen. Am Ende steht eine Lösung, die Projekte unterstützen soll in der Senkung von Risiken (architektonische Entscheidungen werden fundiert getroffen) und der Reduzierung von Kosten (initiale Kosten, Integrationskosten, Wartungskosten).

# **MoVIS – Ein mobiles, RFID-gestütztes Besucherinformationssystem für Museen**

David Ciechanowicz, Konrad Wolf

Westfälische Wilhelms-Universität Münster  
Lehrstuhl für Informatik  
david.ciechanowicz@gmx.de  
webwolf@freenet.de

Art der Arbeit: Projektseminar

Betreuer der Arbeit: Prof. Gottfried Vossen, Dipl.-Wirt.Inform. Joachim Schwieren

**Abstract:** *MoVIS* ist ein .NET basiertes, mobiles Besucherinformationssystem für Museen. Ein Museum wird damit in die Lage versetzt, den für die gezielte Präsentation der Exponate benötigten Content auf einfache Art selbst zu erstellen und zu verwalten. Dem Museumsbesucher wird eine PDA-basierte Lösung zur Verfügung gestellt, mit welcher durch Scannen von an den Exponaten angebrachten, passiven RFID-Tags kontextabhängige, Benutzer-adaptierte Informationen zu den Exponaten zugänglich gemacht werden.

## **1 Motivation**

Im Rahmen des Projektseminars *MoVIS* (Mobile Visitor Information System) wurde von einem 12-köpfigen Team ein auf RFID-Technologie basierender, mobiler Museumsguide für PDAs entwickelt [SV07]. Ziel des Projektes ist es, ein hochintuitives und vor allem wirtschaftliches System für Museumsbesucher und auch -betreiber zu entwerfen [RTA05]. Die Ergebnisse des Projektseminars werden im Folgenden anhand des Funktionsumfangs der Lösung, sowie der zugrundeliegenden technischen Architektur vorgestellt.

## **2 Charakterisierung des MoVIS-Systems**

### **2.1 Funktionsumfang**

*MoVIS* ermöglicht eine gezielte Bereitstellung von kontextabhängigen, Benutzer-adaptierten Informationen für unterschiedliche Gruppen und Typen von Museumsbesuchern. Abhängig von gewählter Sprache und Wissensstand erhält der Besucher individuelle und relevante Informationen zu einem gewünschten Exponat. Dazu muss lediglich der PDA mit dem integrierten 125 KHz RFID-Reader an dem



entsprechenden RFID-Tag des Exponats vorbeiführt werden [SV07]. Der Besucher hat dann die Wahl zwischen unterschiedlichen Medientypen wie einfachen Texten (ggf. mit Audio hinterlegt), Videos/Diashows, Bildern oder auch interaktiven Quizfragen. Die einzelnen Exponatinformationen sind inhaltlich miteinander verknüpft. Das bedeutet, dass zu einem Exponat thematisch verwandte Exponate oder auch übergeordnete Themen aufgelistet werden können. Texte können Hyperlinks enthalten, die entweder auf ein anderes Exponat verweisen, einen Begriff näher erläutern oder aber den bestehenden Text um weiterführende Informationen ergänzen [RC03] [CR96]. Der Umfang der Informationen ist demnach äußerst flexibel und abhängig von verfügbarer Zeit und Interesse des Besuchers. Neben dem freien, *explorativen Modus* bietet sich dem Besucher außerdem die Möglichkeit, eine vorgefertigte *Tour* auszuwählen, um sich themenorientiert vom System durch das Museum führen zu lassen [CLL04].

Der Museumsbetreiber verwendet zur Bereitstellung und Pflege der benötigten Inhalte ein benutzerfreundliches *Content Management System (CMS-System)*, welches in Form eines .NET Rich Clients realisiert wurde und über Remoting auf den zentralen Datenbankserver zugreift. Die Exponatinformationen können dabei mit unterschiedlichen Medientypen realisiert werden. Auch gesteuerte Touren können flexibel mit dem System modelliert werden. Neben Bereitstellungsfunktionen stehen weiterhin Excel-basierte Auswertungsmöglichkeiten von während des Betriebs anonym gesammelten Tracking-Daten zur Verfügung [BE87] [KL93].

2.1 Architektur

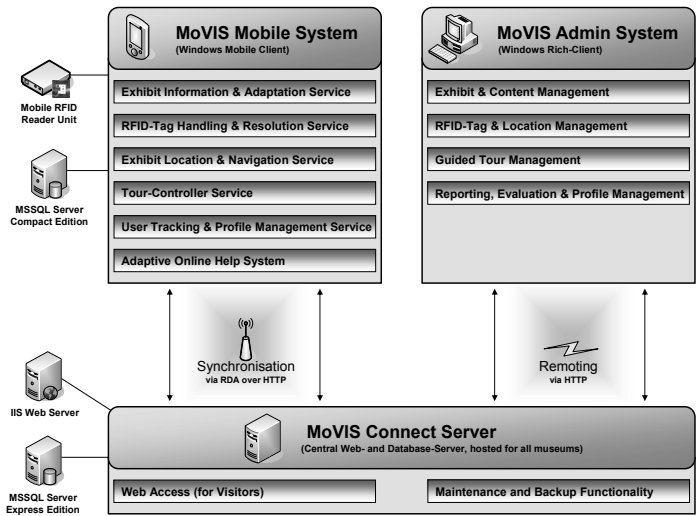


Abbildung 1: Drei-Komponenten-Architektur von MoVIS

Die MoVIS-Architektur weist die drei Kernkomponenten *Mobile Client*, *Content Management System (CMS-System)* und *Connect Server* auf. Der *Connect Server* stellt neben einem MS SQL-Server 2005, welcher als zentraler Speicher von Content und

Datenbanken eines jeden Mandaten (Museum) fungiert, auch die Plattform für einen IIS Web-Server dar. Letzterer bietet dem *Mobile Client* die Möglichkeit zur Synchronisation via RDA über HTTP [WW03]. Im Gegensatz zu anderen Systemen wird bewusst auf die Anforderung einer Always-On WLAN-Verbindung verzichtet, da alle Daten lokal auf dem PDA gespeichert werden. Dies bietet mehr Stabilität und einen deutlichen Kostenvorteil, da keine teure Infrastruktur für den Betrieb erforderlich ist. Das *CMS-System* greift via .NET-Remoting über HTTP auf den *Connect Server* zu [Ma03]. Die Contenterstellung, der -upload und die mittels des objektrelationalen Mappers ODX realisierte Datenbankverwaltung erfolgt über das unter .NET 2.0 entwickelte *CMS-System* [ODX].

