

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 co-operation with GI and to publish the annual GI Award dissertation.

Broken down into the fields of

- Seminars
- 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 978-3-88579-443-1



Gesellschaft für Informatik (Hrsg.): Informatiktag 2010

GI-Edition

Lecture Notes in Informatics



Gesellschaft für Informatik (Hrsg.)

Informatiktag 2010

Fachwissenschaftlicher Informatik-Kongress

19. und 20. März 2010

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



Gesellschaft für Informatik (Hrsg.)

Informatiktage 2010

Fachwissenschaftlicher Informatik-Kongress

19. und 20. März 2010

**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-9

ISSN 1614-3213

ISBN 978-3-88579- 443-1

Volume Editor

Gesellschaft für Informatik e.V.

Ahrstraße 45

53175 Bonn

E-Mail: gs@gi-ev.de

Redaktion: Ludger Porada

E-Mail: ludger.porada@gi-ev.de

Series Editorial Board

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

Hinrich Bonin, Leuphana-Universität Lüneburg, Germany

Dieter Fellner, Technische Universität Darmstadt, Germany

Ulrich Flegel, SAP Research, Germany

Ulrich Frank, Universität Duisburg-Essen, Germany

Johann-Christoph Freytag, Humboldt-Universität Berlin, Germany

Thomas Roth-Berghofer, DFKI

Michael Goedicke, Universität Duisburg-Essen

Ralf Hofestädt, Universität Bielefeld

Michael Koch, Universität der Bundeswehr, München, Germany

Axel Lehmann, Universität der Bundeswehr München, Germany

Ernst W. Mayr, Technische Universität München, Germany

Sigrid Schubert, Universität Siegen, Germany

Martin Warnke, Leuphana-Universität Lüneburg, Germany

Dissertations

Dorothea Wagner, Universität Karlsruhe, Germany

Seminars

Reinhard Wilhelm, Universität des Saarlandes, Germany

Thematics

Andreas Oberweis, Universität Karlsruhe (TH)

© Gesellschaft für Informatik, Bonn 2010

printed by Köllen Druck+Verlag GmbH, Bonn

Wissenschaftliche Tagungsleitung

Gottfried Vossen, Universität Münster

Programmkomitee

Christoph Beierle, FernUniversität Hagen

Hermann Engesser, Informatik Spektrum

Otto Spaniol, RWTH Aachen

Alfred Zimmermann, Fachhochschule Reutlingen

Workshops - Themen und Partner

SOA Architekturmuster

Capgemini sd&m AG

Web 2.0 - Technologien

iteratec Gesellschaft für iterative Softwaretechnologien mbH

Schnittstellendesign - Sprachen, Formate, Protokolle

PPI Financial Systems

IT-Sicherheitsmanagement und Informationsflusskontrolle

Sirrix AG

Java EE 6 – Live-Programmierung

Micromata

Neue Shopkonzepte

Deutsche Telekom AG

Übersicht über das Rahmenprogramm und Partner der Informatiktage 2010

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 2008

Wissenschafts- und Absolventen-Workshop

Leitung: Prof. Dr. Alfred Zimmermann, Fachhochschule Reutlingen

Workshop-Spezial: Software-Ingenieure als Unternehmer

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

Hauptvortrag:

Dr. Alexander Löser, TU Berlin

Web 2.0 Business Analytics: Qualität durch Datenmenge?

Talk-Show des Computer Clubs 2

Mit Peter Schaar, dem Datenschutzbeauftragten der Bundesrepublik

Postersessions

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:

Capgemini sd&m AG

Deutsche Telekom AG

iteratec Gesellschaft für iterative Softwaretechnologien mbH

Micromata

PPI Financial Systems

Sirrix AG

Medienpartner:

IX

Computer Club 2

Staufenbiel Media GmbH

Vorwort zum Tagungsband 2010

Die Förderung der Zusammenarbeit von Wissenschaft und Wirtschaft war und ist eines der zentralen Anliegen der GI. Wichtigstes Instrument dafür sind unsere Fachtagungen, die von unseren Fachbereichen und Fachgruppen organisiert und fachlich ausgerichtet werden. Darüber hinaus gibt es aber auch direkte Maßnahmen, mit denen das Kennenlernen von Studierenden und Unternehmen unterstützt wird. Es sind dies vor allem die Networking-Aktivitäten in unseren Regionalgruppen, bei denen Studierende explizit angesprochen und in die fachliche Arbeit einbezogen werden. Ganz wichtig empfinde ich auch die Flaggschiffe der GI-Nachwuchsförderung: Die Informatiktage und den informatiCup, der inzwischen sehr attraktive GI-Studierendenwettbewerb. Bei diesen Veranstaltungen gehört die Zusammenarbeit zwischen Hochschulen und Wirtschaft praktisch zum Programm.

Zentraler Programmpunkt der Informatiktage sind die praxisorientierten Workshops, die von Unternehmen der IT-Branche angeboten werden. Die Studierenden, werden vom Hochschulnetzwerk der GI zur Teilnahme ausgewählt und arbeiten in den Workshops an konkreten Aufgaben, die von den Firmen vorgeschlagen werden. Davon profitieren beide Seiten: Die Studierenden erlernen Methodenkompetenzen und gewinnen Einblicke in konkrete planerische Prozesse der Unternehmen, die Unternehmen erhalten ihrerseits Einblicke in die Problemlösungs- und Fachkompetenzen der Studierenden. Nicht selten ergeben sich für die Unternehmen ganz neue Blickwinkel auf aktuelle Problemlagen. Nirgendwo sonst erhalten Unternehmen so tiefe Einblicke in den Leistungsstand der Informatik in deutschen Hochschulen.

Aber nicht nur die Arbeit in den Workshops macht die Informatiktage attraktiv, auch das Rahmenprogramm verdient Aufmerksamkeit. Der Schwerpunkt ist in diesem Jahr das Thema „Datensensibilität“. Im Gespräch wird sich uns der Datenschutzbeauftragte der Bundesrepublik, Peter Schaar, vorstellen. Street-View, ELENA, die Gesundheitskarte und „Mobil Apps“ sind die Stichworte, die die Aktualität des Themas und den Handlungsdruck verdeutlichen.

Den Unternehmen danke ich für ihre Bereitschaft, die Informatiktage zu fördern, dem B-IT für seine erneute Gastfreundschaft und den Vertrauensdozent/inn/en und dem Programmkomitee für die inhaltliche Arbeit. Ich wünsche den Teilnehmerinnen und Teilnehmern der Informatiktage viel Erfolg und hoffe, dass Sie sowohl untereinander als auch in Gesprächen und Diskussionen mit den vertretenen Unternehmen viele Anregungen für ihre berufliche und fachliche Karriere erhalten.



Prof. Dr. Stefan Jähnichen (GI-Präsident)

Inhaltsverzeichnis

Grundlagen der Informatik

Bruns D.

Formal Semantics for the Java Modeling Language 15

Gemsa A.

Schematized Visualizations of Shortest Paths in Road Networks 19

Künstliche Intelligenz

Besold T.

Theory and Implementation of Multi-Context Systems Containing Logical and Sub-Symbolic Contexts of Reasoning 23

Holz D.

Naive but Efficient – Using Greedy Strategies for Exploration, Inspection and Search 27

Janning R.

Konstruktion von Common Sense-Ontologien durch Analyse natürlicher Sprache 31

Kämpgen B.

Deskriptives Data-Mining für Entscheidungsträger: Eine Mehrfachfallstudie 35

Lindner F.

Perzeptuelle Verankerung ausgedehnter Landmarken während der instruierten Navigation 39

Neumann J.P.

Ein modularer Lösungsansatz für das University-Course-Timetabling-Problem 43

Schneider M.

Umsetzung eines auf Potentialfeldern basierenden Pfadplanungsalgorithmus für die Jenaer Robot Soccer Engine 47

Softwaretechnik

Geller J.

XML Import/Export für FBD in OpenPCS 51

Heinzemann C.

Verifikation von Protokollverfeinerungen 57

Heym A.

Ausarbeitung und Umsetzung eines Ansatzes für modellbasiertes Testen in betrieblichen Informationssystemen 61

Hoyer J., Kruthoff-Bruewer R.

Peer-to-Peer Kommunikation zwischen Multimediaanwendungen in einem Placeshiftszenario 65

Kohring A.

IQ Profil: Eine MediaWiki Erweiterung zur Bewertung von Informationsqualität der Artikel 69

Ljaci N.

Integration von MockUp-Konzepten in die Spezifikation grafischer Bedienoberflächen 73

Niebuhr B.

Testfallgenerierung aus UML-Modellen beschrieben in UML-Testing-Profile 77

Noack T.

Systematisches Abwickeln und strukturiertes Ablegen von Qualitätszielen in der Individualsoftwareentwicklung 81

Pasdziernik M.

Testgetriebene Entwicklung eines Mehrphasen-Compilers 85

Prätsch S.

Ein Modellprüfungsansatz für die Erreichbarkeitsprüfung von Process Algebra Nets 89

Roth S.

Context- and Domain-Aware Semantic Web Services 93

Schönfeld F.

Der Physik-Engine Editor ped 97

Wagner T.

Prototypische Realisierung einer Integration von Agenten und Workflows 101

Mensch-Computer-Interaktion

Apel S.

Ellipsis - Containerbasiertes Agentensystem in einer Hochverfügbarkeitsstruktur 105

Diouf K.

Social Software auf PVR-Systemen 109

Döhring M.

Automatically Tagging Online Community Content using
Probabilistic Topic Models 113

Maiero J.

Image-based Tracking 117

Möckel S.

CARD-PAL – Ein kartenbasierter Ansatz zur Visualisierung und
Formalisierung statischer Aspekte während der Anforderungsanalyse 121

Neidhart A.

Ein Konzept und Entwicklung eines Prototyps zur Auswertung von
Volleyballspielen auf Grundlage der Videoaufzeichnungen 125

Richlick T.

Virtuelle interaktive Prospekte mit Hilfe der Nintendo Wiimote 129

Rutetzki C.

Integration von syntaxbasierter Nutzerunterstützung in Sketching-Editoren 133

Westermann T.

I'm Home: Smartphone-enabled Gestural Interaction with
Multi-Modal Smart-Home Systems 137

Datenbanken und Informationssysteme

Boden B.

Synthese von Subspace Clustering und Subgraph Mining zur Analyse von
Graphen mit Knotenbeschriftungen 141

Otto M.

Konzeption und Implementierung einer Regression Datenbank zu
Unterstützung der IP-Entwicklung 145

Terwelp C.

Konstruktion und Optimierung von Diskriminierungsnetzwerken 149

Tu V.H.

Ansätze zur Integration unterschiedlicher Data Warehouse-Architekturen 153

Technische Informatik

Beier T.

Berechnung von Korrelation in Echtzeit auf der Cell BE Plattform 157

Dorji T.

Decodability Model for Dynamically Adaptive Prioritization Algorithm
in H.264/SVC Video Streaming 161

Franke D.

Verifikation der Java-Echtzeitfähigkeit für den Einsatz in zeitkritischen Systemen 165

Hauser S.

Entwurf eines generischen, applikationsspezifischen, transportgesteuerten
Prozessor-Modells in VHDL und Validierung auf einem FPGA 169

Kulas M.

Entwurf und Realisierung eines Adapters für UniLoG zur Lastgenerierung
an IP-basierten Schnittstellen 173

Messmer R.

Performant-Optimierung eines Echtzeit-Multimedia-Netzwerkes 177

Paveesuda C.

Service Brokering for Efficient Service Selection in Wireless Mesh Networks 181

Sapkaew W.

Performance Analysis of Transmission Power Control in Wireless Mesh Networks 185

Schleiden C.

Exploiting Object Orientation to Parallelize and Optimize C++ Applications 189

Informatik in der Lebenswissenschaft

Bischoff S.

Konzeption und Umsetzung einer RIA zur untersuchungsbegleitenden Erfassung
von RNFLT-Scans und Untersuchung von Klassifikatoren für die diagnostische
Unterstützung bei neurodegenerativen Erkrankungen am Beispiel der
Multiplen Sklerose 193

Brink M.

A Skeletal Approach for Parallel Medical Imaging 195

Glaß M.

Detektion von Strukturen in Fluoreszenzmikroskopiebildern 199

Grimm R.

Automatic Patient Pose Estimation 203

Haupt C.

Markush Structure Reconstruction -A Prototype for their Reconstruction from Image and Text into a Searchable, Context Sensitive Grammar based Extension of SMILES 207

Kullick R.

Realisierung eines Clients zur Visualisierung von verteilten XML-Dokumenten für ein Informationssystem in der Landwirtschaft 211

Lenk K.

Quantitative measurement of regional wall motion in the right heart ventricle 215

Schneider M.

Model-Based Respiratory Motion Compensation for Image-Guided Cardiac Interventions 219

Graphische Datenverarbeitung**Friedl M., Nagl F., Schäfer A., Tschentscher A.**

Shader-Based-Image-Processor - Ein Framework für echtzeitfähige, grafikartenbasierte Bilderverarbeitung 223

Burggraf T.

PhotoGet – Bilderbereitstellung einfach gemacht 227

Fröhlich B.

Semantische Segmentierung 231

Heppner C.

Ein Bildequalizer zur Lehreunterstützung 235

Hönig M.

Umsetzung eines 3D-Action-Adventures mit der Open Source 3D-Grafik-Engine Irrlicht 239

Knopp B.

Modellbasierte Poseschätzung aus dichten Volumendaten 243

Marklein M.

Efficient Game State Management für Massive Multiplayer Online Games 247

Meissner M.

GPU-basiertes Volumen-Rendering mit Hilfe von Octree gestütztem Raycasting 251

Nagl F., Röse J.

Konzeption und Umsetzung eines 3D-Bildbetrachters 255

Wirtschaftsinformatik

Elsler R., Horstmann M., Körner M.

Negolook: Eine Integration clientseitiger Verhandlungsunterstützung in betriebliche Kommunikationssysteme 259

Gutbrod M., Schiele F., Zinsenhofer C.

Serviceorientiertes-Modell eines TOGAF basierten Portals für Enterprise Architecture Management (EAM) 263

Johannes H.

Retooling in der Automobilindustrie: Betriebswirtschaftliche Vorteile durch den Einsatz einer Software 267

Langner B.

PhotoBird – Entwicklung eines Corporate Designs anhand einer multimedialen Anwendung 271

Urmann M.

Realisierung einer Anbindung von SAP ERP zu SAP EWM auf Basis SAP SCM und Dokumentation von Technik, Systemen und Schnittstellen zur Abbildung eines ausgewählten Geschäftsprozesses 275

Wehr D.

Fachliche Modellierung von multidimensionalen Datenstrukturen im Business Intelligence Umfeld einer öffentlichen Verwaltung 279

Yildiz B.

Ein intelligentes Benachrichtigungssystem für das Incident Management 283

Sicherheit – Schutz und Zuverlässigkeit

Blaß T.

Generic Parallel Computation of Large Garbled Circuits using JaMP 287

Hanspach M.

Dynamisierung von Probability-Impact-Matrizen 291

Somorovský J.

Streaming-based Processing of Secured XML Documents 295

Formal Semantics for the Java Modeling Language

Daniel Bruns
Karlsruhe Institute of Technology
bruns@kit.edu

A common critique of formal methods in software development practise is, that they are not readily understandable and thus not widely used (see for instance [Nam97]). The Java Modeling Language (JML) was created in an attempt to bridge that gap. By building upon the syntax of Java it is meant to be easily accessible to the common user – who might *not* be skilled in formal modeling. Due to this advantage, JML has quickly become one of most popular specification languages to use with both static and runtime analysis of programs. JML specifications are written in a Java-like expression language as comments straight into source files. It provides both in-code assertions as well as method contracts and class invariants which are indicated by appropriate preceding keywords. Fig. 1 shows a method specification. However, the official reference [LPC⁺09] mostly lacks a clear definition of semantics. In turn, several tools implementing JML syntax use their own interpretations. Past approaches to formal semantics have rather been documentations of those tools than providing a unified reference.

```
/*@ requires x >= 0;  
  *@ ensures \result == (\product int z;  
  *@          0 < z & z <= x; z); @*/  
public int factorial (int x) {  
    if (x <= 0) return 1;  
    else return x * factorial(x-1);  
}
```

Figure 1: Method specification with pre- and post-condition for a factorial calculation. Note that it is only partial; the result of negative inputs is not specified. Special expressions for the method result and a product over integers are used.

The present work makes use of purely mathematical notations, such as predicate logic with sets and functions, to describe a simple machine model which provides the semantical entities for a rigorous evaluation of JML artifacts. Predominantly, the semantics given throughout this work are based on the informal descriptions found in the JML Reference Manual. As many features are defined in terms of Java, we relied on the Java language specification [GJSB05] where needed – without formalizing Java itself. As a result, most structures which are commonly used in the specification of sequential Java are covered by this formalization. As this work had revealed several conceptual flaws or inconsistencies within the language design, we have put further efforts in pushing a sound official documentation.

Foundations First of all, we define a formalization of the virtual machine on which a Java program runs. A (closed) program defines a *universe* \mathcal{U} of primitive values and typed semantical objects. We define a *system state* $s \in \mathcal{S}$ as a triple (h, σ, χ) where (i) $h : \mathcal{U} \times \mathcal{I} \rightarrow \mathcal{U}$ is a function representing the *heap* by mapping pairs of objects and identifiers to elements of the universe, (ii) $\sigma : \mathcal{I} \rightarrow \mathcal{U}$ is a function representing the *stack* by evaluating local variables and (iii) χ is a sequence of named methods and respective receiver objects representing a *call stack*. An executable fragment of a program (say, a method body) invoked in a pre-state s_0 yields a (possibly infinite) sequence R (the *run*) of states which are passed through in execution, a termination witness Ω (an instance of `Throwable` where `null` denotes a successful execution) and a return value ρ .

Expressions The syntax of JML expressions is essentially the same as in Java, but enriched with quantification and special constructs, such as `\result` to access the return value of a method invocation, or `\old` to refer to pre-state values of expressions (in a method's post-state). We use an evaluation function $val_{s_0}^{s_1} : Expr \rightarrow \mathcal{U}$ which depends on two system states – a pre-state s_0 and a post-state s_1 . Formulae in this sense are just `boolean` typed expressions. Several features (adapted from Java) however, make JML expressions harder to handle than classical first order logic: (i) Expressions may have side-effects under certain circumstances, e.g. `new Object() == o` is a legal assertion. This means for expressions not only yielding a value but also a transition from one state to another. We describe this using a function $\omega : Expr \rightarrow \mathcal{S}^S$. As a result, for *any* expression the order of sub-expressions does matter. (ii) The exceptional behavior of Java is imitated, which can be interpreted as a covert third truth value. E.g. `0/0 == 0/0` is not semantically valid, even though it appears to be a logical tautology. In order for a boolean expression be considered *valid*, it has not only to yield the value *true* but also be well-defined. This is evaluated by the function $wd_{s_0}^{s_1} : Expr \rightarrow \{true, false\}$. (iii) To the above two items *short-circuit* evaluation applies as in Java. As an example, a boolean disjunction is evaluated in the following way:

$$\begin{aligned} val_{s_0}^{s_1}(A \parallel B) = true & \quad \text{iff} \quad val_{s_0}^{s_1}(A) = true \quad \text{or} \quad val_{\omega(A)(s_0)}^{\omega(A)(s_1)}(B) = true \\ wd_{s_0}^{s_1}(A \parallel B) = true & \quad \text{iff} \quad wd_{s_0}^{s_1}(A) = true \quad \text{and} \\ & \quad val_{s_0}^{s_1}(A) = true \quad \text{or} \quad wd_{\omega(A)(s_0)}^{\omega(A)(s_1)}(B) = true \end{aligned}$$

As a result, `true || 0/0 == 0/0` is a valid proposition, while its commutation is not. While this distinction is perfectly reasonable from a programmer's view, additional effort has to be taken to incorporate this into classical logic calculi. Fortunately, the definition *val* is independent of *wd*. This offers the possibility to encode *wd* as a formula of classical logic and to view it as an additional proof obligation while leaving the definition of truth untouched.

Specifications JML extends the design-by-contract paradigm [Mey92] by introducing method contract facilities which allow not only to specify pre- (requires α) and post-conditions (ensures β) for successful executions, but also post-conditions for exceptional termination cases (signals (e) γ) as well as necessary pre-conditions for non-

termination (diverges δ). Assuming the pre-condition to hold¹ in state s_0 (i.e. $val_{?}^{s_0}(\alpha) = true$), there are different conditions to hold depending on the mode of termination (with the above definition of R , Ω and ρ):

Termination mode	Necessary condition
None ($ R = \infty$)	$val_{?}^{s_0}(\delta) = true$
Exceptional ($R = \langle s_0, \dots, s_k \rangle, \Omega \neq \text{null}$)	$val_{s_0}^{s_k \oplus \{e \mapsto \Omega\}}(\gamma) = true$
Normal ($R = \langle s_0, \dots, s_k \rangle, \Omega = \text{null}$)	$val_{s_0}^{s_k \oplus \{\text{result} \mapsto \rho\}}(\beta) = true$

Invariants are evaluated in a similar way, but their semantics are based on the *visible state* paradigm [Poe97]. This essentially means that objects have to respect invariants as long as they are not the receiver of a currently active method invocation. This is a stronger requirement than the common *observable state* semantics, which means to assert the invariant just in a method's post-state. That may lead to invariant violations upon re-entrant method calls. It turns out that the property of a state being visible is not deducible from itself, but from the call stacks of the run as a whole. The characterization in this thesis has shown several inconsistencies in the very design of visible states, such as virtually any constructor would be unable to even establish an invariant.

Model fields JML also features the notion of *model fields* [CLSE05] which are not part of the program but added in specification for means of abstraction. Yet there has been no common understanding in what they should represent. One viewpoint is that model fields abstract from the current machine state. Another one is that they may refer to objects which are not referenced by the program itself, thus defining a definitional extension to the logic. In the former case, their use would be very similar to that of logical axioms, while in the latter one has to provide extended evaluation facilities (e.g. define a second heap which holds values to the model object's fields). In any way, specification features must not interfere with the actual program. As this is still a hot topic, significantly more work is needed to reach a reference definition.

Tool Support The KeY tool [BHS07] for formal verification of Java Card programs provides a JML interface (among others). JML specifications are translated into formulae of dynamic logic [Eng05]. As KeY implements only a subset for which sound formal semantics exist, its interpretation of the language comes very close to the anticipated (informal) meaning. As the need for treatment of a substantial part of Java (including generics, concurrency, etc.) rises, there is yet still a demand for more sound JML semantics.

¹For simplicity, we only show *val* while *wd* has to hold for the same parameters. ? does denote an unspecified state here.

References

- [BHS07] Bernhard Beckert, Reiner Hähnle, and Peter H. Schmitt, editors. *Verification of Object-Oriented Software: The KeY Approach*. LNCS 4334. Springer-Verlag, 2007.
- [Bru09] Daniel Bruns. Formal Semantics for the Java Modeling Language. Diploma thesis, Universität Karlsruhe, June 2009.
- [CLSE05] Yoonsik Cheon, Gary T. Leavens, Murali Sitaraman, and Stephen Edwards. Model Variables: Cleanly Supporting Abstraction in Design By Contract. *Software — Practice & Experience*, 35(6):583–599, May 2005.
- [Eng05] Christian Engel. A Translation from JML to JavaDL. Studienarbeit, Fakultät für Informatik, Universität Karlsruhe, February 2005.
- [GJSB05] James Gosling, Bill Joy, Guy Steele, and Gilad Bracha. *The Java Language Specification, Third Edition*. The Java Series. Addison-Wesley, 2005.
- [LPC⁺09] Gary T. Leavens, Erik Poll, Curtis Clifton, Yoonsik Cheon, Clyde Ruby, David R. Cok, Peter Müller, Joseph Kiniry, Patrice Chalin, Daniel M. Zimmerman, and Werner Dietl. JML Reference Manual. Draft. Revision 1.235, Available from <http://www.jmlspecs.org/>, September 2009.
- [Mey92] Bertrand Meyer. Applying “Design by Contract”. *IEEE Computer*, 25(10):40–51, October 1992.
- [Nam97] Baudouin Le Charlier Namur. Specifications are Necessarily Informal or: The Ultimate Myths of Formal Methods. Technical Report BU-CEIS-9703, Bilkent University, Faculty of Engineering, 1997.
- [Poe97] Arndt Poetzsch-Heffter. *Specification and verification of object-oriented programs*. Habilitationsschrift, Technische Universität München, January 1997.

Schematized Visualizations of Shortest Paths in Road Networks

Andreas Gemsa*

In der heutigen mobilen Gesellschaft sind Routenplanungsdienste aus dem privaten sowie wirtschaftlichen Bereich kaum noch wegzudenken. Dabei gibt es grundsätzlich zwei verschiedene Ansätze. Routenplanung kann in einem mobilen Endgerät durchgeführt werden, das den Fahrer während der Fahrt über den Routenverlauf unterrichtet. Ein anderer Ansatz ist es, dem Fahrer vor der Fahrt eine Routenbeschreibung mitzuteilen. Das kann durch Online-Routenplanungsdienste wie zum Beispiel Google Maps [GMa] oder Map24 [Map] erfolgen. Typischerweise wird dem Benutzer eine Routenübersicht mit Hilfe einer Karte und eine textuelle Beschreibung der Route geliefert. Die Karte zeigt die Route farblich hervorgehoben und der mitgelieferte Text beschreibt den genauen Verlauf. Üblicherweise besteht eine Route (gewisser Länge) aus drei Abschnitten: Einem, aber oft komplexen Abschnitt in der Stadt um den Start der Route, einem relativ langen Stück Autobahn, Bundesstraße und/oder Landstraße und schließlich einem kurzen Abschnitt um das Reiseziel [MHBB06]. Bei den erwähnten Systemen werden diese drei Abschnitte mit dem gleichen Maßstab dargestellt. Das macht es für den Benutzer schwierig, Details in der Umgebung des Reisestarts sowie des Reiseziels auszumachen. Deshalb sind solche Darstellungen nur bedingt hilfreich. Würde man einen Menschen bitten eine Wegbeschreibung zu zeichnen, so würde sich dieses Bild von dem der Online-Routenplanungsdienste deutlich unterscheiden. Viele Informationen, die für den Routenverlauf unerheblich sind, sind auf dieser Art Zeichnungen nicht dargestellt. Die Darstellung konzentriert sich stark auf die Informationen die für die Route wichtig sind. Dabei werden die exakten geographischen Gegebenheiten der Strecke verfälscht. Das betrifft sowohl die Orientierung als auch die Länge der benutzten Straßenabschnitte. Normalerweise sind die komplexeren Bereiche in der Umgebung von Start und Ziel der Route vergrößert gezeichnet.

In dieser Diplomarbeit wird ein mehrstufiges Verfahren erarbeitet, das in der Lage ist solche Zeichnungen automatisch zu generieren. Dabei bildet ein Algorithmus, der monotone Routenabschnitte schematisiert eine wichtige Rolle.

Argumente, die für eine schematische Routendarstellung sprechen, finden sich in der Psychologie. Da Menschen Richtungsänderungen in sieben Kategorien (scharf links, links, halb links, geradeaus, halb rechts, rechts, scharf rechts) wahrnehmen [Kli03], genügt es Richtungsänderungen auch nur in diesen Kategorien anzudeuten. Weiterhin ist eine Reduzierung der dargestellten Information aus zwei Gründen von Vorteil. Zum Einen muss das Gehirn nicht mehr selbst entscheiden welche Informationen wichtig sind und zum Anderen muss das von der Kapazität sehr begrenzte Kurzzeitgedächtnis [Mil56] weniger Daten verwalten. Ein letzter Aspekt ist der Erhalt der orthogonalen Ordnung [MELS95].

*Institut für Theoretische Informatik, Karlsruher Institut für Technologie, gemsa@iti.uka.de

Menschen haben in der Regel eine grobe Vorstellung, die sogenannte mentale Karte, von der Strecke zwischen Start und Ziel. Es ist sinnvoll diese Relation in Form der orthogonalen Ordnung beizubehalten.

Ein Aspekt der Schematisierung von beliebigen Pfaden wurde bereits von Brandes und Pampel [BP09] untersucht. Sie haben gezeigt, dass für beliebige Pfade das Finden einer neuen Einbettung, die rektilinear ist und die orthogonale Ordnung der Eingabeeinbettung erhält, \mathcal{NP} -schwer ist. Aus diesem Grund beschränken wir uns in dieser Arbeit auf Pfade, die eine x - oder y -monotone Eingabeeinbettung haben und geben hierfür einen polynomiellen Algorithmus an. Für das grundsätzliche Problem Routen skizzenhaft darzustellen wurde von Agrawala und Stolte bereits ein Ansatz vorgestellt, der auf simulated annealing beruht [AS01]. Aufgrund der verwendeten heuristischen Lösungsstrategie können jedoch keine Gütegarantien für die Ausgabe zugesichert werden.

Für die Schematisierung wurde von uns ein spezieller Algorithmus entworfen, der sich unter anderem der dynamischen Programmierung bedient. Die Eingabe des Algorithmus besteht aus einem einfachen Pfad $P = (v_1, \dots, v_n)$ und einer x - oder y -monotonen Einbettung π . In dem konkreten Fall dieser Arbeit ist die Einbettung π ein Teil des geographischen Verlaufs der Route. Des Weiteren wird dem Algorithmus eine Menge von zulässigen Kantenrichtungen \mathcal{C} übergeben (z.B. $\{0^\circ, 45^\circ, 90^\circ, \dots, 315^\circ\}$). Jede Kante e des Pfades hat, bezogen auf π , eine bevorzugte Richtung $\omega(e)$ aus der Menge \mathcal{C} . Diese bevorzugte Richtung ist die Richtung in \mathcal{C} , die die geringste Differenz zur Kantenrichtung von e in π aufweist. Ziel des Algorithmus ist es eine Einbettung ρ zu berechnen, die nur Kantenrichtungen aus \mathcal{C} verwendet, die orthogonale Ordnung von π erhält, und nur einer minimalen Anzahl von Kanten *nicht* ihre jeweilige präferierte Richtung zuweist. Wir geben einen polynomiellen Algorithmus an, der mit Hilfe von dynamischer Programmierung, einer Transformation von Kanten des Pfades auf Punkte in der euklidischen Ebene und der Verwendung von Range Trees [dBvKOS00], eine Laufzeit von $\mathcal{O}(n^5 \log n)$ hat. Dabei bestimmt der Algorithmus zunächst nur, ob eine Kante mit ihrer präferierten Richtung eingebettet werden kann oder nicht. Im Anschluss wird ein lineares Programm genutzt, welches die tatsächlichen Längen der Kanten bestimmt. In einem letzten Schritt werden dann die Koordinaten aller Knoten in der Ausgabe-Einbettung ρ ermittelt.

Die grundsätzliche Idee hinter dem in der Diplomarbeit vorgestellten mehrstufigen Verfahren zur automatischen Erzeugung von Routen-Skizzen, ist wie folgt: Das Ziel ist den Großteil der Route zu schematisieren. Zunächst wird ein relativ kleiner Teil des Start- und Endbereichs abgetrennt. Dieser wird nicht schematisiert, da typischerweise gerade zu Beginn und Ende der Strecke viele Richtungsänderungen erfolgen. Eine Schematisierung dieses Bereichs wäre nur von geringem Nutzen und möglicherweise sogar kontraproduktiv. Der verbliebene Teil des Pfades wird anschließend mit Hilfe des Douglas Peucker Algorithmus [DP73] zunächst vereinfacht. Allerdings wird beim Vereinfachen darauf Wert gelegt, dass es an keiner Stelle der Route zu inkonsistenten Richtungsänderungen an Kreuzungen kommen kann. Da die verbliebene Teilroute in der Regel nicht x - oder y -monoton ist, werden mit einer greedy Heuristik maximale x - oder y -monotone Teilpfade gesucht, die vom Schematisierungs-Algorithmus dann schematisiert werden. Sobald alle Teilpfade schematisiert wurden, werden sie wieder zusammengefügt. Da sowohl Kantenlängen als auch Kantenrichtungen verändert wurden, kann es sein, dass es zwischen

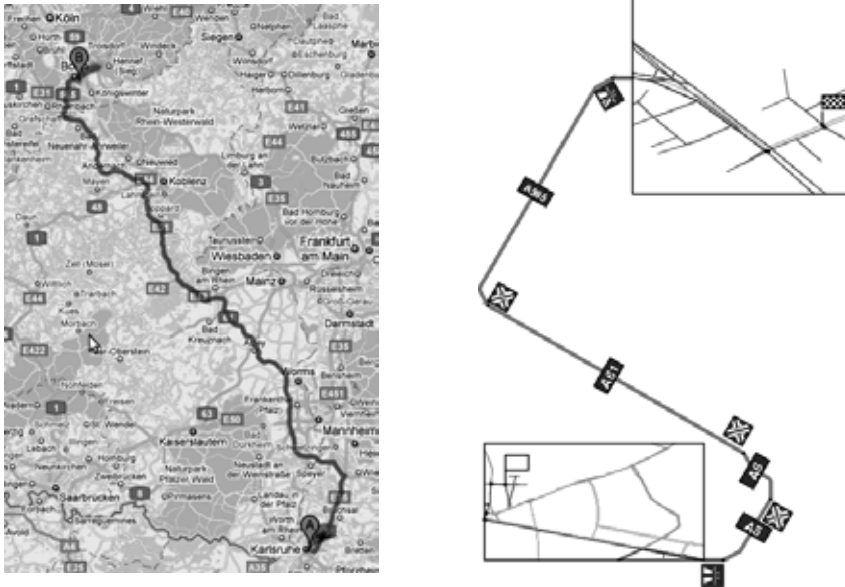


Abbildung 1: Route vom Institut für theoretische Informatik in Karlsruhe bis zum B-IT in Bonn

zwei schematisierten Teilpfaden zu einem ursprünglich nicht vorhandenen Schnitt kommt. In der Arbeit wird ein Verfahren erläutert, das solch unerwünschte Schnitte verhindert. Nachdem die Route nun, bis auf die zu Anfang abgetrennten Bereiche, wieder vollständig ist, wird die Ausgabe mit zusätzlichen Informationen ergänzt, die dem Nutzer helfen sollen (z.B. Piktogramme für Richtungsänderungen, Autobahnwechsel usw.). Die Positionierung der Piktogramme basiert auf dem *Label Placement Problem*, was für sich genommen bereits ein \mathcal{NP} -schweres Problem ist [KT97]. Wir verwenden ein Verfahren von Yamamoto et al. [YCL05] welches heuristisch eine möglichst große unzusammenhängende Knotenmenge des Konfliktgraphs bestimmt. Es ist möglich, dass dieses Verfahren nicht alle Konflikte, also Überschneidungen der Piktogramme, verhindert. Abschließend werden die Teile, die zu Beginn abgetrennt wurden passend wieder angefügt.

In dieser Diplomarbeit wird ein Algorithmus entwickelt, der sich diversen Techniken der theoretischen Informatik bedient. Ein wichtiger Bestandteil bildet dabei die Technik des dynamischen Programmieren. Aber auch der Einsatz von linearer Programmierung und geometrischer Datenstrukturen ist wichtig. Dieses Verfahren wird in einem Kontext – der automatischen Generierung von Routen-Skizzen – erläutert, welches eine praktische Anwendung demonstriert. Weiterhin werden Lösungsstrategien für Probleme entwickelt, die beim Erstellen einer Routen-Skizze auftreten können. Alle Teile des mehrstufigen Verfahrens wurden implementiert und auf dem deutschen Straßennetz evaluiert. Abbildung 1 zeigt ein die Route von Karlsruhe nach Bonn. Auf der linken Seite ist eine Ausgabe von Google Maps und auf der rechten Seite, zum Vergleich, eine automatisch generierte Routen-Skizze. Die Daten des deutschen Straßennetzes wurde freundlicherweise von der PTV AG [ptv] zur Verfügung gestellt.

Literatur

- [AS01] Maneesh Agrawala and Chris Stolte. Rendering effective route maps: improving usability through generalization. In *SIGGRAPH '01: Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, pages 241–249, New York, NY, USA, 2001. ACM.
- [BP09] Ulrik Brandes and Barbara Pampel. On the Hardness of Orthogonal-Order Preserving Graph Drawing. In *Proc. of the 16th International Symposium on Graph Drawing*, volume 5417 of *Lecture Notes on Computer Science*, pages 266–277. Springer-Verlag, 2009.
- [dBvKOS00] Mark de Berg, Marc van Kreveld, Mark Overmars, and Otfried Schwarzkopf. *Computational Geometry Algorithms and Applications*. Springer-Verlag; second edition, February 2000.
- [DP73] D.H. Douglas and T.K. Peucker. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. *Canadian Cartographer*, 10:112–122, 1973.
- [GMa] Google Maps : <http://maps.google.com>.
- [Kli03] Alexander Klippel. *Wayfinding Choremes - Conceptualizing Wayfinding and Route Direction Elements*. PhD thesis, University of Bremen, 2003.
- [KT97] Konstantinos G. Kakoulis and Ioannis G. Tollis. On the Edge Label Placement Problem. In *Proc. Symp. 4th Internat. Symp. on Graph Drawing (GD'96)*, volume 1190, pages 241–256, 1997.
- [Map] Map24 : <http://www.map24.de>.
- [MELS95] Kazuo Misue, Peter Eades, Wei Lai, and Kozo Sugiyama. Layout Adjustment and the Mental Map. *Journal of Visual Languages and Computing*, 6(2):183–210, 1995.
- [MHBB06] Tobias Meilinger, Christoph Hölscher, Simon J. Büchner, and Martin Brösamle. How Much Information Do You Need? Schematic Maps in Wayfinding and Self Localisation. In *Spatial Cognition*, number 4387 in *Lecture Notes in Computer Science*, pages 381–400, 2006.
- [Mil56] George Miller. The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information. *The Psychological Review*, 63:81–97, 1956.
- [ptv] PTV AG - <http://www.ptv.de>.
- [YCL05] Missae Yamamoto, Gilberto Câmara, and Luiz Antonio Nogueira Lorena. Fast Point-Feature Label Placement Algorithm for Real Time Screen Maps. In *GeoInfo*, pages 122–138, 2005.

Theory and Implementation of Multi-Context Systems Containing Logical and Sub-Symbolic Contexts of Reasoning

Tarek R. Besold

Chair of Computer Science 8: Artificial Intelligence

University of Erlangen-Nuremberg

sntabeso@i8.informatik.uni-erlangen.de

Abstract: In [Bes09], we propose an extension of the heterogeneous multi-context reasoning framework by G. Brewka and T. Eiter, which, in addition to logical contexts of reasoning, also incorporates sub-symbolic contexts of reasoning. Also a very basic, yet applicable algorithm for solving this task is given. Moreover, we present a proof of concept implementation of an MCS, already integrating different techniques for reducing computational complexity.

1 Introduction

Whenever we write an axiom, a critic can say that the axiom is true only in a certain context. [McC87]

The quotation above summarizes an important problem in knowledge representation and knowledge engineering: There is no way to write globally true statements about realistic problem domains. Multi-context systems (MCS) are a formalization of simultaneous reasoning in multiple contexts. Different contexts are inter-linked by bridge rules which allow for a partial mapping between formulas/concepts/information in different contexts. Recently there have been a number of investigations of MCS reasoning (for instance, see [RS05] or [BRS07]), with [BE07] being one of the latest contributions. There, the authors describe reasoning in multiple contexts that may use different logics locally. Starting from this point, we expanded the notion of multi-context systems and integrated contexts of logical reasoning and contexts of sub-symbolic reasoning into a single MCS. Possible applications of such reasoners are numerous, e. g. could shortcomings of statistical methods be remedied with declarative knowledge and vice versa.

2 Integrating Logical and Sub-symbolic Contexts of Reasoning

Departing from the theory stated in [BE07], we developed the notion of generalized symbolic/sub-symbolic MCS by introducing the crucial concept of a “reasoner” as a replacement for the narrower concept of a “logic” as used by Brewka and Eiter.

Definition 1 A reasoner is a 5-tuple $R = (\mathbf{Inp}_R, \mathbf{Res}_R, \mathbf{ACC}_R, \mathbf{Cond}_R, \mathbf{Upd}_R)$ where \mathbf{Inp}_R is the set of possible inputs to the reasoner, \mathbf{Res}_R is the set of possible results of the reasoner, $\mathbf{ACC}_R : \mathbf{Inp}_R \mapsto 2^{\mathbf{Res}_R}$ defines the actual reasoning (assigning each input a set of results in a decidable manner), \mathbf{Cond}_R is a set of decidable conditions on inputs and results, $\mathbf{cond}_R : \mathbf{Inp}_R \times \mathbf{Res}_R \mapsto \{0, 1\}$, and \mathbf{Upd}_R is a set of update functions for inputs, $\mathbf{upd}_R : \mathbf{Inp}_R \mapsto \mathbf{Inp}_R$.

As also shown in [BM10], this generalized notion additionally covers sub-symbolic reasoners as e. g. Neural Nets. Using the concepts introduced in Definition 1, we were able to also carry over the other relevant notions from [BE07] (as e. g. bridge rules and belief states, up to the concept of equilibrium of an MCS) to the symbolic/sub-symbolic case, adding auxiliary theoretical relining where needed. Conclusively we showed that the newly introduced notion of generalized MCS in fact is a generalization of the heterogeneous nonmonotonic MCS as shown in [BE07].

3 Computing Equilibria for Finite MCS

Using the newly developed theory of generalized MCS, we designed an algorithm for finding the equilibria of finite¹ MCS of the generalized type. As the algorithm is based on complete enumeration and testing, we moreover had to introduce some theoretical tools, amongst which the concept of a “bridge rule model” – offering a possibility to expatiate on the actual reasoning by signaling explicitly which bridge rules within an MCS are active with respect to a given belief state – is the most important one.

Definition 2 Let \mathbf{Br} be a set of n bridge rules of an MCS. A bridge rule model is an assignment $\mathbf{Br} \mapsto \{0, 1\}^n$ that represents for each bridge rule in \mathbf{Br} whether it is active or not.