Die Datenverwaltung des unter dem .NET Compact Framework 2.0 entwickelten *Mobile Client* basiert auf einem MS SQL-Server 2005 Compact Edition. Datenbankabfragen sind mittels direkter SQL-Abfragen realisiert, da Mapper wie ODX hier aus Performance- und Komplexitätsgründen unzweckmäßig wären. Für das Auslesen von auf RFID-Tags gespeicherten Informationen wird das ID-Services.NET-Framework benutzt, welches in erster Linie eine abstrakte Zugriffsmethode auf die verwendete RFID-Hardware bietet. Die als User-Controls realisierten Komponenten des *Mobile Client* dienen der Kapselung häufig benötigter Funktionen und gewährleisten eine hohe Wiederverwendbarkeit. Für nicht vom .NET Compact Framework angebotene Funktionen wird neben P/Invoke als Zugriffstechnik auf die unter Windows Mobile / CE vorhandene API auch auf COM-Komponenten wie den Windows Media Player oder das HTMLViewer-Control zurückgegriffen.



Abbildung 2: MoVIS PDA, RFID-Tag und RFID-Reader

### 3 Fazit

MoVIS stellt mit seiner Drei-Komponenten-Architektur ein besonders einfach zu handhabendes mobiles Besucherinformationssystem für Museen dar. Der *Connect Server* übernimmt die Aufgabe der zentralen Datenhaltung während das *CMS-System* als Anwendung zur Content-Erstellung dient. Den Kernbereich der Entwicklung stellt der *Mobile Client* dar, mit welchem die Museumsbesucher interagieren. Durch eine Erweiterung der Funktionalität in Richtung eines Besucherinternetportals zur Vor- und Nachbereitung eines Museumsbesuches soll das Interesse für den Einsatz von MoVIS in der Praxis weiter gesteigert werden. Eine Exponat-zu-Exponat Navigation würde darüber hinaus die Orientierung in größeren Museen vereinfachen. Nach ersten Tests im Rahmen des Kultur-Events „Nacht der Museen 2007“ in Münster wird derzeit ein umfangreicher Piloteinsatz vorbereitet, bei dem das System unter realistischen Bedingungen erprobt werden soll.

## Literaturverzeichnis

- [Be87] Beer, V.: Great expectations: Do museums know what visitors are doing? In *Curator* 30, 3 (1987), S. 206-215.
- [CLL04] Chou, L.; Lee, C.; Lee, M.; Chang, C.: A Tour Guide System for Mobile Learning in Museums. In *WMTE '04: Proc. 2nd IEEE international Workshop on Wireless and Mobile Technologies in Education*, IEEE Computer Society, Washington DC, 2004, S. 195.
- [CR96] Chen, C.; Rada, R: Interacting with Hypertext: A Meta-Analysis of Experimental Studies. *Human-Computer Interaction*, 1996, v. 11, S. 125-156.
- [KI93] Klein, H.-J.: Tracking Visitor Circulation in Museum Settings. In *Environment and Behavior*, 25 (1993), S.782-800.
- [Ma03] MacDonald, M.: Microsoft .NET Distributed Applications. Integrating XML Web Services and .NET Remoting, Washington, 2003.
- [ODX] ODX.NET - Distributed ORM Framework; available on <http://www.codeplex.com/odx/Wiki/View.aspx?title=Documentation> (10. Januar 2008)
- [RC03] Romero, L.; Correia, N.: HyperReal: a hypermedia model for mixed reality. In *HYPERTEXT '03: Proc. 14th ACM Conf. on Hypertext and Hypermedia*, ACM Press, New York 2003, S. 2-9.
- [RTA05] Raptis, D.; Tselios, N.; Avouris, N.: Context-based design of mobile applications for museums: a survey of existing practices. In *MobileHCI '05: Proc. 7th Int. Conf. on Human Computer Interaction with Mobile Devices & Services*, ACM Press, New York 2005, S. 153-160.
- [SV07] Schwieren, J.; Vossen, G.: Implementing Physical Hyperlinks for Mobile Applications Using RFID Tags. In *IDEAS '07: Proc. 11th International Database Engineering and Applications Symposium*, IEEE Computer Society, Los Alamitos 2007, S. 154-163.
- [WW03] Wigley, A.; Wheelwright, S.: Microsoft .NET Compact Framework Core Reference, Washington, 2003.

# Entwicklung eines hochverfügbaren Backup Storage Area Network

Paul Enns

Private FernFachhochschule Darmstadt

Paul.Enns@t-systems.com

Art der Arbeit: Diplomarbeit  
Betreuer/in der Arbeit: Prof. Dr. Detlev Zimmermann

**Abstract:** Die vorliegende Arbeit konzentriert sich auf den Bereich der SAN (Storage Area Network) Technologie. Herkömmlicherweise werden Disk- und Bandspeicher über eine gemeinsame SAN-Infrastruktur gesteuert und genutzt. Für den Kunden eines führenden IT-Dienstleisters kommt es hierbei, trotz redundanter SAN-Infrastruktur, häufig zu unternehmenskritischen Störungen. Die Ursachen der genannten Probleme kann man generell in zwei Bereiche unterteilen: Unterschiede in den Small-Computer-System-Interface(SCSI)-Zugriffsmethoden bei Disk- und Bandlaufwerkstechnologien, eingesetzte SAN-Technologie. In dieser Arbeit wird ein Speichernetz erarbeitet, das die bestehenden Schwachstellen im Bereich der Datensicherung beseitigt, zusätzliche Stabilität sowie diverse neue Funktionalitäten für die Systemlandschaft schafft und damit dem betroffenen Kunden einen hohen Nutzwert bietet. Dieses Verfahren ist besonders, da es zum ersten Mal zeigt, wie Speichernetze mehrerer weit entfernter Standorte über das hochleistungsfähige Fibre-Channel-Protokoll gekoppelt werden und so eine effiziente Datensicherung über weite Entfernungen auf Bandspeicher ermöglicht.