Using bridge rule models, we may give a complete and correct algorithm for finding the equilibria of a finite MCS: For a given bridge rule model and an MCS we first apply all the bridge rules activated in the bridge rule model yielding $\mathbf{inp}'_1 \dots \mathbf{inp}'_n$. Then we compute the set of results for each context i given \mathbf{inp}'_i by applying $\mathbf{ACC}(\mathbf{inp}'_i)$, yielding a set of results \mathbf{res}'_i for each i , being of finite cardinality as MCS was said to be finite. Thus, testing whether $(\mathbf{inp}_i, \mathbf{res}'_i)$ is an equilibrium for all j , we obtain the set of equilibria for the given bridge rule model. Iterating the procedure over the (finite) set of all bridge rule models and joining the resulting sets of equilibria finally yields the set of all equilibria.

Some first attempts for remedying the complexity problems the naive generate and test approach leverages (e. g. by exploiting structural properties within the set of bridge rules, or introducing a constraint functionality for classes of bridge rules, allowing to explicitly state interdependencies between groups of bridge rules) are presented in [Bes09] as well.

¹From the application point of view, finite MCS are the most considerable subclass of the set of all MCS, as e. g. almost all real life scenarios yield finite only MCS. For details again see e. g. [BM10].

4 Implementation

In order to show the real life applicability of the theory developed, we implemented a proof of concept MCS framework software, mainly based on the principles of the algorithm just stated, enriched with some techniques for reducing computational complexity.

As programming language – due to its high flexibility and interoperability with JAVA and the JAVA VM – we had chosen SCALA. Moreover, inter alia for the implementation of the already mentioned constraint mechanism for bridge rules and bridge rule classes, we made use of the lparse/smodels combination as implementation of the stable model semantics for logic programmes (vide e. g. [Syr00] and [SNS02]).

We applied the MCS framework software in two real life scenarios, an image classification scenario, and a special purpose application example from another research project, inter alia dealing with museum data completion and consistency checking:

In the first application example, we used it for adding time categorization functionality to the original setting, allowing the classifier to also consider the daytime an image was taken, when doing the classification. This yielded a significant improvement (for an α -level of 5%) in the rate of correct classification.

In the second real life scenario, we used the MCS framework software to provide modal time logic-like functionality (allowing for the computation of Allen's relations between time intervals, vide [All83]) in a semantics enhanced content management system, built for the WissKI research project.² We perform a form of inference fusion (e. g. vide [HCA03] and [HAC03]): Starting from a description logics (DL) based initial representation of data, we transfer parts of it (containing temporal information) to another formalism. Then, we use smodels as a reasoner in order to obtain additional results (in concrete for - if possible - establishing an ordering within the expressed time statements), and finally combine the results of the reasoner and the original data into another enhanced DL representation.

5 Conclusion

The work presents a generalization of heterogeneous MCS that allows for the use of sub-symbolic contexts of reasoning alongside logical contexts of reasoning. An exhaustive algorithm for enumerating all equilibria of an MCS is given. On the practice-oriented side, the proof of concept implementation prototypically shows two out of many scenarios where MCS may beneficially be used in real life scenarios.

Concludingly we are convinced that MCS will play a key role in the further integration and advancement of simultaneous reasoning in multiple contexts. We think that also with the further development of the semantic web and related technologies, this kind of information processing will get even more important and will find application in numerous scenarios.

²Research project "WissKI - Wissenschaftliche KommunikationsInfrastruktur", funding provided by the German Research Council (DFG). The project is being carried out at the Chair for Computer Science 8: Artificial Intelligence of the University of Erlangen-Nuremberg, together with collaborators from the GNM Nuremberg and the ZFMK Bonn. By October 31, 2009, with [LKHS09] a presentation article for the project is in press.

Acknowledgements

I want to thank Dr.-Ing. Stefan Mandl for the numerous valuable discussions, his support and supervision during the creation of [Bes09].

References

- [All83] James F. Allen. Maintaining knowledge about temporal intervals. *Commun. ACM*, 26(11):832–843, 1983.
- [BE07] Gerhard Brewka and Thomas Eiter. Equilibria in Heterogeneous Nonmonotonic Multi-Context Systems. In *Proceedings of the National Conference on Artificial Intelligence*, volume 22, pages 385–390. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, 2007.
- [Bes09] Tarek Richard Besold. Theory and Implementation of Multi-Context Systems Containing Logical and Sub-Symbolic Contexts of Reasoning. Master’s thesis, Department Mathematik & Lehrstuhl für Informatik 8: Künstliche Intelligenz, FAU Erlangen-Nürnberg, 2009.
- [BM10] Tarek Richard Besold and Stefan Mandl. Integrating Logical and Sub-Symbolic Contexts of Reasoning. In Joaquim Filipa, Ana Fred, and Bernadette Sharp, editors, *Proceedings of ICAART 2010 - Second International Conference on Agents and Artificial Intelligence*. INSTICC Press, 2010. The full version of the paper is also available under http://www8.informatik.uni-erlangen.de/inf8/Publications/bridging_mcs_original.pdf.
- [BRS07] Gerhard Brewka, Floris Roelofsen, and Luciano Serafini. Contextual Default Reasoning. *Proc. IJCAI-07, Hyderabad, India*, 2007.
- [HAC03] Bo Hu, Ins Arana, and Ernesto Compatangelo. Facilitating DL-based Hybrid Reasoning with Inference Fusion. *Knowledge-Based Systems*, Vol. 16(5):253–260(8), July 2003.
- [HCA03] Bo Hu, Ernesto Compatangelo, and Ins Arana. Inference Fusion: A Hybrid Approach to Taxonomic Reasoning. In I. Russell and S. Haller, editors, *Proceedings of the 16th international FLAIRS conference*. AAAI, 2003.
- [LKHS09] Karl-Heinz Lampe, Siegfried Krause, Georg Hohmann, and Bernhard Schiemann. Wissen vernetzt: Vom Wandel der Dokumentation in Museen der Natur- und Kulturgeschichte. *KI - Künstliche Intelligenz*, 4/09, Special Issue on "Cultural Heritage and A.I.", 2009.
- [McC87] John McCarthy. Generality in Artificial Intelligence. *Communications of the ACM*, 30:1030–1035, 1987.
- [RS05] Floris Roelofsen and Luciano Serafini. Minimal and Absent Information in Contexts. In *International Joint Conference on Artificial Intelligence*, volume 19, page 558. Lawrence Erlbaum Associates LTD, 2005.
- [SNS02] Patrik Simons, Ilkka Niemelä, and Timo Soininen. Extending and implementing the stable model semantics. *Artif. Intell.*, 138(1-2):181–234, 2002.
- [Syr00] Tommi Syrjnen. *Lparse 1.0 User’s Manual*, 2000.

Naïve but Efficient – Using Greedy Strategies for Exploration, Inspection and Search

Dirk Holz
dirk.holz@ieee.org

Abstract: For operating in initially unknown and dynamic environments, autonomous mobile robots need abilities to explore their workspace and construct an environment model as well as to perform searches in that model and re-explore the environment to keep the model up-to-date. This paper focuses on the efficiency of using simple *frontier-based* greedy strategies for exploration and search that provide an autonomous mobile robot with these abilities.

1 Introduction

Autonomous mobile robots need internal representations or *maps* of their environment in order to act in a goal-directed manner, plan actions and navigate effectively. Exploring and mapping are fundamental prerequisites when operating in initially unknown environments (when a map is not available). In addition, when the environment is dynamic, i.e., changes over time, robots need to actively re-explore and inspect their workspace in order to update the map in regions where changes have taken place.

Exploration is related to well-known problems from the field of computational geometry, namely art gallery, illumination and shortest watchmen problems. Since the original art gallery problem is NP-complete [Agg84] and requires complete knowledge about the environment, exploring an unknown environment is usually performed in a reactive or greedy fashion. Instead of planning all locations where the robot needs to acquire sensory information, a greedy exploration strategy solely plans one step ahead by determining a next best view (NBV) that provides new information about the environment while minimizing some objective function. Over the last decades different exploration strategies have been proposed [AG05, BMSS05, SR05, SB03]. Comparative evaluations of different strategies have been presented in [LR97] and [Ami08]. This paper focuses on the efficiency of frontier-based exploration strategies [Yam97] and presents two improvements together with achieved results (Chap. 2) as well as how to apply them in inspection and search tasks (Chap. 3).

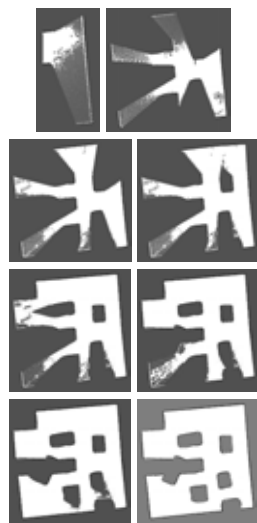


Figure 1: From left to right and top to bottom: frontier cells in the progress of exploration and mapping.

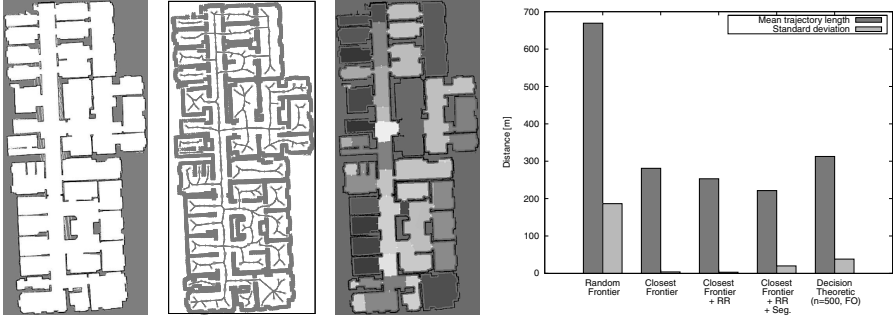


Figure 2: Map segmentation for roomwise exploration. Shown are (from left to right): the (final) input map, the Voronoi diagram with the critical points, the refined segmentation, and the measured path lengths for the different strategies (*RR.* = repetitive re-checking, *Seg.* = map segmentation).

2 Frontier-based Exploration and Extensions

Frontier-based exploration strategies usually operate on grid maps [ME85] that distinguish between known free regions (Fig. 1: white) and unknown regions (Fig. 1: gray). Frontiers (Fig. 1: red/dark gray) are transitions between cells known to be free and unknown regions. Always selecting the frontier being closest to the robot as the NBV, yields a reasonably short total path length while being computationally inexpensive as shown by [KTH01]. However, if the closest frontier does not lie within the same room, a room might need to be explored twice. To account for that, we segment the so far built map into individual rooms and prefer those frontiers that lie in the same room thereby exploring the environment room-wise. Referring to Fig. 2, we 1) construct the Voronoi diagram for the free space in the map, 2) determine *critical points* and 3) split the map at the critical points [Thr98, WSB08]. We define critical points to be local minima with respect to the distances to the closest Voronoi site, nodes of degree 2, and to be itself adjacent to a junction node or adjacent to another node that is adjacent to a junction node since [Thr98] yields too many segments and [WSB08] is too restrictive in the vicinity of doors.

In order to further shorten the robot’s trajectory we repetitively re-check, during navigation, whether the currently approached frontier is still a frontier and start to approach the next NBV if not. As can be seen in Fig. 2, exploring closest frontiers yields a shorter trajectory than the decision-theoretic strategy from [GBL02] and both extensions, *map segmentation* and *repetitive re-checking* further improve the achievable results.

3 Using Greedy Exploration Strategies for Inspection and Search

In contrast to exploration, inspection and searches are carried out when the robot has already fully explored the environment and constructed a complete map. Inspection problems can be solved by 1) solving the corresponding art gallery problem to determine a minimum set of vehicle poses and 2) solving a Traveling Salesman problem to determine the shortest route through the complete set of poses. Although this procedure is likely to

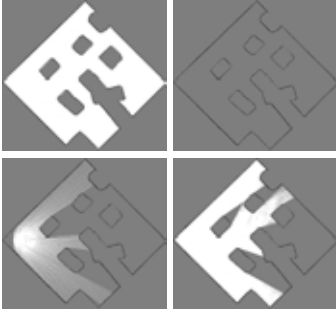


Figure 3: Using frontier-based exploration for inspection. From left to right: fully explored environment, cleared free space, and the map after the first update and after 50% inspection time.

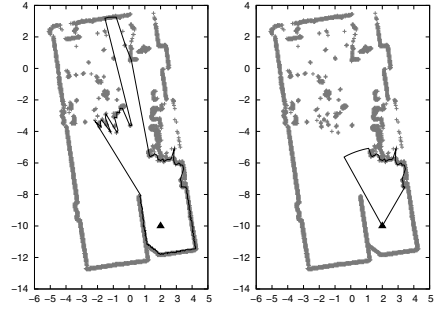


Figure 4: Incorporating field and range of view of the actively used sensor in inspection and search tasks. The map is updated only using the information in the resulting area.

find optimal solutions, it is computationally expensive. Another possibility being computationally efficient is to simply re-explore the environment using the aforementioned frontier-based exploration strategy, though it might not yield optimal solutions. In order to apply it, we simply reset the known free regions in a local copy of the map as being unknown and start exploring and updating the local copy (Fig. 3).

For search tasks, for example when searching for an object using a camera, the limited field of view (and range of vision) can be incorporated by projecting the respective volumes into the map plane and only updating the covered regions (Fig. 4).

4 Results

An example trajectory, recorded in a simulated environment, of a robot consecutively exploring and inspecting is shown in Fig. 5. With 66m and 54m both parts of the robot's trajectory are reasonably short. An example of exploring a real-world environment is given in Fig. 6.

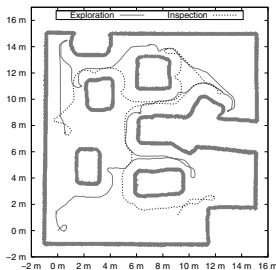


Figure 5: Consecutive exploration and inspection in a simulated environment.

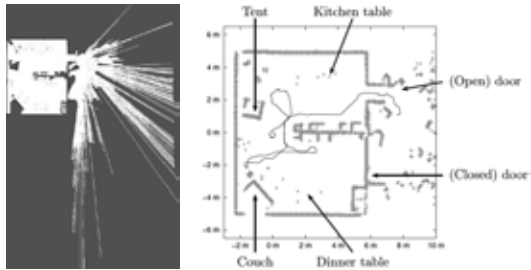


Figure 6: Exploring a real-world environment (the RoboCup@Home arena at the GermanOpen 2009). Left: final map, right: trajectory of the robot.

For mapping we use a matching algorithm that is described, in detail, in [HB10]. Videos of a robot using the algorithms presented herein are available at <http://www.b-it-bots.de/media>. The robot won the RoboCup@Home World Championship 2009.

References

- [AG05] F. Amigoni and A. Gallo. A Multi-Objective Exploration Strategy for Mobile Robots. In *Proc. of the IEEE Int. Conf. on Robotics and Automation (ICRA)*, pages 3861–3866, 2005.
- [Agg84] A. Aggarwal. *The art gallery theorem: its variations, applications and algorithmic aspects*. Ph. D. Thesis, John Hopkins University, Baltimore, MD, USA, 1984.
- [Ami08] F. Amigoni. Experimental Evaluation of Some Exploration Strategies for Mobile Robots. In *Proc. of the IEEE Int. Conf. on Robotics and Automation (ICRA)*, pages 2818–2823, 2008.
- [BMSS05] W. Burgard, M. Moors, C. Stachniss, and F. Schneider. Coordinated Multi-Robot Exploration. *IEEE Transactions on Robotics*, 21(3):376–386, 2005.
- [GBL02] H. H. González-Baños and J.C. Latombe. Navigation Strategies for Exploring Indoor Environments. *International Journal of Robotics Research*, 21:829–848, 2002.
- [HB10] D. Holz and S. Behnke. Sancta Simplicitas – On the efficiency and achievable results of SLAM using ICP-Based Incremental Registration. In *Proc. of the IEEE Int. Conf. on Robotics and Automation (ICRA)*, 2010. Accepted/To appear.
- [KTH01] S. Koenig, C. Tovey, and W. Halliburton. Greedy Mapping of Terrain. In *Proc. of the IEEE Int. Conf. on Robotics and Automation (ICRA)*, pages 3594–3599, 2001.
- [LR97] D. Lee and M. Recce. Quantitative evaluation of the exploration strategies of a mobile robot. *International Journal of Robotics Research*, 16(4):413–447, 1997.
- [ME85] H. Moravec and A. E. Elfes. High Resolution Maps from Wide Angle Sonar. In *Proc. of the IEEE Int. Conf. on Robotics and Automation (ICRA)*, pages 116–121, 1985.
- [SB03] C. Stachniss and W. Burgard. Exploring Unknown Environments with Mobile Robots using Coverage Maps. In *Proc. of the Int. Conf. on Artificial Intelligence (IJCAI)*, pages 1127–1134, 2003.
- [SR05] R. Sim and N. Roy. Global A-Optimal Robot Exploration in SLAM. In *Proc. of the IEEE Int. Conf. on Robotics and Automation (ICRA)*, pages 661–666, 2005.
- [Thr98] S. Thrun. Learning Metric-Topological Maps for Indoor Mobile Robot Navigation. *Artificial Intelligence*, 99(1):21–71, 1998.
- [WSB08] K. M. Wurm, C. Stachniss, and W. Burgard. Coordinated Multi-Robot Exploration using a Segmentation of the Environment. In *Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, pages 1160–1165, 2008.
- [Yam97] B. Yamauchi. A Frontier Based Approach for Autonomous Exploration. In *Proc. of the IEEE Int. Symp. on Computational Intelligence in Robotics and Automation (CIRA)*, pages 146–151, 1997.

Konstruktion von Common Sense-Ontologien durch Analyse natürlicher Sprache

Ruth Janning

FernUniversität in Hagen
Lehrgebiet Wissensbasierte Systeme
rj@jannsite.de

Art der Arbeit: Bachelor of Science Abschlussarbeit

Betreuer/in der Arbeit: Prof. Dr. C. Beierle, Dr. B. Messing

Abstract: Die Begriffe 'Semantic Web', 'RDF' und 'OWL' und damit einhergehend 'Ontologien' sind in den letzten Jahren zu wahren Modeworten geworden. Doch nicht nur speziell im Web werden Ontologien benötigt, sondern allgemein zur Wissensrepräsentation, im Besonderen für künstlich intelligente Anwendungen. Damit diese menschliche Schlussfolgerungen domänenübergreifend nachahmen können, benötigen sie eine Art Hintergrundwissen, welches auch das sogenannte Alltagswissen (*Common Sense*) enthält. Ontologien bieten eine gute Möglichkeit für dessen Repräsentation. Es gibt einige verschiedene Ansätze speziell für die Konstruktion von Common Sense-Ontologien. Doch sind diese meist mit großer 'Fleißarbeit' verbunden. Walid S. Saba, Software Engineer und Knowledge Engineer, sowie an verschiedenen Universitäten in der Lehre im Bereich Computer Science tätig gewesen, versucht dagegen ein Verfahren zur Konstruktion von Common Sense-Ontologien zu entwickeln, mit dem die Struktur von Alltagswissen nicht neu entwickelt, sondern lediglich aufgedeckt werden muss. Dazu benutzt er als Leitfaden die natürliche Sprache. Die vorliegende Arbeit bezieht sich auf einen Artikel, in dem Saba [Sa07] sein Verfahren vorstellt, und schlägt einen Algorithmus zu dessen Implementierung und Analyse vor.

1 Common Sense-Ontologien

Ein Mensch benötigt Hintergrundwissen, um bestimmte rationale Aktionen durchführen zu können. Natürlich kann das menschliche Hintergrundwissen nicht in all seinen Feinheiten in einer maschinennahen Wissensrepräsentation erfasst werden. Es muss von 'irrelevantem' Wissen abstrahiert werden. Ontologien bieten die Möglichkeit, relevantes Wissen zu strukturieren und formal zu repräsentieren, um dieses schließlich einem intelligenten System zur Verfügung zu stellen. Um menschliches Schließen über Domänengrenzen hinweg mit Hilfe von maschinellen intelligenten Systemen nachzubilden, benötigt ein solches System eine Repräsentation des Allgemeinwissens, das ein Mensch sich von Geburt an nach und nach aneignet. Genau diesen Zweck versuchen Common Sense-Ontologien zu erfüllen.

2 Konstruktion von Common Sense-Ontologien nach Saba

Es gibt eine enge Beziehung zwischen dem Verstehen natürlicher Sprache und Wissensrepräsentation mit Schließen. Denn nach dem *Understanding as Reasoning-Paradigma* [Sa07] ist das Verstehen natürlicher Sprache zum größten Teil ein Common Sense Reasoning-Prozess auf einem pragmatischen Niveau. Für das Verstehen von Sprache bzw. die erfolgreiche Realisierung von NLU-Programmen ist daher nach Saba der Gebrauch einer typbasierten ontologischen Struktur notwendig, die unser allgemeines Weltbild und die Art, wie wir darüber sprechen, widerspiegelt. In der Künstlichen Intelligenz gab es schon einige Versuche, große Ontologien aus Alltagswissen zu entwickeln. Besonders zu nennen ist hier CYC [Cy09]. Allerdings musste die große Menge an Alltagswissen, die CYC nun schon beinhaltet, zunächst über Jahre hinweg nach und nach manuell erweitert werden. Saba hingegen nimmt an, dass die Struktur von Alltagswissen 'nur' aufgedeckt und nicht neu erfunden werden muss, wobei die natürliche Sprache als Leitfaden benutzt werden sollte.