## 1. Motivation

Während Festplatten in den meisten Fällen die Eigenschaft haben, ihre Daten über einen wahlfreien Zugriff (Random access) anzusprechen, werden Magnetbänder über den sequentiellen Zugriff (Sequential access) beschrieben bzw. gelesen.<sup>1</sup>

In der aktuellen SAN-Architektur werden die Disksubsysteme und Bandlaufwerke überwiegend über denselben Host Bus Adapter (HBA) angesteuert.

Das Ansprechen der verschiedenen Speichertechnologien über denselben HBA verursacht oft folgende Schwierigkeiten:

---

<sup>1</sup> Vgl.: <http://www.t10.org/scsi-3.htm>, 06.07.2007

Beim Auftreten von Lese- oder Schreibproblemen auf dem externen Festplattenspeicher eines Computersystems sendet der HBA ein SCSI-Reset-Kommando an die Logical Unit Number (LUN) des betroffenen Datenspeichers. Wegen der von Festplatten- und Bandlaufwerken gemeinsam genutzten Ressource HBA wird der entsprechende SCSI-Reset an alle für den HBA sichtbaren Ziele versendet. Während Festplatten auf Grund ihres wahlfreien Zugriffs den Lese- bzw. Schreibvorgang nach einem SCSI-Reset einfach wiederholen können, führt das entsprechende Ereignis bei Bandlaufwerken zum Abbruch des sequentiellen Datenstroms und zum Zurückspulen des Magnetbandes. Infolgedessen kann es passieren, dass beim nächsten Laden desselben Magnetbandes die ursprünglichen Daten überschrieben werden.

Es scheint im ersten Moment, als könnte das genannte SCSI-Reset-Problem durch den einfachen Einbau von zusätzlichen, speziell für Datensicherungszwecke genutzten HBAs gelöst werden. Die Praxis aber hat gezeigt, dass auch bei der Nutzung dedizierter HBAs wegen der immer noch gemeinsamen Ressource SAN-Switch ähnliche Notlagen entstehen.

Ein weiteres Problem entsteht bereits, wenn eine einzelne SAN-Komponente wie ein SAN-Switch oder ein SAN-Router einen technischen Defekt hat und so genannte Panic-Meldungen erzeugt. Diese Meldungen werden an das gesamte physikalische SAN gestreut und verursachen letzten Endes Unterbrechungen der Datenströme oder gar einen Neustart mancher SAN-Komponenten. Infolgedessen verlieren die HBAs die Verbindung zu den Tape-Libraries, woraufhin keine Datensicherung mehr möglich ist.

## **2. Bewertung von Alternativen**

Für eine rationale Entscheidungsfindung ist es notwendig, alle relevanten Speichertechnologien auf dem Markt zu erfassen und eine Bewertung sowie einen Vergleich dieser Technologien bis auf die Protokollebene durchzuführen. Die Abbildung 1 veranschaulicht die bestehenden Unterschiede bezüglich ihrer Schnittstellen zum Netzwerk sowie ihrer Zugriffsverfahren auf die Daten. Dabei ist zu erkennen, dass das SCSI over IP(iSCSI)-SAN und Fibre Channel(FC)-SAN nach dem ähnlichen Prinzip gestaltet sind. Bei beiden befindet sich zwischen der Anwendung und der physikalischen Platte ein Netzwerk, über das ein blockorientierter Datenaustausch und damit eine direkte Datensicherung zwischen Server und Tape-Library möglich sind.

Durch die Vielzahl der am Markt verfügbaren Lösungen, die sich sowohl in Qualität als auch im Preis stark unterscheiden, ist die Entscheidungsfindung für ein Unternehmen nicht einfach. Daher wird zunächst ein Instrument erarbeitet, welches es gestattet, Alternativen möglichst an unternehmens- bzw. dienstleistungsspezifischen Parametern orientiert zu evaluieren und auszuwählen. Mit Hilfe dieses Instrumentes wird herausgefunden, welche Konstellation von Speichernetz und Datensicherungssystem die optimale Lösung für diese Datensicherungsumgebung darstellt.

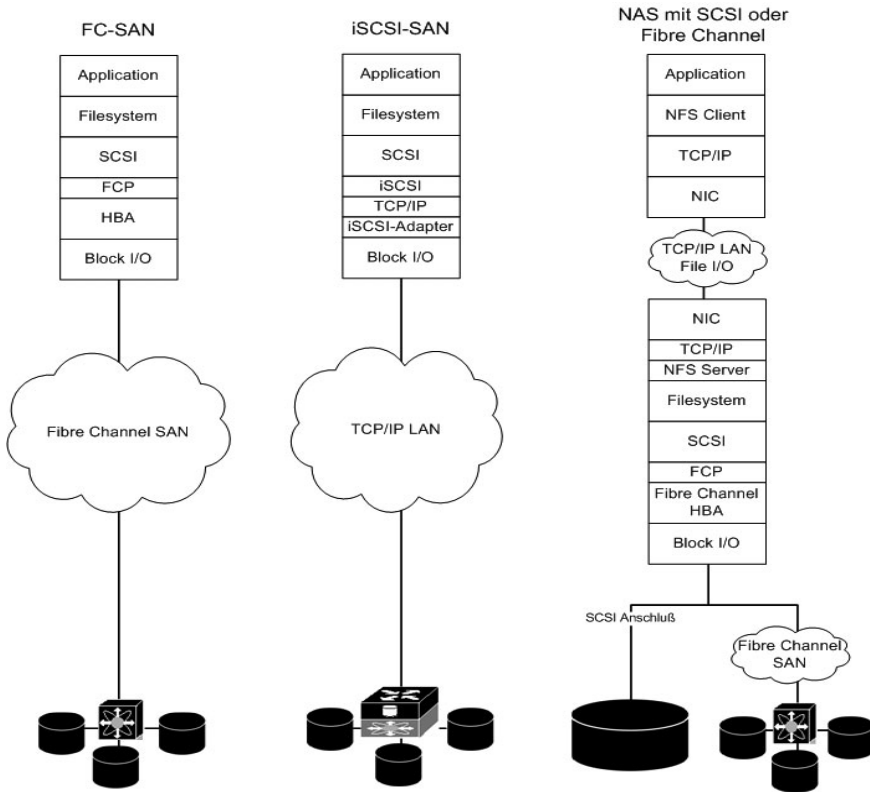


Abbildung 1: Auszug aus der Gegenüberstellung von Speichernetzen

### 3. Lösungsansatz und Implementierung

Unter Berücksichtigung eines Anforderungskataloges wird zwischen zwei Rechenzentrenstandorten, die über 110 km voneinander entfernt liegen, ein speziell für Datensicherungszwecke angepasstes FC-SAN konzipiert und umgesetzt (Abbildung 2). Das Backup-SAN ist hochverfügbar ausgelegt, so dass der Ausfall einer SAN-Komponente keine oder eine nur geringe Auswirkung auf die Verfügbarkeit der Datensicherung hat.

Mit Hilfe der zusätzlichen VSAN(Virtuelle SAN)-Technologie und der Inter-VSAN-Routing(IVR)<sup>2</sup> Funktionalitäten des Herstellers CISCO werden innerhalb eines physikalischen SAN mehrere logische SAN-Segmente abgebildet, um eventuelle Störungen auf ein Segment einzugrenzen und um nicht das gesamte Backup-SAN zu gefährden.

<sup>2</sup> Vgl.: <http://www.cisco.com/en/US/products/hw/ps4159/index.html>, 16.08.2007

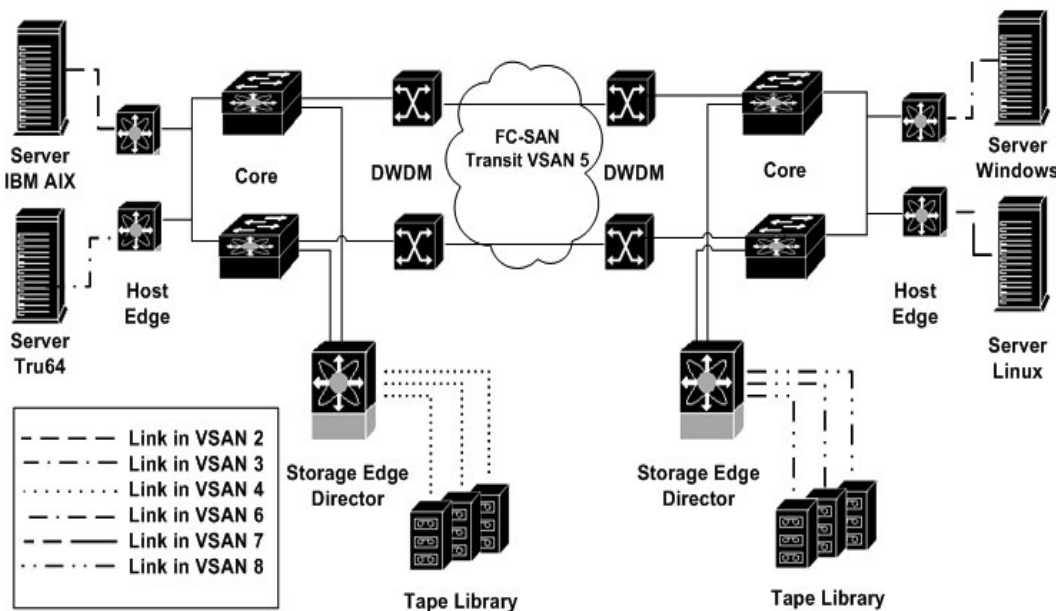


Abbildung 2: Schematische Darstellung des Backup Storage-Area-Network

## 4. Zusammenfassung und Ausblick

In dieser Arbeit wird eine Lösung entwickelt, welche durch die Separierung der Datenströme der Diskspeicher von dem der Bandspeicher eine höhere Stabilität der gesamten Datensicherung ermöglicht. Außerdem wird veranschaulicht, wie durch den Einsatz von innovativen Technologien standortübergreifende und hochleistungsfähige Backupverfahren erzielt werden können. Darüber hinaus wird gezeigt, wie Entscheidungen in einem Unternehmen unter Einsatz von Methoden zur Entscheidungsfindung effektiver getroffen werden können.

Mit diesem Ergebnis ist es nun möglich, neue Strategien im Bereich Datensicherung sowie Disaster-Recovery umzusetzen und die damit entstandenen Synergieeffekte zu nutzen.