Sabas [Sa07] Basisstrategie für die Generierung einer entsprechenden Ontologie beruht auf Frege's *Kompositionalitäts-prinzip* (siehe [Du81]). Nach diesem Prinzip wird der Sinn eines gegebenen Satzes abgeleitet aus unserem Wissen über die Bedeutungen der einzelnen vorkommenden Worte zusammen mit der Untersuchung, wie diese in dem Satz miteinander verknüpft sind. Die Bedeutungen der Worte (bzw. die Konzepte und die korrespondierende ontologische Struktur) und die Beziehungen zwischen ihnen können nach Saba durch eine Analyse ihres Gebrauchs in der täglichen Sprache gewonnen werden. Saba führt ein Prädikat $App(p,c)$ ein, wobei p eine Eigenschaft (Adjektiv) oder eine Aktion (Verb) darstellt und c ein Konzept (Substantiv). Das Prädikat, angewandt auf ein bestimmtes p und ein bestimmtes c , erhält den Wert *wahr*, wenn 'es Sinn macht, von der Eigenschaft bzw. Aktion p von c zu sprechen'. Sind nun zwei Mengen gegeben, von denen die eine Adjektive (bzw. Verben) und die andere Substantive enthält, können die jeweiligen Adjektiv-Substantiv-Kombinationen (bzw. Verb-Substantiv-Kombinationen) mit Hilfe von $App(p,c)$ analysiert werden. Als Ergebnis erhält man schließlich eine Hierarchie, deren Knoten aus Mengen von Substantiven bzw. Konzepten bestehen, die gleiche Eigenschaften besitzen und einem bestimmten Oberkonzept angehören. Der Prozess zur Erschließung der ontologischen Struktur nach Saba [Sa07] kann wie folgt formal beschrieben werden:

1. Es seien eine Menge von Konzepten $C = \{c_1, \dots, c_m\}$ und eine Menge von Aktionen (Eigenschaften) $P = \{p_1, \dots, p_n\}$ gegeben.
2. Außerdem sei ein Prädikat $App(p, c)$ gegeben, wobei $c \in C$ und $p \in P$. $App(p, c)$ erhält den Wert *wahr*, wenn die Aktion (oder Eigenschaft) p auf Objekte vom Typ c sinnvoll anwendbar ist.
3. Für jede Aktion (Eigenschaft) $p \in P$ wird eine Menge $C_p = \{c \mid App(p, c)\}$ generiert, die alle Konzepte c enthält, für die $App(p, c)$ den Wert *wahr* erhält.

4. Anschließend wird die gesuchte Hierarchie durch Analyse der Teilmengen-Beziehungen zwischen den verschiedenen generierten Mengen systematisch entwickelt.

Hat man erst einmal eine ontologische Struktur aufgedeckt, die die beschriebenen Eigenschaften besitzt, wird die Erschließung der Semantik einiger natürlich-sprachlicher Phänomene wie Polysemie, Metaphern, Nominalkomposita und Intensionale Verben sehr einfach.

3 Ontologie-Generierung

Saba [Sa07] erwähnt in seinen Artikel nur die oben genannten 4 Schritte, ohne genauer auf deren Umsetzung einzugehen. Ein Vorschlag für die praktische Umsetzung ist, die Analyse der Konzept-Aktivität/Eigenschaft-Kombinationen mit Hilfe von $App(p, c)$ in Schritt 2 durch eine Tabelle zu visualisieren, die Mengen C_p aus Schritt 3 durch Auslesen der zu p gehörenden Spalte zu gewinnen und Schritt 4 durch einen Algorithmus (siehe Abbildung 1) zu realisieren, der zu einem Knoten (bzw. zu dessen zugehöriger Menge von Konzepten) die GröÙte seiner Teilmengen sucht und diese als linken Sohn des Knotens in die Hierarchie einfügt. Der Grund dafür, dass die größte Teilmenge ausgewählt wird, liegt darin, dass die einen Knoten repräsentierenden Mengen nach unten hin immer kleiner werden sollen, das heißt auf immer speziellere Konzepte abzielen sollen. Der rechte Sohnknoten ist dann die Komplementärmenge des linken Sohnes, bezogen auf die Menge des Vaterknotens. Eine komplexere Frage ist, wie Schritt 1 und 2 des Verfahrens, also die Wahl der Mengen von Adjektiven bzw. Verben und Substantiven und die Erzeugung der entsprechenden Tabelle, realisiert werden können. Vorstellbar ist hier beispielsweise, eine Analyse von vorhandenen Texten (z.B. aus den zahlreichen Quellen des Internets) durch ein maschinelles System durchführen zu lassen, indem es mit Hilfe von Computer-Lexika nach Adjektiven, Verben und Substantiven sucht und eventuell mit Data Mining-Techniken prüft, welche Kombinationen von Adjektiven bzw. Verben mit Substantiven wie häufig vorkommen. Häufig vorkommende Kombinationen könnten dann in der Tabelle an der entsprechenden Stelle ein '+' erhalten.

Bei der Suche nach Beispielen für die vorliegende Arbeit sowie bei Implementierung und Testen des Verfahrens wurde deutlich, dass dessen Anwendung nicht immer ganz eindeutig ist, da es zum Beispiel in manchen Fällen schwer fällt, zu entscheiden, wann ein Adjektiv oder Verb auf ein Substantiv 'sinnvoll anwendbar' ist oder nicht. Eine weitere Schwierigkeit ergibt sich daraus, dass ursprüngliche und übertragene Bedeutungen von Worten (Metaphern) gesondert behandelt werden müssen, d.h. an verschiedenen Stellen in der Hierarchie auftauchen sollen.

Es bleibt tatsächlich noch einiges zu tun, bis das von Saba [Sa07] angestrebte Ziel einer Konzept-Algebra, die mit einer typbasierten Ontologie des Alltagswissens arbeitet, erreicht wird. Man darf gespannt sein, welche Fortschritte weitere Arbeiten in dieser Richtung noch bringen werden.

Methode: generateHierarchy()

Ausgabe: Konzept-Hierarchie

Algorithmus:

1. Initialisierung des Wurzelknotens der Konzept-Hierarchie durch Zuordnung der Menge, die alle Konzepte enthält.
2. Aufruf der rekursiven Methode getSubSets(<Menge>) mit der Menge des Wurzelknotens als Argument:
getSubSets(Wurzelknotenmenge)
3. **return** Konzept-Hierarchie

Methode: getSubSets(<Menge>)

Gegeben: Ermittelte Mengen C_p von Konzepten (siehe Schritt 3 der Ontologie-Generierung)

Eingabe: Menge, deren größte Teilmenge sowie deren Komplementärmenge, bezogen auf die Menge, zu bestimmen und als linker und rechter Sohn einzufügen sind

Algorithmus:

1. Alle echten Teilmengen von <Menge> finden.
2. Größte Menge dieser Teilmengen ermitteln (falls es davon mehrere gibt, diejenige auswählen, deren zugehörige Aktivität/Eigenschaft alphabetisch kleiner ist, als alle anderen).
3. *LinkerSohnknoten* = ermittelte größte Menge.
4. *RechterSohnknoten* = <Menge> \ *LinkerSohnknoten*.
5. **if** *LinkerSohnknoten* $\neq \emptyset$
then
 - *LinkerSohnknoten* als linken Sohn des durch <Menge> repräsentierten Knotens in die Hierarchie einfügen.
 - Teilmengen des linken Sohns ermitteln:
getSubSets(*LinkerSohnknoten*).
 - *RechterSohnknoten* als rechten Sohn des durch <Menge> repräsentierten Knotens in die Hierarchie einfügen.
 - Teilmengen des rechten Sohns ermitteln:
getSubSets(*RechterSohnknoten*).

Abbildung 1: Algorithmus zur Generierung der Hierarchie aus den Konzept-Mengen (Schritt 4)

Literaturverzeichnis

- [Cy09] Cycorp Inc.: Cycorp, Inc.. <http://www.cyc.com/>.
- [Du81] Dummett, M.: Frege: Philosophy of Language. Harvard University Press, Cambridge, MA., 1981.
- [Sa07] Saba, Walid S.: Language, logic and ontology: Uncovering the structure of commonsense knowledge. In: International Journal of Human-Computer Studies 65, März 2007; S. 610-623.

Deskriptives Data-Mining für Entscheidungsträger: Eine Mehrfachfallstudie

Benedikt Kämpgen

benedikt.kaempgen@informatik.uni-wuerzburg.de

Abstract: In der dieser Zusammenfassung zu Grunde liegenden Arbeit wird ein Ansatz zum Entscheidungsträger-verständlichen Deskriptiven Data-Mining vorgestellt und seine Anwendung in zwei Projekten demonstriert. Denn häufig nutzt die Wissensentdeckung in Daten ihr Potenzial nicht aus, was hauptsächlich auf Barrieren zwischen dem Endnutzer und dem Entwicklerteam zurückzuführen ist.

1 Problem und Motivation

Obwohl es sich in verschiedensten Anwendungsbereichen lohnt, Wissen aus Daten zu extrahieren und obwohl mittlerweile große Unternehmen in Data-Mining-Expertisen investieren¹, scheitern bis zu 60% der Projekte (vgl. [MSMFB09]). Als Hauptursache werden falsche Vorstellungen des Endnutzers, der von den Ergebnissen profitieren soll, und falsche Vorgehensweisen des Entwicklerteams beim Realisieren der Ergebnisse (vgl. [CKCK05]) identifiziert. In der vorhandenen Literatur wird das gemeinsame Verständnis von Endnutzer und Entwicklerteam nur unzureichend berücksichtigt. Prozessmodelle wie CRISP-DM [KM06] behandeln nicht, wie Aufgaben des Data-Mining durchgeführt werden. Methodologien wie Catalyst [Pyl03] enthalten solche Informationen, empfehlen jedoch oft veraltete oder nicht frei verfügbare Werkzeuge. Und auch Data-Mining-Fallstudien, z.B. Vorgehensbeschreibungen [BG08], sind entweder zu technisch ausgelegt, um als Endnutzer-verständlich zu gelten oder zu oberflächlich, um für ein Entwicklerteam nachvollziehbar zu sein.

Um die Barrieren zwischen Endnutzer und Entwicklerteam zu vermindern, wird in der Arbeit ein erweitertes Prozessmodell durch eine Methodologie konkretisiert und diese in zwei Projekten mit einem Framework an Werkzeugen angewendet – für eine leichter zugängliche und erfolgreichere Anwendung von Data-Mining-Techniken.

¹IBM's SPSS Deal May Spark BI Market Consolidation, <http://www.kdnuggets.com/news/2009/n16/41i.html>, Januar 2010

2 Entscheidungsträger-verständliche Methodologie

Im Lösungsansatz der Arbeit wird speziell das Beschreiben und Erklären von Daten für Endnutzer thematisiert – Deskriptives Data-Mining für Entscheidungsträger. Dem Ansatz nach stellen die Projektbeteiligten gemeinsam Anforderungen, um Fragestellungen zu einem Problem des Entscheidungsträgers zu behandeln. Eine Anforderung enthält einen Bericht: dieser besteht aus ein oder mehreren Tabellen, deren Inhalte in “reporting type questions” [KMPZ04] beschrieben und auch in Beispielen veranschaulicht werden können. Ein Export eines solchen Berichts bietet eine Repräsentation der Wirklichkeit. Eine Anforderung kann außerdem ein Muster enthalten: es beschreibt durch “deeper analytic questions” [KMPZ04] einen verständlichen Sachverhalt, der innerhalb eines Berichts entdeckt werden soll. Ein konkretes Muster kann Informationen bieten, die der Entscheidungsträger benötigt, um sein Problem zu lösen.

Um dem Entscheidungsträger das Ergebnis in Form einer zusammenhängenden Geschichte [Pyl03, S. 509], unterstützt durch die angeforderten Berichte und Muster, übergeben zu können, sind vier Entwicklungsaufgaben durchzuführen:

- Die Qualität der Daten wird durch Standardverfahren aus der Literatur sichergestellt.
- Ein typisches Data-Warehouse wird zur Speicherung und Abfrage der Daten aufgebaut, auch zum direkten Nutzen für den Entscheidungsträger.
- Die angeforderten Berichte, bestehend aus einer Datenabfrage und Layoutinformationen werden erstellt. Die Berichte werden auf den Daten ausgeführt und exportiert.
- Die angeforderten Muster werden mit Data-Mining-Techniken in den Berichten gesucht und interpretiert.

Der Projektlebenszyklus [GARM⁺08] drückt sich dadurch aus, dass Entwicklungsergebnisse häufig zu iterativen Änderungen an den Anforderungen und Aufgaben führen.

Die Aufgaben werden mittels Komponenten durchgeführt: Mit einer ETL-Komponente lassen sich Daten extrahieren, transformieren und integrieren. Die benötigten Datenbankmodelle (ER-Modell, Multidimensionales Modell [LMTS06]) werden von einer Data-Warehouse-Komponente unterstützt. Eine Reporting-Komponente ermöglicht unter Verwendung von standardisierten Abfragesprachen (SQL, MDX²) die Berichte und ihre Exporte. Auf Berichtsexporte (CSV, ARFF [BFH09]) werden mit Hilfe einer Data-Mining-Komponente Techniken (Diagramme [HK06, S. 54-61], Korrelationskoeffizienten [HK06, S. 67], Subgruppenentdeckung [Atz07, S. 17] etc.) ausführbar und entdeckte Muster interpretierbar. Eine Dokumentationskomponente ermöglicht dem Team, Erfahrungen zu formalisieren und später – auch in weiteren Projekten – zu nutzen. Für jede der Komponenten finden sich Werkzeuge am Closed-Source- und Open-Source-Markt.

²Multidimensional Expressions, [http://msdn.microsoft.com/en-us/library/aa216767\(SQL.80\).aspx](http://msdn.microsoft.com/en-us/library/aa216767(SQL.80).aspx), Januar 2010

3 Anwendung der Methodologie

Relevante Aspekte für potenzielle Entscheidungsträger und Entwicklerteams werden in zwei Fallstudien beschrieben. Im ersten behandelten Projekt werden die im Rahmen des Bologna-Prozesses neu eingeführten Bachelorstudiengänge an der Universität Würzburg mittels elektronisch erfasster Prüfungsdaten bewertet. Nicht zuletzt zahlreiche Bildungsstreiks der Studierenden Ende 2009 haben die Relevanz dieses Themas bestätigt. Fragestellungen, die in diesem Projekt durch Anforderungen behandelt werden, sind beispielsweise:

- 1) Wie entwickeln sich die wichtigsten Kennzahlen der Studiengänge über die Zeit?
- 2) Wie entwickeln sich die wichtigsten Kennzahlen der Prüfungsveranstaltungen?
- 3) Welche relevanten Leistungsdaten weisen die aktuell Studierenden auf?

Die Ergebnisse aus dem Projekt, aus datenschutzrechtlichen Gründen nicht offengelegt, werden den Dekanen der Universität zur Verfügung gestellt und empfehlen weitere Schritte, um die Daten auf den angestrebten Standardisierten Hochschulraum hin zu analysieren.

In einem zweiten Projekt wird der Nutzen eines Lehrsystems der Universität Würzburg anhand von Benutzungsdaten beurteilt. Das über das Internet erreichbare System ermöglicht den Studierenden, den Prüfungsstoff mit praxisnahen Problemfällen der Dozenten zu erlernen. Jedes Jahr muss ein Antrag gestellt werden, um es mittels Studiengebühren finanzieren zu können, weshalb eine solche Leistungskontrolle wichtig ist. Fragestellungen, die unter anderem behandelt werden, sind folgende:

- 1) Wie zufrieden sind die Lernenden mit dem Lehrsystem?
- 2) Werden Hilfsfunktionen des Lehrsystems wie beabsichtigt genutzt?
- 3) Welchen Einfluss hat das Lernen über das System auf Prüfungsergebnisse?

Die letzte Frage konnte dabei nur rudimentär ausgewertet werden. Die Untersuchungen lassen sich für aktuelle Daten wiederholen; bei entsprechender Erweiterung mit dem Potenzial, den jährlichen Finanzierungsantrag zu vereinfachen.

4 Diskussion und Ausblick

Eine gemeinsame Betrachtung der Fallstudien bekräftigt den Schluss, dass die Methodologie auf beliebige Fachbereiche übertragen werden kann. Bereits die zweifache Anwendung offenbart vielversprechende Weiterentwicklungen, projektübergreifende Lessons-Learned und Vorschläge für Best-Practices; den notwendigen Kontext erzeugen die Komponenten. Die verwendeten Werkzeuge stellen ein Framework dar, über das weitere Projekte mit geringerem Aufwand durchgeführt werden können als jeweils über ein individuell programmiertes System, allerdings mit einer begrenzten Flexibilität. Die Verständlichkeit der Methodologie für Entscheidungsträger mittels der Fallstudien zu evaluieren, ist auch aus repräsentativen Gründen nicht möglich. Jedoch stellt ein solches Framework eine Grundlage dar, um die praxisrelevante Data-Mining-Schnittstelle zwischen dem Entscheidungsträger und dem Entwicklerteam weiter zu untersuchen.

Literatur

- [Atz07] Martin Atzmüller. *Knowledge-Intensive Subgroup Mining: Techniques for Automatic and Interactive Discovery*. IOS Press, Amsterdam, The Netherlands, The Netherlands, 2007.
- [BFH09] Remco R. Bouckaert, Eibe Frank und Mark Hall. *WEKA Manual for Version 3-7-0*, 6 2009.
- [BG08] Michael Brydon und Andrew Gemino. Classification trees and decision-analytic feed-forward control: a case study from the video game industry. *Data Min. Knowl. Discov.*, 17(2):317–342, 2008.
- [CKCK05] Krzysztof Cios, Lukasz Kurgan, Krzysztof J. Cios und Lukasz A. Kurgan. Trends in Data Mining and Knowledge Discovery. In In: *Pal N.R., Jain, L.C. and Teoderesku, N. (Eds.), Knowledge Discovery in Advanced Information Systems*, Seiten 200–2. Springer, 2005.
- [GARM⁺08] P. González-Aranda, Ernestina Menasalvas Ruiz, Socorro Millán, Carlos Ruiz und Javier Segovia. Towards a Methodology for Data Mining Project Development: The Importance of Abstraction. In *Data Mining: Foundations and Practice*, Seiten 165–178. 2008.
- [HK06] Jiawei Han und Micheline Kamber. *Data Mining: Concepts and Techniques*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, second. Auflage, 2006.
- [KM06] Lukasz A. Kurgan und Petr Musilek. A survey of Knowledge Discovery and Data Mining process models. *Knowl. Eng. Rev.*, 21(1):1–24, 2006.
- [KMPZ04] Ron Kohavi, Llew Mason, Rajesh Parekh und Zijian Zheng. Lessons and Challenges from Mining Retail E-Commerce Data. *Mach. Learn.*, 57(1-2):83–113, 2004.
- [LMTS06] Sergio Luján-Mora, Juan Trujillo und Il-Yeol Song. A UML profile for multidimensional modeling in data warehouses. *Data Knowl. Eng.*, 59(3):725–769, 2006.
- [MSMFB09] Oscar Marbán, Javier Segovia, Ernestina Menasalvas und Covadonga Fernández-Baizán. Toward data mining engineering: A software engineering approach. *Information Systems*, 34(1):87 – 107, 2009.
- [Pyl03] Dorian Pyle. *Business Modeling and Data Mining*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2003.

Perzeptuelle Verankerung ausgedehnter Landmarken während der instruierten Navigation

Felix Lindner

Universität Hamburg
Department Informatik
lindner@informatik.uni-hamburg.de

Art der Arbeit: Diplomarbeit

Betreuer/in der Arbeit: Dr. Carola Eschenbach, Prof. Dr. Jianwei Zhang

Abstract: Jedes situierte System, das interne Repräsentationen der externen Welt vorhält oder aufbaut, muss das Symbol-Grounding-Problem lösen. Die Verknüpfung von Symbolen und Perzepten, die dieselben physikalischen Objekte in der Welt denotieren, wird in der Literatur unter der Bezeichnung „perzeptuelle Verankerung“ aktuell diskutiert. Der Beitrag skizziert eine Modellierung der grundlegenden Strukturen für die perzeptuelle Verankerung ausgedehnter Landmarken in der Architektur eines instruierten navigierenden Agenten.