## Literaturverzeichnis

- Judd Josh, Krueger Dan: *Principles of SAN Design*, West Conshohocken/Penn. 2005/2006  
 Kembel Robert: *Fibre Channel A Comprehensive Introduction*, Tucson/Calif. 2003  
 Troppens Ulf, Erkens Rainer: *Speichernetze*, Heidelberg (dpunkt) 2003  
<http://www.t10.org/scsi-3.htm>, Abruf 06.07.2007  
<http://www.t11.org/index.html>, Abruf 08.07.2007  
<http://www.cisco.com/en/US/products/hw/ps4159/index.html>, Abruf 16.08.2007

# Transformation von EPK nach BPEL

Marcel Matzat, Danilo Schmiedel

Hochschule für Technik, Wirtschaft und Kultur - Leipzig  
Fachbereich Informatik, Mathematik und Naturwissenschaften  
mmatzat@imn.htwk-leipzig.de  
dschmied@imn.htwk-leipzig.de

**Art der Arbeit:** Masterprojekt  
**Betreuer der Arbeit:** Prof. Dr.-Ing. T. Kudraß

**Abstract:** Um der Forderung nach Agilität in einem sich ständig ändernden Marktumfeld gerecht zu werden, fühlen sich viele Unternehmen dazu veranlasst, die bestehenden IT-Systeme zu restrukturieren. Während herkömmliche Anwendungsarchitekturen eher technisch ausgerichtet sind, stellt der Ansatz der Serviceorientierten Architekturen (SOA) die fachlichen Aspekte in den Vordergrund. Als Ausdrucksmittel existieren hierfür semi-formale Sprachen, wie etwa die *Ereignisgesteuerten Prozessketten* (EPK). Da diese plattformunabhängigen Modelle zu abstrakt für die technische Ausführung auf einer konkreten Plattform sind, müssen sie in plattformspezifische Prozesse überführt werden. Eine bekannte Sprache zur Beschreibung ausführbarer Geschäftsprozesse stellt die *Business Process Execution Language* (BPEL) dar. Beide Beschreibungsformen befinden sich auf unterschiedlichen Abstraktionsebenen, wobei eine Überführung der einen in die andere Sicht möglichst automatisiert erfolgen soll. Die vorliegende Arbeit liefert einen Überblick in die Transformationsproblematik und stellt anhand einer konkreten Transformationsmethode die Schwierigkeit des Vorhabens dar. Der Fokus der Betrachtungen ist auf die Frage gerichtet, wie eine durchgängige Methode zur Modellierung fachlicher Geschäftsprozesse und deren Ausführung auf einer technischen Plattform aussehen kann.

## 1 Einleitung

Mit der Globalisierung der Märkte und dem resultierenden Konkurrenzdruck sind Unternehmen dazu gezwungen, ihre Geschäftsprozesse effizienter zu gestalten. Monolithische IT-Systeme werden aufgrund mangelnder Flexibilität von Serviceorientierten Architekturen (SOA) verdrängt. Über plattformunabhängige Schnittstellen werden hierbei einzelne Funktionalitäten (Services) zu einem durchgängigen Geschäftsprozess lose miteinander gekoppelt. SOA verkörpert dabei keine Technik, sondern lediglich die Idee, ein breit angelegtes Rahmenwerk zur Erstellung, Verwaltung und Kombination von Softwarediensten zu bilden. Die jeweiligen Services stellen separate Geschäftsfunktionen dar, die über standardisierte und beschriebene Schnittstellen aufgerufen und bei Bedarf an neue Geschäftsabläufe angepasst werden können. Bei der Änderung eines fachlichen Prozesses ist lediglich die erneute Abstimmung der Serviceabfolge notwendig. Demnach ste-



hen bei einer SOA vielmehr die Geschäftsprozesse als die einzelnen Applikationen im Vordergrund. Durch die enge Integration der fachlichen Prozesse mit den technischen Services, soll die flexible und schnelle Anpassung der IT an die sich ständig ändernden Geschäftsprozesse der Organisation gewährleistet werden.

## 2 Grundlagen

Hinsichtlich der Modellierung von fachlichen und technischen Modellen existiert ein breites Angebot an Softwarelösungen, welche unterschiedliche Ansätze zur Unterstützung des Entwicklungsprozesses verfolgen. Neben rein grafisch orientierten Werkzeugen kommen außerdem formale Verfahren zum Einsatz, mit denen verschiedene Informationen im Zusammenhang mit den Prozessmodellen abgebildet werden können. Darunter zählen Petri-Netze, objektorientierte UML-Modelle und spezifische Ansätze der Hersteller von BPM-Werkzeugen, wie beispielsweise die ARIS<sup>1</sup>-Methode [Sch96], [Sch99], [IS07a]. Mit dem ARIS-House of Business Engineering steht ein Konzept zur Verfügung, welches die Einordnung verschiedener Systemkomponenten bzgl. Rolle und Beziehung in ein ganzheitliches Geschäftsprozessmanagement erlaubt. Diese Methode vereinigt eine Vielzahl von Diagrammtypen für die fachliche Modellierung, wobei der Entwurf, die Steuerung und die Ausführung von Geschäftsprozessen auf vier Ebenen strukturiert werden. Im Mittelpunkt des ARIS-Ansatzes steht die Darstellung von Geschäftsprozessen durch die *Ereignisgesteuerten Prozessketten* (EPK). Sie beschreiben den Ablauf eines Prozesses durch die Verbindung von Funktionen, Ereignissen und Operatoren. Des Weiteren erlaubt die Notation die Zuordnung von Rollen und Systemen zu Funktionen. Der Kontrollfluss wird in EPKs mit Hilfe von Verbindungen zwischen den Ereignissen und Funktionen sowie durch festgelegte Regeln abgebildet. Diese bewirken die Verzweigung oder Zusammenführung von Abläufen.

Die EPKs orientieren sich an der fachlichen Problemstellung und können aufgrund fehlender technischer Ausdrucksmöglichkeiten nicht direkt auf einer Plattform ausgeführt werden. Deshalb muss an dieser Stelle eine Überführung in eine technische Beschreibungssprache erfolgen. Eine für diesen Zweck entwickelte Sprache, die von den großen Herstellern wie Oracle, SAP und IBM unterstützt wird, ist die *Business Process Execution Language* [ACD<sup>+</sup>03]. Sie ist eine Kombination aus graph- und block-orientierten Bestandteilen und beschreibt im Wesentlichen das Verhalten ausführbarer Geschäftsprozesse.

## 3 Transformationsmethode

Unter Transformation versteht man im Allgemeinen die Umformung bzw. Umwandlung eines Modells auf Basis einer gegebenen Quellsprache in ein Modell einer Zielsprache. Abhängig vom Abstraktionsgrad beider Sprachen, kann zwischen verschiedenen Einsatzszenarien unterschieden werden (vgl. [Vis03]). Das für die Überführung von EPK nach

---

<sup>1</sup> ARIS - Architektur integrierter Informationssysteme [Sch99]

BPEL relevante Szenario nennt sich *Synthese* und beinhaltet die Transformation, bei welcher der Abstraktionsgrad der Quellsprache höher ist als der der Zielsprache.

### 3.1 Workflow Pattern

Wie bereits erwähnt, stellt die unterschiedliche Semantik von EPK und BPEL eine Schwierigkeit bei der Überführung dar. So gilt beispielsweise für Schleifen in BPEL die Einschränkung, dass diese nur über genau einen Eintritts- und Austritts-Punkt verfügen dürfen. In EPK lassen sich hingegen Modelle realisieren, für welche derartige Beschränkungen nicht gelten. Beide Beschreibungsformen unterstützen zwar die Grundkonzepte, wie Sequenzen, Iterationen, Teilungen (AND / OR) und Vereinigungen (AND / OR), allerdings ist deren Interpretation nicht einheitlich. Daher ist es besonders schwierig, daraus die Unterstützung komplexerer Konstrukte abzuleiten.

Die Voraussetzung für eine Reihe von Transformationsansätzen (z.B.: [Föt05], [Sch07], [IS07b]) bildeten die so genannten *Workflow Patterns* nach van der Aalst und Hofstede [vdAtHB03], welche im Rahmen des Forschungsprojektes *Workflow Patterns Initiative* [Ini07] entstanden sind. Vereinfacht formuliert stellen sie typische Konstrukte dar, die in Workflow-Prozessen vorzufinden sind (z.B.: Sequenz, parallele Ausführung, ...). Zwar werden in den beiden Sprachen nicht alle Pattern unterstützt, die Schnittmenge erlaubt es jedoch die Transformation des Kontrollflusses nach BPEL sicherzustellen. Da für die einzelnen EPK-Elemente keine explizite Entsprechung in BPEL existiert, können die Konstrukte der Quellsprache nicht direkt in Konstrukte der Zielsprache überführt werden. Somit lassen sich in EPK modellierte Abläufe gelegentlich nur durch komplexe BPEL-Aktivitäten ausdrücken.

### 3.2 Top-down-Ansatz

Serviceorientierte Architekturen werden durch wirtschaftliche und technische Merkmale charakterisiert. Da der Mensch und seine Arbeitsprozesse im Vordergrund stehen, sollte die Entwicklung einer SOA in einem *top-down*-Ansatz - also ausgehend von der fachlichen Sicht - erfolgen. Somit ist gewährleistet, dass die Geschäftsprozesse die IT leiten und nicht umgekehrt.

Wie in Abbildung 1 dargestellt, bildet die in unstrukturierter Form vorliegende **Betriebswirtschaftliche Problemstellung** den Ausgangspunkt. In einem aufwändigen Analyseprozess werden daraus die fachlichen Geschäftsmodelle abgeleitet, welche als Grundlage für die spätere **Validierung** und die **Transformation** dienen. Für die Erzeugung eines ausführbaren Prozesses ist das fachliche Modell daraufhin mit technischen Details anzureichern. Hierbei müssen die zu automatisierenden Funktionen im Fachkonzept identifiziert und anschließend den entsprechenden Web-Services zugeordnet werden. Die Übertragung des Datenflusses erfolgt durch zusätzlich eingebrachte Datenbeschreibungselemente (*Datencluster*). Aus früheren Transformationsprojekten (z.B.: [Föt05], [Sch07]) ist bereits be-

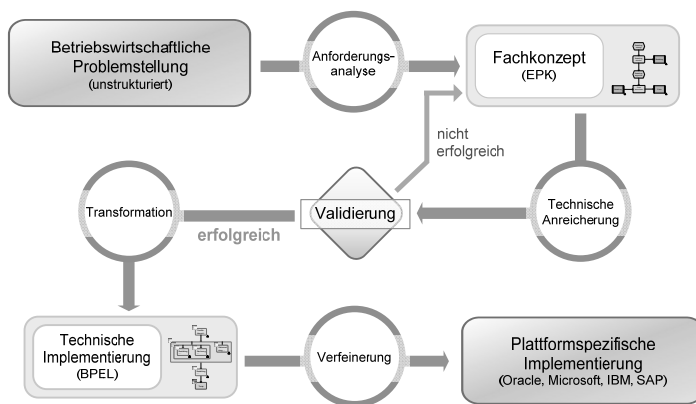


Abbildung 1: Schrittfolge des *top-down*-Transformationsansatzes

kannt, dass nicht alle in der Realität existierenden EPK-Modelle überführbar sind. Daher müssen die strukturierten, fachlichen Spezifikationen vor der Überführung auf die technische Implementierungsebene eingehend untersucht und ggf. überarbeitet werden. Dieser Validierungsschritt kann sowohl manuell als auch automatisiert erfolgen. Die **Transformation** erzeugt die technische Implementierung in Form eines BPEL-Modells. Im letzten Schritt muss dieses für die jeweilige Ausführungsplattform mit technischen Informationen (z.B.: Datenbankadapter, Mail-Funktionalität) verfeinert werden.

Wesentlicher Transformationsanspruch ist die Schließung der Lücke zwischen Fachabteilung und IT, welche sich aus den unterschiedlichen Beschreibungssprachen ergibt. Die einzelnen Modelle ermöglichen es, verschiedene Aspekte eines Systems aus diversen Blickwinkeln zu beschreiben. Um eine Konsistenz zwischen den Spezifikationen zu erhalten, müssen die jeweiligen Sichten miteinander gekoppelt werden. Hierfür existieren verschiedene Ansätze, wobei automatisierte Überführungen gegenüber manuellen Herangehensweisen kürzere Entwicklungszeiten und eine geringere Fehleranfälligkeit versprechen.