1 Motivation

Die Beziehung zwischen Wahrnehmung und internen Repräsentationen der externen Welt [Ha90] spielt insbesondere für die kognitive Robotik eine zentrale Rolle. Um die in einer Wissensbasis symbolisch beschriebenen Entitäten in der realen Welt zu identifizieren, müssen die Symbole mit Repräsentationen auf Perzeptionsebene in referentiell korrekter Weise in Zusammenhang gebracht werden können. Ein Ansatz zur Verknüpfung von Symbolen und Perzepten, die dieselben physikalischen Objekte in der Welt denotieren, wird unter der Bezeichnung *perzeptuelle Verankerung* diskutiert [CS03]. Im Rahmen dieser Diplomarbeit konnten grundlegende Strukturen für die perzeptuelle Verankerung ausgedehnter Landmarken¹ in einem natürlichsprachlich instruierten navigierenden Agenten [TSEHK03] realisiert werden. Mit der perzeptuellen Verankerung erhält der Agent die Fähigkeit, Landmarken während der Navigation wiederzuerkennen, die Geometrien der Landmarken inkrementell zu rekonstruieren und zu einer kognitiven Karte (siehe Abschnitt 3) zusammenzufügen, die als Wissensbasis deliberative Prozesse wie die Aktionsauswahl im Agenten informieren kann.

¹ Landmarken sind Objekte in der Welt, die als Orientierungspunkte während der Navigation genutzt werden können. Alle im Rahmen dieser Arbeit betrachteten Objekte sind Landmarken (Häuser, Bäume, Wege).

2 Ein Verfahren zur multimodalen perzeptuellen Verankerung

Ein instruierter navigierender Agent erhält aus zwei unterschiedlichen Quellen Informationen über seine Umgebung: aus der Instruktion und über seinen Wahrnehmungsapparat. Im Modell, das dem in der Diplomarbeit verfolgten Ansatz zugrunde liegt, werden die Informationen aus diesen beiden Kanälen durch die beiden Datenstrukturen *Instruktionsknoten* (I-Knoten) respektive *Perzeptionsereignis* (P-Ereignis) repräsentiert. Ein I-Knoten enthält eine propositionale Repräsentation einer Landmarke aus der natürlichsprachlichen Instruktion. Ein P-Ereignis enthält neben einer propositionalen Repräsentation der wahrgenommenen Landmarke auch eine geometrische Repräsentation der gesehenen Kontur.

Während der Navigation sind möglicherweise mehrere Wahrnehmungsereignisse mit einer Landmarke assoziiert, die zu einem kohärenten Modell zusammenzufassen sind. Das *Perzeptionsmodell* (P-Modell) einer Landmarke besteht entsprechend aus einem *Perzeptions-Chunk* (P-Chunk), der alle bis dato eingetroffenen P-Ereignisse zu dieser Landmarke enthält und einer geometrischen Repräsentation der Landmarke, die durch eine geeignete Aggregation der in den P-Ereignissen enthaltenen geometrischen Repräsentationen der gesehenen Konturen berechnet wird. Die Historie der P-Modelle wird in einer Datenstruktur namens *Anker* verwaltet. Aus ihm ist der zeitliche Verlauf des Wissenserwerbs zu einer Landmarke ersichtlich. Zusätzlich dient der Anker dazu, die Koreferenz zwischen P-Modell und I-Knoten zu repräsentieren.

Das Beispiel in Abbildung 2.1 zeigt den Zustand eines Ankers in einer Situation s_3 . Zuvor wurden in Situationen s_0 und s_1 dem Anker bereits die P-Ereignisse p_7 und p_{12} zugefügt. In Situation s_3 sind p_{27} und p_{31} neu mit dem Anker assoziiert. Zudem konnte eine Koreferenz zwischen dem P-Modell des Ankers und dem I-Knoten lm_5 festgestellt werden. Der geometrische Anteil des P-Modells setzt sich aus gesehenen Segmenten (hell dargestellt) und imaginierten Segmenten (dunkel dargestellt) zusammen.

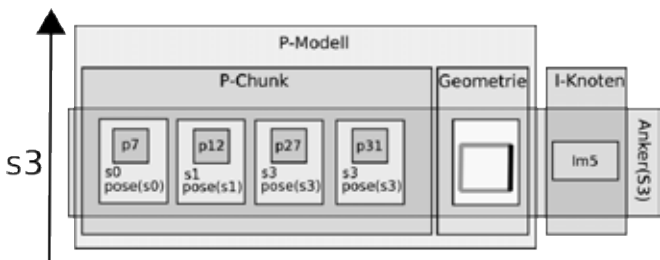


Abbildung 2.1 Die innere Struktur eines Ankers in einer Situation s_3 .

Der Aufbau eines P-Modells basiert auf der Fähigkeit, Landmarken wiederzuerkennen. Für die Entscheidung, welchem Anker ein neues P-Ereignis zugeordnet werden soll, wird die Kompatibilität zwischen P-Ereignis und dem aktuellsten P-Modell eines Ankers berechnet.

Definition (Kompatibilität zwischen P-Ereignis und P-Modell) Ein P-Ereignis ist genau dann kompatibel zu einem P-Modell, wenn die propositionale Beschreibung im P-Ereignis mit der propositionalen Beschreibung im P-Modell kompatibel ist und ein geometrisches Zusammenhangskriterium zwischen dem mentalen Bild des P-Modells und der perzipierten geometrischen Form aus dem P-Ereignis erfüllt ist. \square

Die Kompatibilität der propositionalen Information von P-Ereignis und Anker wird anhand von Ähnlichkeitsmaßen bestimmt, die auch für die Koreferenzauflösung zwischen Instruktion und Perzeption verwendet werden [He03]. Ist die Ähnlichkeit hinreichend hoch, gilt dies als Indiz dafür, dass es sich beim Referenten des P-Ereignisses um eine zuvor schon einmal wahrgenommene Landmarke handelt. Allerdings können verschiedene Landmarken existieren, die durch äquivalente Merkmalsbündel repräsentiert sind. Daher reicht diese Information im Allgemeinen nicht aus, um die Identität einer Landmarke festzustellen [Sa05].

Auf Grundlage der geometrischen Repräsentation des gesehenen Teils der Landmarkenkontur werden geometrische Kriterien angesetzt, anhand derer entschieden werden kann, welchem der Anker das P-Ereignis zuzuordnen ist (oder dass es zu keinem bereits existierenden Anker kompatibel ist). Für die Berechnung der geometrischen Kompatibilität wird zunächst grundsätzlich zwischen solchen Landmarken unterschieden, die geometrisch durch ein geschlossenes Polygon repräsentiert werden (Häuser und Bäume) und solchen Landmarken, die geometrisch durch einen Linienzug repräsentiert sind (Wege). Ein P-Ereignis einer polygonal ausgedehnten Landmarke ist geometrisch kompatibel zu einem P-Modell, wenn sich die geometrischen Repräsentationen von P-Ereignis und P-Modell überlappen (Überlappungskriterium). Das P-Ereignis einer linear ausgedehnten Landmarke ist geometrisch kompatibel zu einem P-Modell, wenn es eine Gerade gibt, so dass die Start- bzw. Endsegmente des Linienzuges des P-Ereignisses und des P-Modells auf dieser Geraden liegen (Kriterium der linearen Fortsetzung).

3 Kognitive Karte zur wissensbasierten Aktionsauswahl

Die in den Ankern aggregierte Landmarkeninformation lässt sich auch als kognitive Karte des instruierten Agenten auffassen. Sie enthält propositionale Information über die perzipierten Landmarken, eine extrapolierte geometrische Repräsentationen der gesehenen Teilstücke sowie den zeitlichen Verlauf der P-Ereignisse.

Anhand dieser Repräsentationen kann der Agent während der Aktionsauswahl Wissen auch über solche Landmarken nutzen, die gegenwärtig nicht im Wahrnehmungsfokus des Agenten liegen. Zudem kann er Annahmen über räumliche Konstellationen machen und gezielte Aktionen zur Informationsbeschaffung ausführen. Des Weiteren erlaubt das Umgebungswissen dem Agenten, verschiedene mögliche Koreferenzauflösungen zwischen natürlichsprachlicher Instruktion und Perzeption zu priorisieren: Ist in der Routeninstruktion eine Landmarkenkonstellation beschrieben, die zu keiner Konstellation aus dem momentanen Wahrnehmungsfokus passt, und kann der Agent aber auf der kognitiven Karte passende Konstellationen finden, so kann er die gesuchte

Konstellation mit einer der Konstellationen aus dem Gedächtnis koreferenzieren. Die Priorisierung bei der Auswahl kann auf Basis einer räumlichen oder einer episodischen Distanz (Zeitpunkt des letzten P-Ereignisses zu dieser Konstellation) geschehen.

4 Zusammenfassung

In jedem situierten System, das Ausschnitte der externen Welt intern repräsentiert, stellt sich die Frage nach einer referentiell korrekten Koreferenzauflösung zwischen den Entitäten des internen Modells und den Entitäten der externen Welt. Ausgehend von einem Rahmenwerk zur perzeptuellen Verankerung wurde im Rahmen dieser Arbeit eine Umsetzung für einen natürlichsprachlich instruierten, navigierenden Agenten geschaffen. Es konnte gezeigt werden, dass die perzeptuelle Verankerung den Navigationserfolg eines instruierten Agenten verbessert. Die Menge der Anker kann als hierarchische kognitive Karte mit einer geometrischen, einer propositionalen und einer episodischen Ebene aufgefasst werden. Das Überblickswissen ermöglicht dem Agenten, über die aktuelle Wahrnehmung hinaus über die Umgebung zu rasonieren.

Das skizzierte Verfahren zur perzeptuellen Verankerung ausgedehnter Landmarken wurde erfolgreich in Simulationsexperimenten evaluiert und konnte anhand eines Robotersystems in einer einfachen realen Umgebungen erprobt werden.

Literaturverzeichnis

- [CS03] Coradeschi, S.; Saffiotti, A. (2003): An introduction to the anchoring problem. *Robotics and Autonomous Systems. Vol. 43, No. 2–3.*, 2003; S. 85–96
- [Ha90] Harnad, S. (1990): The symbol grounding problem. *Physica D: Nonlinear Phenomena*, 42, S. 335–346.
- [He03] Helwich, J. H. (2003): *Graphenbasierte Navigation eines Geometrischen Agenten: Integration von Perzeption und Instruktion*. Diplomarbeit, FB Informatik, Universität Hamburg, Hamburg.
- [Sa05] Santore, J. (2005): *Identifying Perceptually Indistinguishable Objects*. Dissertation, Department of Computer Science and Engineering, University at Buffalo, The State University of New York.
- [TSEHK03] Tschander, L., Schmidtke, H. R., Eschenbach, C., Habel, C., & Kulik, L. (2003): A geometric agent following route instructions. In Freksa, C., Brauer, W., Habel, C. und Wender, K. F., Hrsg.: *Spatial Cognition III*, S. 89–111. Springer, Berlin.

Ein modularer Lösungsansatz für das University-Course-Timetabling-Problem

J. Philippe Neumann
stjeneum@stud.h-da.de

Peter Altenbernd
p.altenbernd@fbi.h-da.de

Fachbereich Informatik, Hochschule Darmstadt

Abstract: Die Erstellung von universitären Stundenplänen ist eine für Menschen komplexe und zeitaufwändige Aufgabe, für die es bereits eine weite Palette an unterschiedlichsten und meist sehr komplexen algorithmischen Ansätzen gibt. Wir möchten hier einen modularen, populationlosen Ansatz vorstellen, der auf der Kombination einfacher Teilalgorithmen beruht, sich durch eine geringe Komplexität auszeichnet und trotzdem hochwertige Ergebnisse erzielt.

1 Einleitung

Das University-Course-Timetabling-Problem gehört zu der Familie der Scheduling-Probleme und betrachtet die Erstellung von Stundenplänen für Veranstaltungen an Universitäten [CBSRD06]. Hierbei ist es das Ziel, sowohl gültige als auch gute Stundenpläne zu erstellen. Die Erstellung solcher Stundenpläne ist oft von einer großen Zahl von Randbedingungen (sogenannten *Constraints*) eingeschränkt, welche verschiedene Prioritäten haben. So dürfen zum Beispiel mehrere Veranstaltungen des gleichen Dozenten nie für die gleiche Uhrzeit eingeplant werden, damit der Stundenplan valide bleibt (sogenannte *Hard-Constraints*). Auf der anderen Seite sollte es beispielsweise vermieden werden, dass Studenten mehr als eine bestimmte Anzahl von Veranstaltungen an einem Tag besuchen müssen. Sollte dies trotzdem eingeplant werden, wird der Stundenplan zwar nicht ungültig, verliert aber an Güte (sogenannte *Soft-Constraints*).

Selbst bei einer geringen Zahl von Constraints und Veranstaltungen ist die Befriedigung dieser eine für einen Menschen äußerst komplexe und zeitaufwändige Aufgabe. Am Fachbereich Informatik der Hochschule Darmstadt umfasst diese Aufgabe jedes Semester um die 100 Dozenten und 200 Veranstaltungen, welche in 30 Räumen über 5 Tage und 7 Slots je Tag verteilt werden müssen [dPHK⁺09]. Aufgrund der hohen Komplexität dieser Aufgabenstellung stellt ein schlichtes Ausrechnen einer Lösung auch bei geringen Problemgrößen keine Option dar. Aus diesem Grund werden primär Heuristiken zur Lösung solcher Probleme verwendet [CBSRD06]. Diese Heuristiken sind oft sehr komplex, benötigen große Mengen an Rechenzeit oder sind nicht flexibel genug für die Stundenplanung im Fachbereich Informatik der Hochschule Darmstadt.

Im Rahmen dieser Arbeit untersuchen wir die Eignung verschiedener Heuristiken für die Lösung des University-Course-Timetabling-Problems. Hierbei konzentrieren wir uns auf

die Kombination einfacher, nicht populationsbasierter Algorithmen und die Ausnutzung ihrer Stärken sowie die Umgehung ihrer Schwächen. Die Beschränkung auf diese Art von Algorithmen ist hierbei durch die unserer Erfahrung nach höhere Laufzeit und Komplexität populationsbasierter Algorithmen motiviert. Die Implementierung des Ansatzes hat sich als dazu in der Lage erwiesen, schnell hochqualitative Lösungen zu erzeugen ohne auf Flexibilität oder geringe Komplexität zu verzichten.

2 Lösungsansatz

Zum Verständnis des Lösungsansatzes ist vorest das Verständnis der verwendeten Heuristiken nötig. *Stochastisches Hill-Climbing* [RSORS] wird zur iterativen Verbesserung einer Lösung genutzt. Hierbei wird die *Fitness* (Güte) der aktuellen Lösung bestimmt. Wir verwenden hierfür die Stärke der Verletzung der Soft-Constraints. Es werden zufällig Änderungen an der Lösung vorgenommen, danach wird wiederum die Fitness bestimmt. Hat sich die Fitness nach der Veränderung nicht verschlechtert, wird die veränderte Lösung zur neuen Ausgangslösung. Hill-Climbing-Algorithmen dieser Art liefern zwar sehr schnell starke Verbesserungen, neigen aber auch dazu, lange vor dem Erreichen einer optimalen oder annähernd optimalen Lösung in lokale Maxima (genannt Fitness-Plateaus) zu geraten. In diesen Plateaus gibt es keine oder nur schwer zu findende verbessernde Lösungen. Es ist daher notwendig, eine schlechtere Lösung zu akzeptieren, um in einen neuen Zustand zu gelangen [YdlM93]. Eine weitere verwendete Heuristik ist die *simulierte Abkühlung* (*Simulated Annealing* [KGV83]), welche ähnlich dem Hill-Climbing arbeitet, aber auch schlechtere Lösungen akzeptieren kann. Die Wahrscheinlichkeit, dass eine schlechtere Lösung akzeptiert wird, verringert sich allerdings im Laufe der Zeit.

Um unsere Ziele zu erreichen, haben wir uns für einen mehrphasigen Ansatz entschieden, dessen Teilalgorithmen leicht ersetzt werden können, sollten sie nicht die gewünschte Leistung erbringen. Die Motivation für die Wahl eines mehrphasigen Ansatzes war hierbei die starke Vereinfachung des Algorithmus. Hierbei werden in der ersten Phase alle Hard-Constraints befriedigt und spätere Phasen dürfen diese nicht mehr verletzen. Dies schränkt zwar den Suchraum ein, vereinfacht das gesamte weitere Vorgehen allerdings und vermeidet es, gute Lösungen zu erzeugen, die ihre Güte nur dadurch erreichen konnten, dass sie nicht valide sind.

In der ersten Phase werden die vorgegebenen Veranstaltungen mit dem Ziel, einen gültigen Stundenplan mit vernachlässigbarer Güte zu erhalten, in den Stundenplan einsortiert. Es werden also ausschließlich Hard-Constraints und keine Soft-Constraints beachtet. Wir wenden hierzu verschiedene, meist randomisierte Algorithmen an und übernehmen das Ergebnis des jeweils besten.

In Phase 2 wird der gültige Stundenplan aus Phase 1 durch stochastisches Hill-Climbing zu einem guten Stundenplan weiterentwickelt. Hierzu werden zwei grundlegende Operationen zur zufälligen Veränderung des Stundenplans eingesetzt. *Swap Operation* vertauscht die Platzierungen von zwei Veranstaltungen. *Replace Operation* entfernt eine Veranstaltung und versucht, diese an einem anderen Ort neu zu platzieren. Als dritte Operation

nutzen wir die neuartige, abwechselnde Anwendung beider grundlegender Operationen (*Fast Operation Cycling*), was wiederum als Grundlage für Simulated Annealing dient.

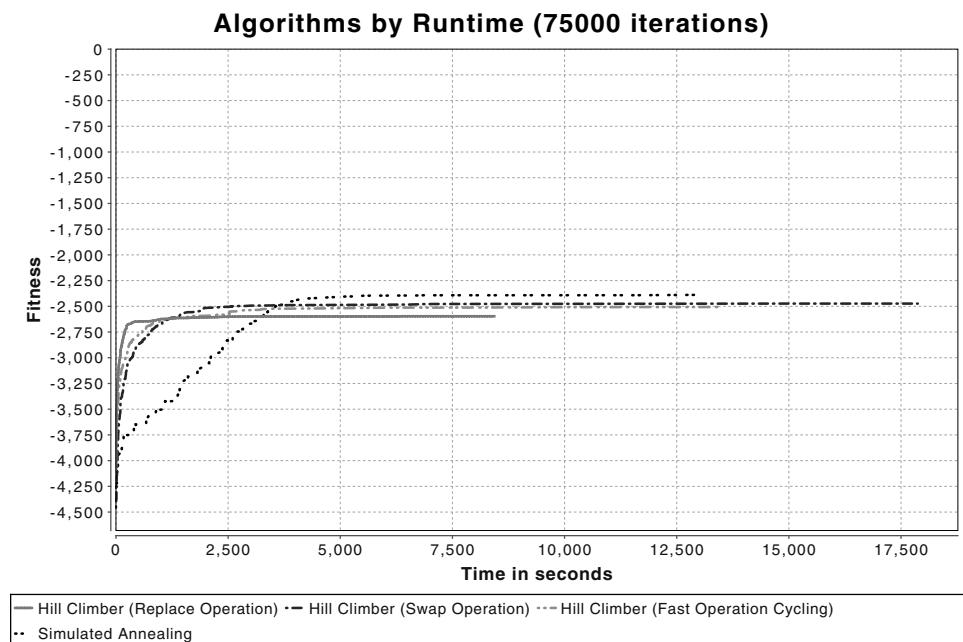


Abbildung 1: Entwicklung der Fitness der verschiedenen Ansätze. Die Fitness stellt hierbei die Summe der Soft-Constraint-Verletzungen da. Eine Fitness von 0 kann hierbei nie erreicht werden, da dafür alle Soft-Constraints verletzungsfrei sein müssten, diese aber teils entgegengesetzte Ziele haben und zuwenige Slots zur Verfügung stehen.

Wie in Abbildung 1 erkennbar, erreichen beide grundlegenden Operationen unterschiedliche lokale Maxima, wobei Swap Operation immer deutlich bessere Maxima erreicht. Auffällig hierbei ist, dass Fast Operation Cycling und das darauf basierende Simulated Annealing immer mit Swap Operation vergleichbare Ergebnisse erreichen, dafür aber deutlich weniger Zeit benötigen als die zugrunde liegenden Operationen.

3 Fazit und Ausblick

Durch die Verbindung einfacher Heuristiken in einem mehrphasigen System ist es uns nach vergleichsweise kurzer Entwicklungszeit gelungen, hochqualitative Lösungen für ein komplexes Scheduling-Problem zu erhalten. In Zukunft hoffen wir das System durch diverse technische und algorithmische Verbesserungen zu beschleunigen. Zusätzlich hoffen wir durch den Einsatz von erschöpfenden Suchen und der Kombination von Simulated Annealing und Random-Restart-Hill-Climbing [ID07] bessere Fitnesswerte zu erhalten.

Literatur

- [CBSRD06] Marco Chiarandini, Mauro Birattari, Krzysztof Socha und Olivia Rossi-Doria. An effective hybrid algorithm for university course timetabling. In *Journal of Scheduling*, Jgg. 9, Seiten 403–432, 2006.
- [dPHK⁺09] Alexander del Pino, Philip Heck, Christopher Klink, Johannes Klug, J. Philippe Neumann und Nicolai Reuschling. Constructing University Timetables using an Extended Artificial Immune System. In *Interactive Systems and Technologies*, 2009.
- [ID07] David Iclanzan und Dan Dumitrescu. Overcoming hierarchical difficulty by hill-climbing the building block structure. In *GECCO '07: Proceedings of the 9th annual conference on Genetic and evolutionary computation*, Seiten 1256–1263, New York, NY, USA, 2007. ACM.
- [KGV83] S. Kirkpatrick, CD Gelatt und MP Vecchi. Optimization by simulated annealing. *Science*, 220(4598):671–680, 1983.
- [RSORS] A. Rosete-Suárez, A. Ochoa-Rodríguez und M. Sebag. Automatic graph drawing and stochastic hill climbing. In *Proceedings of the Genetic and Evolutionary Computation Conference*, Jgg. 2, Seiten 1699–1706.
- [YdlM93] Deniz Yuret und Michael de la Maza. Dynamic Hill Climbing: Overcoming the limitations of optimization techniques. In *The Second Turkish Symposium on Artificial Intelligence and Neural Networks*, Seiten 208–212, 1993.

Umsetzung eines auf Potentialfeldern basierenden Pfadplanungsalgorithmus für die Jenaer Robot Soccer Engine

Max Schneider

Max.Schneider@informatik.uni-erlangen.de

Abstract: Das Fußballspiel an sich verlangt von dem Spieler Fähigkeiten zum Planen, Lernen, Reagieren und die Fähigkeit zum Teamspiel. Beim Roboterfußball kommen noch technische Aspekte wie Sensorik, Motorik, aber auch für Menschen selbstverständliche Prozesse wie Selbstlokalisierung, Pfadplanung und dergleichen ins Spiel. Die am Lehrstuhl für Rechnerarchitektur der Friedrich-Schiller-Universität entwickelte Jenaer Robot Soccer Engine stellt eine virtuelle Umgebung zur Verfügung, in die ein reales Spielfeld mit den sich darauf befindenden Objekten (Roboter, Spielball) abgebildet wird. Anhand der durch ein globales Kamerasystem aufgenommenen Bildern werden Informationen wie Position und Ausrichtung der Roboter oder des Spielballs an ein Strategieplanungsmodul geliefert. Dieses Modul wertet die Daten aus und generiert anhand von entsprechenden Strategiefunktionen Befehle für die Roboter eines Teams, durch die ein "intelligentes" Verhalten der Roboter in der jeweiligen Spielsituation erreicht werden soll. Eine grundlegende Funktion der Strategieplanung ist die Planung eines kollisionsfreien Pfades. Für die Jenaer Robot Soccer Engine wurde ein auf Potentialfeldern basierender Pfadplanungsalgorithmus auf seine Eignung für die Erstellung von kollisionsfreien Pfaden in Echtzeit untersucht und implementiert. Dieser Algorithmus wird in diesem Papier vorgestellt.

1 Potentialfeldgeführte Pfadplanung

Das grundlegende Prinzip des Potentialfeldverfahrens basiert auf der Repräsentation des Roboters, für den ein Pfad bestimmt werden soll, als ein Teilchen in einer als Potentialfeld bezeichneten Umgebung. Die in diesem Potentialfeld wirkenden Kräfte beeinflussen die Bewegungen des Teilchens derart, dass dieses einem kollisionsfreien Pfad folgend, zu seiner Zielposition findet. Die allgemeine Vorgehensweise basiert auf folgenden Schritten: Zunächst wird die Zielposition zum Ursprung eines attraktiven Potentials gemacht, dessen resultierende Kraft den Roboter von jeder Position innerhalb der Versuchsumgebung zu der Zielposition zieht. Das attraktive Potentialfeld ist dabei so aufgebaut, dass der Ursprung ein globales Minimum darstellt und die Potentiale der anderen Positionen mit der zunehmenden Entfernung von dem Ursprung ansteigen. Da sich Hindernisse in der Umgebung des Roboters befinden, müssen diese auch im Potentialfeld berücksichtigt werden. Dies geschieht durch die Einführung lokal wirkender abstoßender (hoher) Potentiale. Durch die lokale Wirkung der repulsiven Potentiale, wird der Roboter nur in der Nähe von Hindernissen durch die resultierende Kräfte abgestoßen und umfährt so die Objekte. Das

anziehende und die abstoßenden Potentiale werden kombiniert und ergeben das die reale Umgebung des Roboters repräsentierende Potentialfeld [Lat91].

Anhand der generierten Potentiale und der daraus resultierenden Kräfte wird anschließend ein Pfad für den betreffenden Roboter generiert. Das Vorgehen dabei ist iterativ und läuft folgendermassen ab: In der Anfangsposition q_{start} wird mittels des negativen Gradienten der Potentialfunktion die auf den Roboter in dieser Position wirkende Kraft F bestimmt. Diese Kraft gibt die Bewegungsrichtung für den Roboter an. Aus der Kombination der aktuellen Position und des Kraftvektors wird nun die nächste Position q' berechnet. Danach wird die in der Position q' wirkende Kraft F' bestimmt und aus der Kombination der beiden Werte die nächste Position q'' berechnet. Dies wird solange wiederholt, bis eine Position mit einem minimalen Potentialwert erreicht wurde. Bei diesem Minimum handelt es sich im besten Falle, um das globale Minimum, das der Zielposition entspricht. Es kann sich aber auch um lokale Minima handeln, die durch Überlagerung der Einzelpotentiale in der Nähe von Hindernissen entstehen können [Lat91, S. 295]. Führt der Gradientenabstieg in ein lokales Minimum, wird eine Routine zu deren Behandlung aktiviert. Hierbei wird die Nachbarkonfiguration des lokalen Minimums mit nächstgrößerem Potential ausgewählt und überprüft, ob der Gradientenabstieg wieder zum lokalen Minimum führt oder nicht. Wenn ja, wird die Position als "besucht" markiert und der nächste Nachbar untersucht. Zeigt die in der aktuellen Konfiguration wirkende Kraft in eine andere Richtung, wird die Pfadplanung mit Gradientenabstieg an dieser Konfiguration fortgesetzt [Deu04, Top99].

1.1 Zusätzliche Potentiale

Neben den beiden grundlegenden Potentialen, lässt sich die Pfadplanung durch folgende Potentiale verbessern. Zunächst kann das Orientierungspotential einbezogen werden, mit dem der Pfadverlauf so beeinflusst werden kann, dass der Roboter in der Zielposition automatisch eine bestimmte Ausrichtung einnimmt, ohne dass eine zusätzliche Rotation erforderlich ist. Desweiteren können die aktuellen Ausrichtungen der Hindernisse, aber auch des Roboters selbst in die Pfadplanung einbezogen werden, um die Wahrscheinlichkeit von Kollisionen - die in dynamischen Umgebungen, wie dem Robot Soccer nicht vollständig vermieden werden können - weiter zu reduzieren. Die Einbeziehung der Ausrichtung der beweglichen Hindernisse führt dazu, dass die generierten Pfade, hinter diesen Objekten und nicht direkt durch deren Fahrbahn verlaufen. Die Berücksichtigung der Ausrichtung des Roboters begünstigt die Pfadplanung in der Art und Weise, dass ein Hin- und Herschwingen, zwischen zwei um ein stationäres Hindernis führende Pfade, vermieden wird, indem der aktuell genommene Pfad bevorzugt wird [Top99].

1.2 Inverses Planen

Je nach Konstellation der Objekte auf dem Spielfeld können generierte Pfade starke Krümmungen aufweisen, so dass sie nicht effizient genug durch den Roboter abgefahren werden

können. Liegt z.B. ein Hindernis auf einer direkten Linie zwischen dem Roboter und seiner Zielposition, führt die potentialgeführte Pfadplanung auf der Basis des negativen Gradientenabstieges zunächst direkt auf das Ziel - und somit auch in Richtung des Hindernisses - zu. Sobald jedoch der Einflussbereich des repulsiven Potentials erreicht wird, kommt es zu einem Knick im Pfad. Dies liegt daran, dass der Gradientenabstieg nicht in Richtung hoher Potentiale - in Richtung Hindernismitte - verläuft, sondern entlang der Grenze des Einflussbereiches. Erreicht die Pfadplanung eine Konfiguration q_{frei} in der keine repulsive Kraft die Fortbewegung in Richtung des Ziels blockiert, erfolgt wieder der direkte Abstieg zum Ziel. Um die Generierung solcher ineffizienter Pfade zu verhindern, erfolgt die Pfadplanung im inversen Sinne. D.h. anstatt die Zielposition als Ursprung des attraktiven Potentials zu wählen, wird die Position der Roboters als das Ziel und die eigentliche Zielposition als Startpunkt der Pfadplanung angenommen. Hierbei entsteht ein Knick in Hindernisnähe auf der Seite des eigentlichen Zielpunktes, dies spielt jedoch keine Rolle, da die Pfadplanung für jedes neue von der Kamera der Robot Soccer Engine gelieferte Bild neu durchgeführt wird. Somit wird der Pfad immer wieder aktualisiert und wenn der Roboter die Position q_{frei} erreicht, führt ein neu generierter Pfad (mit Startpunkt in der Zielposition und dem Zielpunkt in Position q_{frei} auf direktem Wege zum Ziel.

1.3 Ergebnisse

Die Effizienz der in C++ implementierten potentialfeldgeführten Pfadplanung wurde in mehreren Testphasen in einem System mit einem Intel DualCore 2.13 GHz mit 1 GB Arbeitsspeicher untersucht. Dabei wurden verschiedene Kombinationen von Faktoren wie der Reichweite des Einflussbereiches des repulsiven Potentialfeldes, der Diskretisierungsstufe¹, dem Hinzunehmen von Ausrichtungspotentialen und des Potentials für die Spielfeldgrenzen getestet. Tabelle 1 stellt einige der im Durchschnitt erzielten Messzeiten bei der Generierung von Pfaden, zwischen zwei Punkten, die sich an entgegengesetzt gelegenen Seiten des Spielfeldes befinden, dar.

Diskretisierungsstufe	Gemessene Zeiten
3	17 ms
5	7.44 ms
7	5.27 ms

Tabelle 1: Ausführungszeiten der Potentialfeldfunktion. Gemessen in einem System mit einem Intel DualCore 2.13 GHz mit 1 GB Arbeitsspeicher.

¹Die Diskretisierungsstufe gibt an wieviele Punkte eines Feldes in einem Punkt des diskretisierten Feldes zusammengefasst werden, z.B. gibt Diskretisierungsstufe Drei an, dass in einem Punkt des diskretisierten Feldes eine Fläche von 3×3 Punkten des ursprünglichen Feldes vereint werden. Bei der in der Jenaer Robot Soccer Engine verwendeten Spielfeld-Auflösung von 600×400 Pixel führt eine Stufe Drei Diskretisierung zu einem Feld der Größe 200×133 Pixel.

2 Zusammenfassung

Ausgehend von den erzielten Ergebnissen lässt sich schlussfolgern, dass die auf der Überlagerung von Einzelpotentialen beruhende Erstellung eines Potentialfeldes und die anschließende Generierung eines Pfades, sehr schnell durchgeführt werden können, so dass immer ein aktueller und gültiger Pfad zur Verfügung steht. Desweiteren lässt die Erweiterbarkeit der potentialfeldgeführten Pfadplanung, durch bestimmte Eigenschaften beschreibende Potentiale (Ausrichtung des Roboters, Orientierung im Ziel usw.), das Hinzunehmen bzw. Entfernen von Potentialen zu, um auf bestimmte Spielsituationen angepasste Pfade erstellen zu können.

Die interne Struktur der benötigten Potentialfelder bietet neben der Erweiterbarkeit um weitere Potentiale, zusätzlich die Möglichkeit der Parallelisierung der Potentialberechnung, wodurch die Performance des Algorithmus weiter gesteigert und die Ausführungszeiten somit reduziert werden können. So wurde in einem weiteren Projekt der potentialfeldgeführte Pfadplanungsalgorithmus mittels CUDA (Compute Unified Device Architecture) [CUDA] auf Grafikkarten der Tesla-Reihe von NVIDIA umgesetzt. Auf einer Tesla C1060 benötigte die Generierung von Pfaden in einem 640×640 großen Feld ohne eine zusätzliche Diskretisierung maximal vier Millisekunden. Für Robot Soccer lässt diese Optimierung also einen noch größeren zeitlichen Rahmen für die Ausführung anderer Funktionen, wie z.B. der Berechnung von Ballabfangpositionen, optimalen Schusspositionen usw.

Literatur

- [CUDA] CUDA - Zone, Homepage, [http : //www.nvidia.de/object/cuda_learn_de.html](http://www.nvidia.de/object/cuda_learn_de.html)
- [Deu04] Deutsch, C.: Pfadplanung für einen autonomen mobilen Roboter, Master's Thesis, Technische Universität Graz, Fakultät der Elektrotechnik, 2004
- [Lat91] Latombe, J.-C.: Robot Motion Planning, Kluwer Academic Publishers, 1991
- [Top99] Topor, A.: Roboterfußball: Pfadplanung in dynamischer Umgebung, Master's Thesis, Universität Freiburg, Fakultät für Angewandte Wissenschaften, 1999

XML Import/Export für FBD in OpenPCS

Josef H.W.Geller

Wilhelm Büchner Hochschule
Fachbereich Elektrotechnik
josefhwg@yahoo.de

Art der Arbeit: Diplomarbeit
Betreuer der Arbeit: Dr. Wolfgang Kliesch

Abstract: Im Bereich der Speicherprogrammierbaren Steuerung (SPS) stehen dem Anwender neben zahlreicher Hardware auch verschiedene Entwicklungsumgebungen zur Auswahl. Jedoch führen proprietäre Datenformate in den zahlreichen Entwicklungsumgebungen für SPS-Programme zu einer geringen Portierbarkeit der Anwendungssoftware untereinander. Ein von der IEC 61131-3 entwickeltes XML-Format soll dabei Abhilfe schaffen. Aus diesem Grunde wurde es im Zuge dieser Arbeit für die Entwicklungsumgebung OpenPCS der Firma infoteam Software GmbH angewandt. Das Resultat daraus ist ein Konverter, der Programme in Funktionsbausteinsprache zwischen dem proprietären Format von OpenPCS und dem Standardformat hin und her transformieren kann.

1 ,XML Format for IEC 61131-3' für FBS

Es gibt eine Vielzahl an Herstellern, die sich im Bereich der SPS einen Namen gemacht haben. Dadurch stehen dem Anwender der Industrieautomatisierung neben zahlreicher Hardware auch mehrere Entwicklungsumgebungen zur Verfügung.

Die meisten dieser Entwicklungsumgebungen orientieren sich an der Norm IEC 61131-3 und weisen somit viele Parallelen und gemeinsame Grundkonzepte auf. Diese finden sich auch in den darin beschriebenen Programmiersprachen wieder: Anweisungsliste (AWL), Strukturierter Text (ST), Kontaktplan (KOP), Ablaufsprache (AS) und Funktionsbausteinsprache (FBS). AWL ist textbasiert und wird als maschinennahe Programmiersprache eingestuft, ähnlich wie Assembler. Bei AWL wird jede Anweisung in einer separaten Zeile angelegt (vgl. Wellers, 1996, S.60). ST hingegen ist den höheren Programmiersprachen zuzurechnen, da seine Befehle umfangreicher und mächtiger sind als bei AWL (vgl. Wellers, 1996, S.68).

Zu den grafischen Programmiersprachen gehören KOP, AS und FBS. Während KOP die Darstellung elektromechanischer Relaisysteme nachahmt und vorwiegend mit booleschen Werten arbeitet (vgl. Wellers, 1996, S.75), wird AS angewendet, wenn Steuerungen zu realisieren sind, bei denen ein schrittweiser Ablauf mit Übergangsbedingungen zu programmieren ist (vgl. Wellers, 1996, S.85).

Die grafischen Programmiersprachen verwenden keine Anweisungen im herkömmlichen Sinn. Hier steckt die Programmierung in der Verknüpfung der Grafikelemente.

Die Speicherung der mithilfe grafischer Programmiersprachen erstellten Programme wird von jedem Hersteller in einem proprietären Format realisiert. Ein direkter Austausch oder die Wiederverwendung von bereits implementierter Software ist also mit einer anderen Entwicklungsumgebung praktisch nicht möglich. Um die Portierbarkeit von Anwendungsprogrammen zwischen den verschiedenen Entwicklungsumgebungen dennoch zu ermöglichen, ist ein standardisiertes Format nötig. Die Organisation PLCopen – verantwortlich für die Weiterentwicklung des IEC 61131-3 Standards – veröffentlichte im Jahre 2005 die erste Version (V1.01) dieses Einheitsformats („XML Formats for IEC 61131-3“, Baier, Thomas u.a.). Das Technische Komitee ‚TC6‘ entwickelte es innerhalb von PLCopen weiter, so dass 2008 die Version 2.0 und schließlich 2009 die Version 2.01 veröffentlicht wurden. (vgl. John, Tiegkamp, 2009, S.17ff).

Den verschiedenen Herstellern steht nun die Möglichkeit offen das XML-basierte Standardformat in ihre Entwicklungsumgebungen zu implementieren.

2 FBDs in OpenPCS

Das Ziel dieser Arbeit ist es, das Standardformat für die Programmierung in FBS für die Entwicklungsumgebung OpenPCS der Firma infoteam Software GmbH nutzbar zu machen. Die in FBS erstellten Programme nennt man auch Funktionsbaustein-diagramme (FBDs). Folgendes Beispiel zeigt die wesentlichen Merkmale eines FBD wie Funktionsbausteine, Variablen und deren Verknüpfungen.

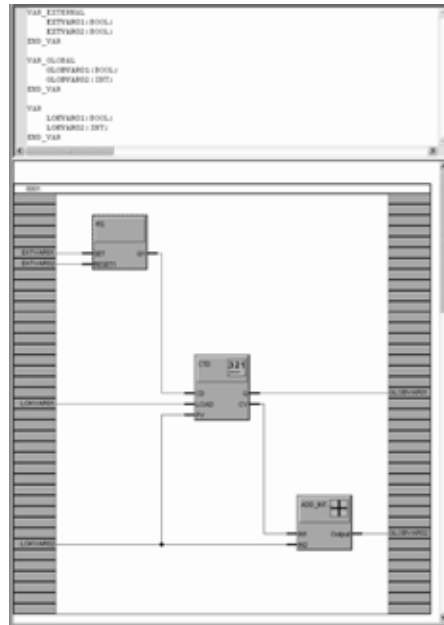


Abbildung 1: Funktionsbausteindiagramm (FBD)

Das originäre Format zum Speichern von FBDs in OpenPCS basiert auf einer ASCII-Serialisierung der intern genutzten Klasseninstanzen der Entwicklungsumgebung selbst. Die darin enthaltenen Parameter konzentrieren sich auf logische Zusammenhänge des Anwenderprogramms und verzichten großteils auf grafische Informationen; dies stellt einen entscheidenden Unterschied zum Standardformat dar, das grafisch viele Freiheiten ermöglicht. Außerdem beruhen die Referenzen, die die logischen Verknüpfungen herstellen, in beiden Formaten auf verschiedenen Konzepten. Es reicht also nicht aus, vorhandene Daten und Parameter nur in einer anderen Struktur abzulegen; sie müssen zuerst transformiert und generiert werden.

3 Import und Export in OpenPCS

Die Möglichkeit, FBDs im standardisierten XML-Format zu exportieren bzw. zu importieren, wurde über einen Konverter realisiert. Implementiert wurde dieser mit C# in Visual Studio als kommandozeilenbasierte Applikation. Dadurch kann sie eigenständig genutzt, aber auch einfach in OpenPCS – der Entwicklungsumgebung, für die sie geschaffen wurde – integriert werden. Es konnte gezeigt werden, dass die einzelnen Elemente eines FBDs im proprietären und standardisierten Format grundsätzlich äquivalent sind, wenngleich die Abbildung des einen Formats auf das andere mehrstufig erfolgen muss und unter anderem einen eigenen Parser für das proprietäre Format benötigt; auch dieser wurde im Rahmen der Arbeit implementiert. Die Klassen zum Einlesen und Abspeichern wurden analog zu den Klassenbibliotheken für Ein- und Auslesefunktionen in C# gehalten, wobei die internen Hilfsklassen auf zwei verschiedenen Ansätze beruhen. Zum einen werden beim Export (in das XML-Format) die einzelnen FBD-Elemente in Instanzen spezieller Klassen aus XML-Knoten geschrieben und dann wie einzelne Äste an den gesamten XML-Baum angefügt. Dabei werden auch zusätzlich notwendige Parameter generiert. Zum anderen werden beim Import (in das proprietäre Format) die Vorlage des Formats und die darin vorgegebene Struktur in einer eigenen Konfigurationsdatei abgelegt; diese ist wiederum in XML geschrieben. Beim Konvertierungsvorgang wird somit aus der Verknüpfung zweier XML-Dateien – der zu übersetzenden Datei und der Konfigurationsdatei – das ursprüngliche Format erzeugt: Es werden die Stärken von XML in beiden Fällen genutzt.

Während des ausführlichen Tests des Konverters konnten die Konzepte verifiziert werden. Für den Export stellte sich die Entwicklungsumgebung Multiprog von KW-Software als besonders nützlich heraus, da sie als bisher einzige verfügbare Programmierungsumgebung das Standardformat unterstützt. In PLCopen erstellte Anwendungsprogramme konnten vollständig in Multiprog eingelesen werden. Beim Import bewährte sich das vorgestellte Konzept, wenn auch die lückenlose Abbildung in das proprietäre Format eine Herausforderung bleibt.