Als entscheidende Anforderungen gelten die Erhaltung der Prozessstruktur, der korrekte Aufruf aller assoziierten Services, die Übertragung des Datenflusses sowie die Generierung einer Interface-Beschreibung für den ausführbaren Geschäftsprozess. Die Vielfalt an unterschiedlichen Ausführungsplattformen untermauert zudem die Forderung nach Standardkonformität für die generierten BPEL-Modelle. Um bereits eingebrachte Änderungen nicht zu verlieren, ist bei der wiederholten Überführung ein Zusammenfügen der verschiedenen Modellversionen (*Merge*) notwendig. Durch die Bereitstellung einer durchgängigen Transformationsmethode bleiben vorhandene Prozessmodelle weiterhin nutzbar und es müssen keine zusätzlichen Experten für eine neue Technologie ausgebildet werden.

Aus der Beschreibung des *top-down*-Ansatzes wird ersichtlich, dass lediglich ein teilautomatisierter Übergang von den betrieblichen Geschäftsmodellen auf die Implementierungsebene möglich ist. Der Forderung bzgl. der Ausführbarkeit des Transformationsergebnisses konnte zwar entsprochen werden, allerdings setzt das Konzept manuelle Vor-

bzw. Nachbearbeitungen voraus. Zu den manuell durchzuführenden Maßnahmen zählen die bereits angesprochenen Service- und Datencluster-Anreicherungen im fachlichen Modell sowie die abschließende technische Verfeinerung durch plattformspezifische Implementierungsdetails. Eine vollautomatisierte Transformation würde bedeuten, dass die EPK zu den Daten- und Service-Informationen mit weiteren technischen Informationen angereichert werden müsste. Neben der Zuordnung aller Ein- und Ausgabedaten wären in diesem Zusammenhang die Bestimmung komplexer Datentypen sowie das Einbringen von Exception-Handleern und Datenmanipulationen erforderlich. Nach Meinung der Autoren sollten diese Tätigkeiten jedoch die Aufgabe der IT-Entwickler bleiben, da derartige Informationen lediglich auf der Implementierungsebene relevant sind.

## Literatur

- [ACD<sup>+</sup>03] T. Andrews, F. Curbera, H. Dholakia, Y. Goland und J. Kleinand F. Leymann. Business Process Execution Language for Web Services, Version 1.1. <ftp://www6.software.ibm.com/software/developer/library/ws-bpel11.pdf>, 2003. Letzter Zugriff: 29.09.2007.
- [Föt05] D. Fötsch. Eine XML-basierte Transformationsmethode für modellgetriebene Web-Service-Architekturen: Vom ARIS-IE-Geschäftsmodell zum BPEL4WS-Prozess. *Leipziger Beiträge zur Informatik: Band III*, 2005. S. 107.
- [Ini07] Workflow Patterns Initiative. Workflow Patterns. <http://www.workflowpatterns.com>, 2007. Letzter Zugriff: 26.08.2007.
- [IS07a] IDS-Scheer. ARIS Solution for Business-Driven SOA Management. [http://www.ids-scheer.de/de/Loesungen/ARIS\\_Loesungen/Business-Driven\\_SOA\\_Management/7808.html](http://www.ids-scheer.de/de/Loesungen/ARIS_Loesungen/Business-Driven_SOA_Management/7808.html), 2007. Letzter Zugriff: 15.08.2007.
- [IS07b] IDS-Scheer. *Handbuch ARIS SOA Architect 7.0*, version 1.0. Auflage, Januar 2007.
- [Sch96] A.-W. Scheer. ARIS-House of Business Engineering. <http://www.iwi.uni-sb.de/Download/iwihefte/heft133.pdf>, September 1996. Letzter Zugriff: 11.01.2008.
- [Sch99] A.-W. Scheer. *ARIS - Vom Geschäftsprozeß zum Anwendungssystem*. Springer-Verlag GmbH, 1999. 3. Auflage.
- [Sch07] Oleg Schmelzle. Transformation von annotierten Geschäftsprozessen nach BPEL. Diplomarbeit, Gottfried Wilhelm Leibniz Universität Hannover, [http://www.se.uni-hannover.de/documents/studthesis/MSc/Oleg\\_Schmelzle-Transformation\\_von\\_annotierten\\_Geschaeftsprozessen\\_nach\\_BPEL.pdf](http://www.se.uni-hannover.de/documents/studthesis/MSc/Oleg_Schmelzle-Transformation_von_annotierten_Geschaeftsprozessen_nach_BPEL.pdf), Mai 2007. Letzter Zugriff: 20.11.2007.
- [vdAtHB03] W.M.P. van der Aalst, A.H.M. ter Hofstede und B. Kiepuszewskiand A.P. Barros. Workflow Patterns. *Distributed and Parallel Databases*, 14(3):5–51, Juli 2003. Letzter Zugriff: 26.08.2007.
- [Vis03] Eelco Visser. A Survey of Rewriting Strategies in Program Transformation Systems. <http://people.cs.uu.nl/visser/ftp/Vis03-survey.pdf>, 2003. Letzter Zugriff: 01.08.2007.