Der flexiblen Adaption von proprietären Formaten steht die Interoperabilität von Standardformaten gegenüber; somit ist langfristig auch in der SPS-Programmierung eine weitere Koexistenz beider Lösungen sehr wahrscheinlich.

Literaturverzeichnis

Bücher

Baier, Thomas u.a.: *Technical Paper. PLCopen Technical Committee 6. XML Formats for IEC 61131-3*, 2. Aufl., Gorinchem: PLCopen 2008.

John, Karl Heinz und Tiegelkamp, Michael: *SPS-Programmierung mit IEC 61131-3. Konzepte und Programmiersprachen, Anforderungen an Programmiersysteme, Entscheidungshilfen*, 4. Aufl., Berlin, Heidelberg: Springer Verlag 2009.

Wellers, Hermann: *SPS- Programmierung nach IEC 1131-3: Normung, Beispiele, Übungen*, Berlin: Cornelsen 1996.

Weiterführende Literatur

St.Laurent, Simon und Fitzgerald, Simon: *XML kurz & gut*, 3.Aufl., Köln, O`Reilly Verlag 2006.

Zacher, Serge: *SPS-Programmierung mit Funktionsbausteinsprache. Automatisierungstechnische Anwendungen*, Berlin und Offenbach: VDE Verlag, 2000.

infoteam Software GmbH: *Datenblatt*, (2008), URL: <http://www.infoteam.de/index.php?id=293>, abgerufen am 01.07.2009.

PLCopen: *XML Logo*, (2008), URL: http://www.plcopen.org/pages/tc6_xml/, abgerufen am 07.07.2009.

Verifikation von Protokollverfeinerungen

Christian Heinzemann

chris227@upb.de

In den letzten Jahren stellt sich vermehrt die Anforderung an Software, sich ihrer veränderlichen Umwelt dynamisch anzupassen. Man spricht dabei von Selbstadaption oder Selbstmanagement. Diese zusätzliche Anforderung führt zu einem weiteren Anstieg der Komplexität von Software ([KM07]). Gleichzeitig findet eine immer stärkere Vernetzung von einzelnen Softwarekomponenten statt. Dadurch entsteht weitere Komplexität. Dies gilt insbesondere auch für Software-intensive Systeme ([GHH⁺08, GH06]). Im Rahmen dieser Arbeit sollen mechatronische Systeme als ein Repräsentant Software-intensiver Systeme betrachtet werden. Mechatronische Systeme sind Systeme des Maschinenbaus, die umfangreiche Funktionalität umsetzen und zur Erfüllung ihrer Aufgaben Elemente aus der Mechanik, der Regelungstechnik, der Elektrotechnik und der Informationstechnik integrieren. Insbesondere die Informationstechnik ist ein Innovationsträger, der es erlaubt, komplexe Funktionen umzusetzen. Beispiele für vernetzte, mechatronische Systeme sind in zahlreichen technischen Systemen zu finden, z.B. in Automobilen, Luft- und Raumfahrt, sowie in modernen Haussteuerungen.

Insbesondere durch die Zunahme an Kommunikation nimmt die Software eine zunehmend wichtigere Rolle ein ([SW07a]). Während die Entwicklung des Regelungsverhaltens in eingebetteten Systemen gut unterstützt wird, ist die Modellierung der Kommunikation zwischen Komponenten nur mangelhaft unterstützt. Diese ist jedoch ein essentieller Bestandteil des Systemverhaltens ([NS08]) und muss deshalb ebenso unterstützt werden, um ein korrektes Gesamtverhalten zu erreichen.

Da mechatronische Systeme häufig in einem sicherheitskritischen Umfeld agieren, wie z.B. bei Steuerungssoftware in Fahrzeugen, ist eine Verifikation des Systemverhaltens notwendig. Ein Problem von aktuellen Verifikationsverfahren wie dem Model Checking ([CGP99]) ist jedoch, dass diese nicht für große Systeme skalieren. Eine Möglichkeit, die Komplexität beherrschbar zu machen, ist die komponentenbasierte Entwicklung ([SW07b]) mit hierarchischen Komponenten. Diese kapseln jeweils einen Teil des Gesamtverhaltens und können über Ports mit ihrer Umwelt kommunizieren. Dies ermöglicht die Verifikation des Gesamtverhaltens über ein kompositionales Model Checking der einzelnen Komponenten und der Kommunikation zwischen den Komponenten ([GTB⁺03]).

Die Kommunikation zwischen den Komponenten wird über Kommunikationsprotokolle definiert. Diese definieren, welche Reihenfolgen von Aufrufen oder Nachrichten bei der Kommunikation zulässig sind. Im Falle der hier betrachteten Echtzeitprotokolle sind außerdem bestimmte Zeitintervalle spezifiziert, in denen die Kommunikation ablaufen muss. Im Rahmen dieser Arbeit wurden Echtzeitprotokolle mit einer dynamischen Anzahl von

Teilnehmern betrachtet. Kommunikationsprotokolle für eine 1:n Kommunikation mit einer dynamischen Anzahl von Teilnehmern können über sogenannte (parametrisierte) Echtzeit-Koordinationsmuster ([Hir08]) spezifiziert und verifiziert werden. Die Muster werden dabei zunächst unabhängig von den konkreten Komponenten entwickelt und beinhalten Rollen, die später an den Ports der Komponenten instanziiert werden.

Die Instanziierung der Muster an einer Komponente und die Delegation eines Protokolls an eine eingebettete Komponente können eine Verfeinerung des Verhaltens nach sich ziehen. Im ersten Fall muss das Muster an die konkrete Komponente angepasst werden, im zweiten Fall ist es möglich, dass innerhalb der eingebetteten Komponente weiteres Verhalten ausgeführt werden muss. Die Überprüfung dieser Verfeinerung war für ein dynamisches Echtzeitprotokoll bislang nicht möglich, ist aber für die vollständige Verifikation des Systemmodells unerlässlich. Im Rahmen dieser Arbeit ([Hei09]) wurde ein Verfahren vorgestellt, mit dem eine korrekte Verfeinerung eines Echtzeitprotokolls mit dynamischer Anzahl von Teilnehmern verifiziert werden kann. Bislang waren Systeme entweder sicherheitskritisch, wie z.B. Steuerungssoftware in Fahrzeugen oder Atomkraftwerken, aber nicht selbstadaptiv oder eine Software war selbstadaptiv, konnte aber nicht in einem sicherheitskritischen Umfeld agieren, wie z.B. Webanwendungen. Durch die erstmals mögliche Verifikation des Gesamtsystems können nun auch selbstadaptive, sicherheitskritische Systeme entwickelt werden.

Ausgehend von einer Verfeinerungsdefinition für Webservices ([HT05]) wurde eine flexible Verfeinerungsdefinition für Echtzeitprotokolle eingeführt, die die Ausführung von zusätzlichem, komponenteninternen Verhalten erlaubt und eine eingeschränkte Relaxierung des Echtzeitverhaltens ermöglicht. Ein Problem bei der Definition der Verfeinerung sind unterschiedliche Formalismen. Das von den Ports umgesetzte Protokoll ist über Realtime Statecharts ([GB03]) definiert, die eine Erweiterung von UML Statecharts um Zeit darstellen. Die Rekonfiguration des Systems wird über Story Diagramme ([Zün01]) beschrieben, die auf Graphtransformationen basieren. Dieses Problem wurde durch die Einführung eines Timed Story Driven Modeling ([HHZ09]) Ansatzes gelöst, der auf Story Diagrammen basiert und einen gemeinsamen Formalismus darstellt, der die zustandsbasierten und zeitlichen Konzepte von Realtime Statecharts mit Story Diagrammen vereint.

Im Rahmen dieser Arbeit wurden verschiedene Ansätze zur Verifikation der Verfeinerung vorgestellt und bewertet. Einer dieser Ansätze wurde prototypisch implementiert und evaluiert. Für die Evaluierung wurde das an der Universität Paderborn entwickelte RailCab System¹ verwendet. Dieses System besteht aus autonomen RailCabs, die führerlos, nur durch Software gesteuert fahren und sich selbstständig zu kontaktlosen Konvois mit dynamischer Länge zusammenschließen können, wie in Abbildung 1 für drei RailCabs dargestellt.

Ein Ausschnitt aus einer Komponentenarchitektur für das RailCab System zeigt Abbildung 2. Die Komponente RailCab repräsentiert das gesamte RailCab und bettet Komponenten für die einzelnen Teile des Gesamtverhaltens ein. Eine dieser Komponenten ist die Komponente Coordinator, die das Verhalten des RailCabs als Anführer eines Konvois definiert. Das Muster ConvoyCoordination definiert das Protokoll zwischen dem Koordinator und

¹<http://www-nbp.uni-paderborn.de/>

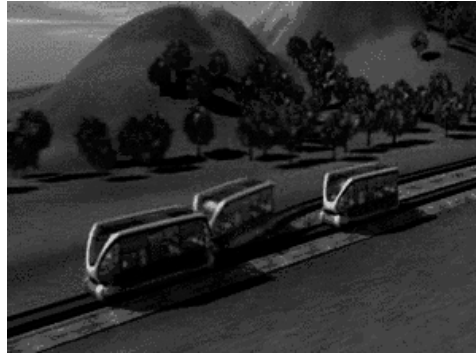


Abbildung 1: Konvoibildung von RailCabs

den übrigen Mitgliedern des Konvois, mit dem der Konvoi aufrechterhalten wird. Der Port coordinator ist ein Multiport, der mehrfach instanziiert werden kann. Für jedes Konvoimitglied muss zur Laufzeit ein Port erzeugt werden, wodurch das Protokoll dynamisch wird. Das Verhalten dieses Ports wird durch Realtime Statecharts beschrieben. Dabei existiert eine Statechartinstanz für jede Portinstanz. Die Coordinator Komponente delegiert das Verhalten ihrerseits an die Komponente PosCalc, die ebenfalls mehrfach instanziiert werden kann. Auch hier muss für jedes Konvoimitglied eine Instanz angelegt werden, die eine Positionsbestimmung durchführt.

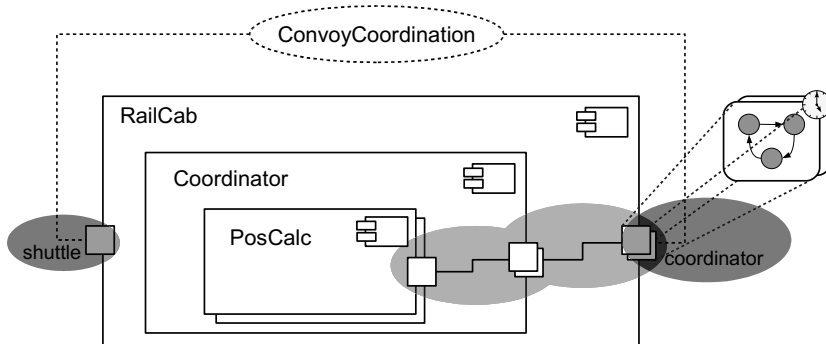


Abbildung 2: Komponentendiagramm mit einem parametrisierten Echtzeit-Koordinationsmuster zur Konvoikoordination. Mögliche Verfeinerungen sind hervorgehoben.

Die möglichen Verfeinerungen bei der Anwendung des Musters sind in dunkelgrau hervorgehoben, die Verfeinerungen bei der Delegation an eingebettete Komponenten in hellgrau. Dabei ist zu beachten, dass durch das hier entwickelte Verfahren zur Verifikation der Verfeinerung auch eine Delegation zwischen unterschiedlichen Strukturelementen möglich ist, d.h. von Multiports auf mehrfach instanziierbare Komponenten mit einfachen Ports.

Literatur

- [CGP99] Edmund M. Clarke, Orna Grumberg und Doron A. Peled. *Model Checking*. The MIT Press, Cambridge, Massachusetts, 1999.
- [GB03] Holger Giese und Sven Burmester. Real-Time Statechart Semantics. Bericht tr-ri-03-239, Lehrstuhl für Softwaretechnik, Universität Paderborn, Paderborn, Germany, June 2003.
- [GH06] Holger Giese und Stefan Henkler. A survey of approaches for the visual model-driven development of next generation software-intensive systems. *Journal of Visual Languages and Computing*, 17(6):528–550, December 2006.
- [GHH⁺08] Holger Giese, Stefan Henkler, Martin Hirsch, Vladimir Roubin und Matthias Tichy. Modeling Techniques for Software-Intensive Systems. In Dr. Pierre F. Tiako, Hrsg., *Designing Software-Intensive Systems: Methods and Principles*, Seiten 21–58. Langston University, OK, 2008.
- [GTB⁺03] Holger Giese, Matthias Tichy, Sven Burmester, Wilhelm Schäfer und Stephan Flake. Towards the Compositional Verification of Real-Time UML Designs. In *Proc. of the 9th European software engineering conference held jointly with 11th ACM SIGSOFT international symposium on Foundations of software engineering (ESEC/FSE-11)*, Seiten 38–47. ACM Press, September 2003.
- [Hei09] Christian Heinzemann. Verifikation von Protokollverfeinerungen. Master Thesis, Software Engineering Group, University of Paderborn, Nov 2009.
- [HHZ09] Christian Heinzemann, Stefan Henkler und Albert Zündorf. Specification and Refinement Checking of Dynamic Systems. In Pieter Van Gorp, Hrsg., *Proceedings of the 7th International Fijaba Days*, Seiten 6–10. Eindhoven University of Technology, The Netherlands, November 2009.
- [Hir08] Martin Hirsch. *Modell-basierte Verifikation von vernetzten mechatronischen Systemen*. Dissertation, University of Paderborn, July 2008.
- [HT05] Reiko Heckel und Sebastian Thöne. Behavioral Refinement of Graph Transformation-Based Models. *Electronic Notes in Theoretical Computer Science*, 127(3):101 – 111, 2005. Proceedings of the Workshop on Software Evolution through Transformations: Model-based vs. Implementation-level Solutions (SETra 2004).
- [KM07] Jeff Kramer und Jeff Magee. Self-Managed Systems: an Architectural Challenge. In *FOSE '07: 2007 Future of Software Engineering*, Seiten 259–268, Washington, DC, USA, 2007. IEEE Computer Society.
- [NS08] Oliver Niggemann und Joachim Stroop. Models for model’s sake: why explicit system models are also an end to themselves. In *ICSE '08: Proceedings of the 30th international conference on Software engineering*, Seiten 561–570, New York, NY, USA, 2008. ACM.
- [SW07a] Wilhelm Schäfer und Heike Wehrheim. The Challenges of Building Advanced Mechatronic Systems. In *FOSE '07: 2007 Future of Software Engineering*, Seiten 72–84, Washington, DC, USA, 2007. IEEE Computer Society.
- [SW07b] Wilhelm Schäfer und Heike Wehrheim. Eingebettete Software. <http://www.transmechatronic.de/index.php?id=119>, 2007.
- [Zün01] Albert Zündorf. *Rigorous Object Oriented Software Development*. University of Paderborn, 2001.

Ausarbeitung und Umsetzung eines Ansatzes für modellbasiertes Testen in betrieblichen Informationssystemen

Annette Heym
Technische Universität München
Institut für Informatik
annettehey@web.de

Abstract: Software wird durch die vermehrte Integration in den Alltag immer wichtiger. Um zu gewährleisten, dass Software zuverlässig und qualitativ hochwertig ist, muss sie getestet werden. Modellbasiertes Testen (MBT) ist eine Technik, die immer mehr an Bedeutung gewinnt und mit der Zeit und damit Kosten im Test eingespart werden können. Daher ist auch Capgemini sd&m daran interessiert, einen Ansatz für MBT in den vorhandenen Prozess der Testfallerstellung zu integrieren. Insbesondere geht es darum, wie man aus einem vorhandenen Modell Testfälle generieren kann. Als Lösung des Problems wurde in dieser Masterarbeit aus einem vorhandenen Systemmodell durch Transformation ein Testmodell erstellt. Nach manueller Erweiterung des Testmodells um Testdaten können daraus Testfälle generiert werden. Das erstellte Konzept wurde prototypisch umgesetzt und anhand eines Beispiels evaluiert.

1 Einleitung

Nach [SL07] hängt der Erfolg eines Softwaresystems von seiner Qualität ab. Um diese Qualität zu sichern muss die Software getestet werden. Damit Testen nicht unnötig Kosten und Aufwand verursacht ist es nötig, schon früh im Softwareentwicklungsprozess zu testen. Hier wird modellbasiertes Testen (MBT) interessant. Nach [UL07] geht es bei MBT darum, automatisch Testfälle aus formalen Modellen, die die Software beschreiben zu generieren. Dadurch kann schon vor der Implementierung getestet werden. Ein Ansatz dafür soll im Rahmen dieser Masterarbeit für Capgemini sd&m entwickelt werden. Systeme bei Capgemini sd&m werden anhand von Modellen spezifiziert, die mit der Unified Modeling Language (UML) erstellt werden. Diese Modelle bilden die Systemspezifikation, die Ausgangspunkt für diese Masterarbeit ist. Im Folgenden werden die Modelle der Systemspezifikation als Systemmodell bezeichnet. Im Vordergrund stehen UML Aktivitätsdiagramme, die bei Capgemini sd&m verwendet werden, um Anwendungsfälle zu beschreiben. Aufgabe ist nun, ein Konzept zu entwickeln, wie ausgehend vom Systemmodell teilweise automatisiert Testfälle erstellt werden können. Das soll anhand eines Prototyps umgesetzt und evaluiert werden. Im Folgenden besprechen wir zunächst die konzeptuelle Lösung der Masterarbeit. Dies beinhaltet Entscheidungen, die während der Erstellung des Konzepts getroffen wurden. Danach gehen wir auf die technische Lösung ein, die die Erstellung des Prototyps sowie die Evaluierung anhand eines Beispiels beschreibt.

2 Konzeptuelle Lösung

Zunächst müssen wir uns ein Konzept überlegen, wie wir aus dem Systemmodell Testfälle generieren können. Im Systemmodell sind nicht alle Informationen vorhanden, die wir für die Erstellung der Testfälle benötigen, wie z.B. Testdaten. Pretschner u.a. haben in [PP05] verschiedene Szenarien zur Erstellung von Testfällen beschrieben. Als erstes ist zu entscheiden, wo die fehlenden Informationen eingefügt werden sollen. Entweder das vorhandene Systemmodell wird mit für den Test notwendigen Informationen erweitert, oder es wird ein separates Modell erstellt, das sogenannte Testmodell, in dem die testrelevanten Informationen enthalten sind. In der Literatur gibt es verschiedene Ansätze für beide Szenarien. In [VLH⁺06] und [HVFR05] wird beispielsweise kein separates Testmodell erstellt, in [ZNSF] und [BDKK08] dagegen schon. Beide Varianten haben Vor- und Nachteile. Zwei separate Modelle bieten den Vorteil, dass die Belange klar getrennt sind, was die Handhabung und Verständlichkeit der Modelle erleichtert. Andererseits ist es ein erheblicher Mehraufwand, zusätzlich zum Systemmodell ein Testmodell zu erstellen. Generiert man allerdings das Testmodell aus dem Systemmodell, so fällt dieser Mehraufwand weg. Aus diesem Grunde haben wir uns für die Variante des separaten Testmodells entschieden. Als nächstes stellt sich die Frage, welche Informationen aus dem Systemmodell ins Testmodell übertragen werden müssen, und wie daraus Testfälle generiert werden können. Ein Überblick über den Prozess der Testfallgenerierung in dieser Arbeit ist in Abbildung 1 dargestellt.

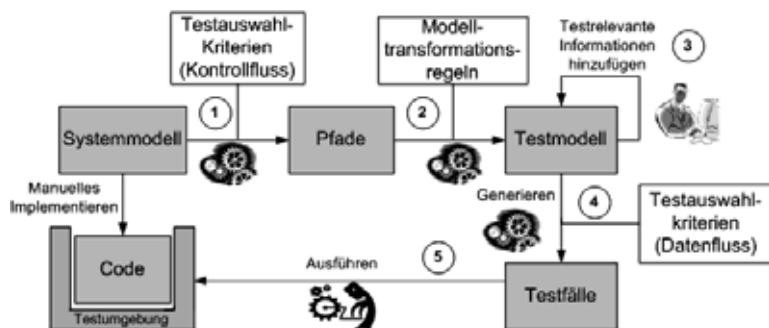


Abbildung 1: Überblick über die Testfallgenerierung

In dieser Masterarbeit werden die Testfälle aus UML Aktivitätsdiagrammen, die Anwendungsfälle beschreiben generiert. Dabei haben wir wiederum zwei verschiedene Varianten identifiziert, um die Aktivitätsdiagramme vom Systemmodell ins Testmodell zu übertragen. Eine Möglichkeit ist, das gesamte Diagramm so wie es im Systemmodell erstellt wurde ins Testmodell zu übertragen. Die zweite Möglichkeit ist, aus dem Aktivitätsdiagramm Pfade auszuwählen, und diese separat ins Testmodell zu übertragen. Dadurch sind im Testmodell schon erste Testfälle sichtbar, was einen großen Vorteil für den Testdesigner darstellt. Deshalb haben wir diese Möglichkeit gewählt. Das entspricht Schritt 1 aus Abbildung 1. Jetzt muss noch entschieden werden, wie die Pfade aus dem Aktivitätsdiagramm selektiert werden sollen. Dafür haben wir das Testauswahlkriterium “all-one-loop-paths” gewählt. Damit werden alle Pfade, die keine Schleife enthalten, so-

wie alle Pfade die einen einmaligen Schleifendurchlauf enthalten selektiert (vgl. [UL07]). Zusätzlich werden Informationen aus der Spezifikation der Dialoge sowie aus dem Datenmodell des Systemmodells benötigt. Diese liefern z.B. wichtige Informationen über die später zu erstellenden Testdaten, und müssen mit in das Testmodell einfließen.