## ***GI-Edition Lecture Notes in Informatics – Proceedings***

- 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 Brömmel, Christoph Busch (Eds.): BIOSIG 2003: Biometric and Electronic Signatures

- P-32 Peter Hubwieser (Hrsg.): Informatische Fachkonzepte im Unterricht – INFOS 2003
- P-33 Andreas Geyer-Schulz, Alfred Taudes (Hrsg.): Informationswirtschaft: Ein Sektor mit Zukunft
- P-34 Klaus Dittrich, Wolfgang König, Andreas Oberweis, Kai Rannenberg, Wolfgang Wahlster (Hrsg.): Informatik 2003 – Innovative Informatikanwendungen (Band 1)
- P-35 Klaus Dittrich, Wolfgang König, Andreas Oberweis, Kai Rannenberg, Wolfgang Wahlster (Hrsg.): Informatik 2003 – Innovative Informatikanwendungen (Band 2)
- P-36 Rüdiger Grimm, Hubert B. Keller, Kai Rannenberg (Hrsg.): Informatik 2003 – Mit Sicherheit Informatik
- P-37 Arndt Bode, Jörg Desel, Sabine Rathmayer, Martin Wessner (Hrsg.): DeLFI 2003: e-Learning Fachtagung Informatik
- P-38 E.J. Sinz, M. Plaha, P. Neckel (Hrsg.): Modellierung betrieblicher Informationssysteme – MobIS 2003
- P-39 Jens Nedon, Sandra Frings, Oliver Göbel (Hrsg.): IT-Incident Management & IT-Forensics – IMF 2003
- P-40 Michael Rebstock (Hrsg.): Modellierung betrieblicher Informationssysteme – MobIS 2004
- P-41 Uwe Brinkschulte, Jürgen Becker, Dietmar Fey, Karl-Erwin Großpietsch, Christian Hochberger, Erik Maehle, Thomas Runkler (Edts.): ARCS 2004 – Organic and Pervasive Computing
- P-42 Key Pousttchi, Klaus Turowski (Hrsg.): Mobile Economy – Transaktionen und Prozesse, Anwendungen und Dienste
- P-43 Birgitta König-Ries, Michael Klein, Philipp Obreiter (Hrsg.): 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 and 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 Ranneberg, 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, Wolffried Stucky (Hrsg.): Datenbanksysteme in Business, Technologie und Web
- P-66 Jörg M. Haake, Ulrike Lucke, Djamshid Tavangarian (Hrsg.): DeLFI 2005: 3. deutsche e-Learning Fachtagung Informatik
- P-67 Armin B. Cremers, Rainer Manthey, Peter Martini, Volker Steinhage (Hrsg.): INFORMATIK 2005 – Informatik LIVE (Band 1)
- P-68 Armin B. Cremers, Rainer Manthey, Peter Martini, Volker Steinhage (Hrsg.): INFORMATIK 2005 – Informatik LIVE (Band 2)
- P-69 Robert Hirschfeld, Ryszard 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 Fährnrich, 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 Poustchi, 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 Mangament & IT-Forensics – IMF 2006



- P-98 Hans Brandt-Pook, Werner Simonsmeier und Thorsten Spitta (Hrsg.): Beratung in der Softwareentwicklung – Modelle, Methoden, Best Practices
- P-99 Andreas Schwill, Carsten Schulte, Marco Thomas (Hrsg.): Didaktik der Informatik
- P-100 Peter Forbrig, Günter Siegel, Markus Schneider (Hrsg.): HDI 2006: Hochschuldidaktik der Informatik
- P-101 Stefan Böttinger, Ludwig Theuvsen, Susanne Rank, Marlies Morgenstern (Hrsg.): Agrarinformatik im Spannungsfeld zwischen Regionalisierung und globalen Wertschöpfungsketten
- P-102 Otto Spaniol (Eds.): Mobile Services and Personalized Environments
- P-103 Alfons Kemper, Harald Schöning, Thomas Rose, Matthias Jarke, Thomas Seidl, Christoph Quix, Christoph Brochhaus (Hrsg.): Datenbanksysteme in Business, Technologie und Web (BTW 2007)
- P-104 Birgitta König-Ries, Franz Lehner, Rainer Malaka, Can Türker (Hrsg.): MMS 2007: Mobilität und mobile Informationssysteme
- P-105 Wolf-Gideon Bleek, Jörg Raasch, Heinz Züllighoven (Hrsg.): Software Engineering 2007
- P-106 Wolf-Gideon Bleek, Henning Schwentner, Heinz Züllighoven (Hrsg.): Software Engineering 2007 – Beiträge zu den Workshops
- P-107 Heinrich C. Mayr, Dimitris Karagiannis (eds.): Information Systems Technology and its Applications
- P-108 Arslan Brömmе, 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 1<sup>st</sup> Conference on Social Semantic Web (CSSW)
- P-114 Sandra Frings, Oliver Göbel, Detlef Günther, Hardo G. Hase, Jens Nedon, Dirk Schadt, Arslan Brömmе (Eds.): IMF2007 IT-incident management & IT-forensics Proceedings of the 3<sup>rd</sup> 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 1<sup>st</sup> International Working Conference on Business Process and Services Computing BPSC 2007
- P-117 Ryszard Kowalczyk (Ed.): Grid service engineering and management The 4<sup>th</sup> International Conference on Grid Service Engineering and Management GSEM 2007
- P-118 Andreas Hein, Wilfried Thoben, Hans-Jürgen Appelrath, Peter Jensch (Eds.): European Conference on ehealth 2007
- P-119 Manfred Reichert, Stefan Strecker, Klaus Turowski (Eds.): Enterprise Modelling and Information Systems Architectures Concepts and Applications
- P-120 Adam Pawlak, Kurt Sandkuhl, Wojciech Cholewa, Leandro Soares Indrusiak (Eds.): Coordination of Collaborative Engineering - State of the Art and Future Challenges
- P-121 Korbinian Herrmann, Bernd Bruegge (Hrsg.): Software Engineering 2008 Fachtagung des GI-Fachbereichs Softwaretechnik
- P-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.)  
9<sup>th</sup> 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-127 Thomas Kühne, Wolfgang Reisig,  
Friedrich Steimann (Hrsg.)  
Modellierung 2008

## ***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.)  
          Informatiktage 2005  
          Fachwissenschaftlicher Informatik-  
          Kongress
- S-3     Gesellschaft für Informatik (Hrsg.)  
          Informatiktage 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.)  
          Informatiktage 2007  
          Fachwissenschaftlicher Informatik-  
          Kongress
- S-6     Gesellschaft für Informatik (Hrsg.)  
          Informatiktage 2008  
          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](mailto:druckverlag@koellen.de)