Jetzt müssen wir uns noch eine geeignete Darstellungsform für das Testmodell überlegen. Das Systemmodell wird bei Capgemini sd&m mithilfe des Modellierungstools Enterprise Architect (EA) von Sparx Systems (siehe <http://www.sparxsystems.com>) erstellt. Daher ist es von Vorteil, wenn das zu erstellende Testmodell ebenfalls in EA darstellbar ist. Aus den in der Literatur aufgeführten Möglichkeiten zur Erstellung von Testmodellen haben wir uns für das UML Testing Profile (UTP) entschieden (<http://utp.omg.org/>). UTP ist ein UML-Profil, das es ermöglicht, alle für den Test eines Systems relevanten Informationen zu modellieren. Mit UTP kann sowohl das Verhalten (die Testfälle) als auch die Struktur (am Test beteiligte Komponenten) beschrieben werden. UTP basiert auf UML und kann problemlos in EA integriert werden.

Damit ist nun auch die Struktur des Testmodells klar definiert. Im nächsten Schritt muss überlegt werden, wie man aus dem Systemmodell das Testmodell erstellen kann. Das erfordert eine Modell-zu-Modell-Transformation mithilfe von Transformationsregeln (siehe Schritt 2 in Abbildung 1). Hier ist zu entscheiden, ob eine Modelltransformationssprache wie z.B. Query/View/Transformation (QVT) (vgl. www.omg.org/spec/QVT/) eingesetzt werden soll, oder ob eine eigene Implementierung sinnvoller ist. Um eine Modelltransformationssprache wie QVT zu verwenden muss das Modell aus EA exportiert und mit QVT weiter verarbeitet werden. Dies ist z.B. über das Eclipse Modeling Framework (EMF) möglich (<http://www.eclipse.org/modeling/emf/?project=emf>). Aufgrund technischer Probleme haben wir uns jedoch für eine eigene Implementierung der Transformation entschieden.

Sobald das Testmodell erstellt ist kann der Testdesigner mit seiner Arbeit beginnen. Seine Aufgabe ist es nun, das Testmodell mit zusätzlichen Informationen zu erweitern (siehe Schritt 3 in Abbildung 1). Das beinhaltet u.a. das Erstellen von konkreten Testdaten für die Testfälle.

Wenn der Testdesigner das Testmodell vervollständigt hat, können wie in Schritt 4 in Abbildung 1 dargestellt Testfälle aus dem Testmodell generiert werden. Hier können nun auch auf den Datenfluss bezogene Testauswahlkriterien angewendet werden. Diese Testfälle werden dann anschließend entweder manuell oder automatisiert auf dem manuell implementierten System ausgeführt (siehe Schritt 5 in Abbildung 1). Diese letzten beiden Schritte sind allerdings nicht mehr Teil dieser Masterarbeit.

3 Technische Lösung und Evaluation

Wie bereits in Abschnitt 2 erwähnt haben wir uns für eine eigene Implementierung der Transformation in Java entschieden. Enterprise Architect bietet für Java die Schnittstelle “Automation Interface” (siehe <http://www.sparxsystems.com>) an, über die man auf von EA bereitgestellte Methoden zum Lesen und Erstellen von Modellen zugreifen kann. Nach dem Einlesen des Systemmodells wird eine Pfadanalyse auf dem Aktivitätsdiagramm aus-

geführt. Dafür wurde die Selektion der Pfade in einem Algorithmus in Java implementiert. Danach werden die ebenfalls in Java implementierten Transformationsregeln auf jeden selektierten Pfad angewendet und dazu je ein Aktivitätsdiagramm im Testmodell erstellt. Zusätzlich zu diesen Aktivitätsdiagrammen gibt es noch weitere Klassendiagramme, die die am Test beteiligten Komponenten und deren Kommunikationswege darstellen.

Anhand eines fiktiven Projektbeispiels von Capgemini sd&m wird der erstellte Prototyp schließlich evaluiert. Exemplarisch wurde ein Anwendungsfall aus dem Systemmodell des Beispiels ausgewählt, der als Eingabe für den Prototypen verwendet wird.

4 Zusammenfassung

Ziel dieser Arbeit war es, ein Konzept zu entwickeln, wie modellbasiertes Testen in den Testprozess von Capgemini sd&m integriert werden kann. Um ein passendes Konzept zu erstellen, wurden wichtige Entscheidungen getroffen. Als Testmodell wurde ein separates Testmodell gewählt, das bereits einzelne Pfade des ursprünglichen Aktivitätsdiagramms enthält. Diese Pfade wurden mithilfe des Testauswahlkriteriums “all-one-loop-paths” selektiert. Die Darstellung des Testmodells erfolgte mithilfe von UTP. Um das Konzept umzusetzen sollte ein Prototyp entwickelt werden. Dieser wurde in Java implementiert und anhand eines Beispiels evaluiert.

Als Weiterführung dieser Arbeit können neben dem “all-one-loop-paths”-Kriterium noch weitere Testauswahlkriterien implementiert werden. Des Weiteren ist es wichtig, eine Lösung zu finden, wie Änderungen im Systemmodell komfortabel in das Testmodell übertragen werden können, ohne dass dabei bereits vorgenommene Ergänzungen am Testmodell erneut durchgeführt werden müssen.

Literatur

- [BDKK08] Dr. C. Brandes, A. Ditze, C. Kollee und Dr. D. Kreische. Modellbasiertes Testen praxisgerecht realisiert: vom UML2-Profil zum Testmanagement-Werkzeug. März 2008.
- [HVFR05] J. Hartmann, M. Vieira, H. Foster und A. Ruder. A UML-based approach to system testing. März 2005.
- [PP05] A. Pretschner und J. Philipps. *Methodological Issues in Model-Based Testing*, Kapitel 10, Seiten 281 – 291. Springer Verlag, 2005.
- [SL07] A. Spillner und T. Linz. *Basiswissen Softwaretest*. dpunkt.verlag, 2007.
- [UL07] M. Utting und B. Legeard. *Practical Model-Based Testing - A Tools Approach*. Morgan Kaufmann Publishers, 2007.
- [VLH⁺06] M. Vieira, J. Leduc, B. Hasling, R. Subramanyan und J. Kazmeier. Automation of GUI Testing Using a Model-driven Approach. Mai 2006.
- [ZNSF] J. Zander-Nowicka, I. Schieferdecker und T. Farkas. Derivation of Executable Test Models from Embedded System Models Using Model Driven Architecture Artefacts.

Peer-to-Peer Kommunikation zwischen Multimediaanwendungen in einem Placeshiftszenario

Ralf Kruthoff-Brüwer, Julius Hoyer

Fachhochschule Osnabrück
Fakultät Ingenieurwissenschaften und Informatik

R.Kruthoff-Bruewer@fh-osnabrueck.de
J.Hoyer@fh-osnabrueck.de

Art der Arbeit: Diplomarbeit

Betreuer/in der Arbeit: Prof. Dr. Karsten Morisse

Abstract: Im Rahmen der Diplomarbeit mit dem Titel „Peer-to-Peer Kommunikation zwischen Multimediaanwendungen in einem Placeshiftszenario“ wurden verschiedene Problematiken die bei der direkten Kommunikation zwischen Endgeräten in einem IP-Netzwerk wie dem Internet auftreten können betrachtet. Speziell Router, die mit Network Address Translation [EF94] (NAT) arbeiten und häufig mit Firewalls ausgestattet sind, machen es schwierig eine direkte Verbindung zwischen zwei Clients aufzubauen. Ziel der Arbeit war es, diese Probleme genauer zu ergründen und Techniken zu finden, die es erlauben eine Peer-to-Peer-Verbindung auch dann zu etablieren, wenn sich ein oder sogar beide Endgeräte hinter einem Router befinden. Als spezieller Anwendungsfall diene ein sog. Placeshiftszenario, in dem Audio/Video-Daten zwischen Endgeräten gestreamt werden sollen.

1 Problembeschreibung

Sollen Daten in einem IP-Netzwerk direkt, also Peer-to-Peer übertragen werden, kommt es häufig zu Problemen beim Verbindungsaufbau. Grund hierfür ist die Verwendung von NAT-Techniken zur Vermittlung zwischen privaten und öffentlichen IP-Netzen. Diese Technik sieht herkömmliche Client/Server-Kommunikation vor, bei der der Server öffentlich erreichbar ist. Allerdings bringt eine Client/Server-Struktur gerade in den Bereichen in denen es darum geht breitrandige Audio/Video-Daten zu übertragen (Placeshifting) große Nachteile im Bezug auf die benötigte Server-Infrastruktur mit sich.

Die durch ein NAT maskierte oder eine Firewall geschützten Clients können zwar eine Verbindung in ein öffentliches Netz aufbauen. Wird jedoch versucht einen Client in einem anderen privaten Netz zu erreichen, schlägt dieser Versuch oft fehl. Wie in Abbildung 1 gezeigt ist es also beiden Endgeräten nicht möglich den jeweils anderen Kommunikationspartner zu erreichen.

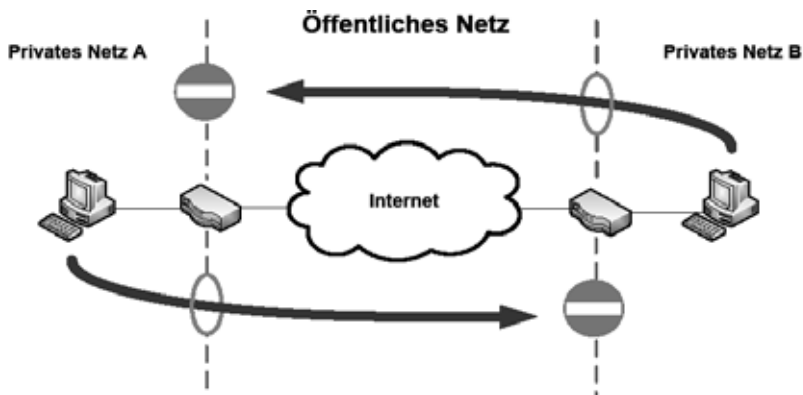


Abbildung 1: Problemstellung

2 Network Address Translation

Durch Network Address Translation wird es möglich, mehrere Endgeräte, die in einem privaten Netzwerk miteinander verbunden sind, über nur eine öffentliche IP-Adresse mit dem Internet zu verbinden. Abbildung 2 zeigt die grundlegende Funktionalität von NAT: Geht eine Anfrage aus dem privaten Netz am NAT-Gerät ein, wird die Quell-Adresse im Paket gegen die Adresse am öffentlichen Interface des NAT ersetzt. In einer Tabelle wird gespeichert, von wem die Anfrage gestellt wurde (private Quell-Adresse), auf welche Adresse diese geändert wurde (globale Quell-Adresse) und wie die Ziel-Adresse lautet, an die das Datenpaket übertragen werden soll. Anschließend wird das Paket mit geänderten Adressdaten an das Ziel versendet. Antwortet der Zielrechner auf die Anfrage, kann das NAT-Gerät anhand der Tabelle bestimmen, von welchem Gerät auf der internen Seite die Anfrage gestellt wurde und diese zuordnen. Wurde kein Eintrag durch eine Anfrage nach außen in die Tabelle geschrieben, werden alle Pakete, die an der Außenseite des NAT eingehen, verworfen. Es können also keine Verbindungen von außen initialisiert werden.

Das grundlegende NAT-Verfahren wird in der RFC 1631 [EF94] aus dem Jahr 1994 beschrieben. In der RFC 3022 [SE01] wurde dieses Dokument dann um weitere NAT Typen erweitert und NAT genauer spezifiziert.

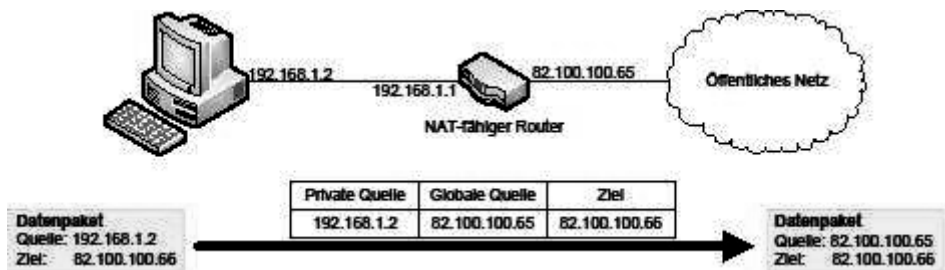


Abbildung 2: Funktionsweise der Network Address Translation

Trotz dieser Spezifikationen werden die verschiedenen NAT-Typen von vielen Herstellern NAT-fähiger Geräte zu großen Teilen unterschiedlich interpretiert und implementiert.

3 NAT-Traversal Techniken

Vorgehensweisen und Techniken, die das Etablieren einer Verbindung von außen durch ein NAT-Gerät hindurch ermöglichen werden als NAT-Traversal Techniken bezeichnet. In diesem Bereich gibt es bereits eine Reihe von Techniken, die auf unterschiedliche Art und Weise versuchen die erläuterten Probleme zu umgehen. Im Rahmen der Arbeit wurden folgende Techniken näher betrachtet:

- Relaying
- Connection Reversal
- NAT Port Mapping Protocol
- UPnP
- Application Layer Gateway
- Session Border Controller
- Simple traversal of UDP over NATs [RWHM03]
- Traversal Using Relay NAT [RMM08]
- Interactive Connection Establishment [Ros07]
- Hole Punching

Einige dieser Techniken wie UPnP setzen jedoch spezielle Eigenschaften des NAT-Geräts voraus. Andere wiederum arbeiten nur mit speziellen Implementierungen des NAT-Verfahrens einwandfrei. Daher war es das Ziel der Arbeit, die bekannten Techniken zu analysieren, zu erweitern und zu kombinieren. Als Resultat sollte eine möglichst hohe Erfolgsquote beim Initialisieren einer Verbindung über NAT Geräte hinweg erreicht werden.

4 Kombination verschiedener NAT-Traversal Techniken

Um das erfolgreiche Herstellen einer Verbindung in möglichst vielen Fällen zu ermöglichen, wurden einige der analysierten Techniken kombiniert, erweitert und implementiert. So wurde z.B. das STUN-Protokoll um Sicherheitsmechanismen erweitert. Außerdem wurden die Techniken des Hole Punching und des Connection Reversal kombiniert.

5 Implementierung

Das erarbeitete kombinierte Verfahren wurde in einem NAT-Traversal-Framework implementiert. Auf Basis dieses Frameworks ist ein Server entstanden, der im ersten Schritt mit Hilfe des STUN-Verfahrens Informationen über die einzelnen Clients sammelt. Soll nun eine Peer-to-Peer-Verbindung zwischen zwei Clients etabliert werden, gibt der Server die Verbindungsmethoden und Parameter vor.

Literaturverzeichnis

- [EF94] Egevang, K.; Francis, P.: The IP Network Address Translator (NAT). RFC 1631 (Informational). <http://www.ietf.org/rfc/rfc1631.txt>. Version: Mai 1994 (Request for Comments).
- [SE01] Srisuresh, P.; Egevang, K.: Traditional IP Network Address Translator (Traditional NAT). RFC 3022 (Informational). <http://www.ietf.org/rfc/rfc3022.txt>. Version: Januar 2001 (Request for Comments).
- [RMM08] Rosenberg, J. ; Mahy, R. ; Matthews, P.: Traversal Using Relays around NAT (TURN). draft-ietf-behave-turn-11. <http://tools.ietf.org/id/draft-ietf-behave-turn-11.txt>. Version: Oktober 2008 (Internet-Draft).
- [Ros07] Rosenberg, J.: Interactive Connectivity Establishment (ICE). draft-ietfmmusic-ice-19. <http://tools.ietf.org/id/draft-ietf-mmusic-ice-19.txt>. Version: Oktober 2007 (Internet-Draft).
- [RWHM03] Rosenberg, J. ; Weinberger, J. ; Huitema, C. ; Mahy, R.: STUN – Simple Traversal of User Datagram Protocol (UDP) Through Network Address Translators (NATs). RFC 3489 (Proposed Standard). <http://www.ietf.org/rfc/rfc3489.txt>. Version: März 2003 (Request for Comments).

IQ Profil: Eine MediaWiki Erweiterung zur Bewertung von Informationsqualität der Artikel

Andreas Kohring

Universität Paderborn
Institut für Informatik
ako@upb.de

Abstract: Um Benutzern von Unternehmenswikis die Möglichkeit zu bieten, auf einen Blick die Informationsqualität eines Wikiartikels zu erkennen, wurde eine Erweiterung für ein MediaWiki erstellt. Diese Erweiterung ermöglicht es den Benutzern zu erkennen, wie gut oder schlecht die Informationsqualität des Artikels bewertet wurde. Die Erweiterung stellt neben neuen Funktionen in der MediaWiki API auch eine JavaFX Implementierung einer GUI zur Verfügung, die über die erweiterte MediaWiki API die Ergebnisse darstellen kann. Es wird die Möglichkeit geboten, dass jeder Benutzer eine individuelle Bewertung abgeben kann. Durch die API Erweiterung wird ebenfalls die Möglichkeit bereit gestellt, von beliebiger Stelle, automatisiert auf die Bewertungen zuzugreifen.

Art der Arbeit: Diplomarbeit

Betreuer: Wolfgang Reinhardt (wolle@upb.de), Johannes Magenheim (jsm@upb.de)

1 Problemstellung

Unternehmen nutzen vermehrt die Möglichkeit von Wikis, um Informationen ihres Unternehmens und ihrer Mitarbeiter zu organisieren. Aktuell setzen über 30% der Unternehmen weltweit Wikis ein [MFG08]. Ziel ist es oft, dass so auch fachfremde Mitarbeiter auf Informationen Zugriff erhalten, die aus anderen Bereichen des Unternehmens stammen. Solche Benutzer haben jedoch nur selten die Möglichkeit zu erkennen, ob die Informationen qualitativ hoch- oder minderwertig sind. Zur Bewertung der Informationsqualität von Wikiartikeln wurde in [Fri09] ein Satz von Kriterien und Indikatoren erarbeitet, anhand derer man die Qualität der Informationen erfassen kann. Darauf aufbauend hat diese Arbeit das Ziel, eine Erweiterung eines Wikis zu erstellen, die auf Basis dieser Kriterien implementiert wird und die Informationsqualität sowohl erfassbar, als auch darstellbar macht. Die entwickelte Erweiterung wird durch eine Benutzerstudie in einem Feldversuch evaluiert werden. Die Evaluation findet in einem typischen Unternehmenswiki statt. Zum aktuellen Zeitpunkt ist diese Evaluationsphase jedoch noch nicht abgeschlossen, und kann daher hier noch nicht weiter dargestellt werden.

2 Das Wiki-Set

Die Arbeit [Fri09] untersucht die Möglichkeit, Informationsqualität von Artikeln in Wikis zu bewerten, die in Unternehmen eingesetzt werden. Sie stellt verschiedene Modelle zur Bewertung von Informationsqualität in unterschiedlichen Medien gegenüber, und bestimmt daraus einen Satz von Kriterien, anhand derer die Qualität der Informationen bewertet werden kann. Es wurden fünf Kriterien identifiziert, die durch verschiedene Indikatoren spezifiziert werden. Der so gefundene Satz von Kriterien und Indikatoren wird als Wiki-Set bezeichnet und in Tabelle 1 aufgelistet.

Kriterium	Indikatoren
Glaubwürdigkeit	Verbindung und Bekanntheit des Autors, Merkmale des Autors, Quellenangaben, Anzahl bisheriger Zugriffe
Relevanz	Artikeltitel, Zusammenfassung, Anzahl bisheriger Zugriffe, Artikellänge
Aktualität	Datum der letzten Änderung, Erstellungsdatum, Korrektheit und Aktualität von Links
Vollständigkeit	Quellenangaben, Artikellänge
Verständlichkeit	Rechtschreibung, Grammatik, Interpunktion, Struktur, Verwendung von Multimedia-Inhalten, weiterführende Links, Hinweise auf erforderliches Vorwissen

Tabelle 1: das Wiki-Set für Informationsqualität

3 Implementierung

Es sind sehr viele, meist quelloffene Wiki-System, sogenannte Wiki-Engines, in Unternehmen verbreitet. Eines der am weitest verbreiteten Wikis ist das MediaWiki [Fou09a]. Daher wurde die Erweiterung für Wikis von diesem Typ entwickelt. Bei der Entwicklung handelt es sich um eine zweigeteilte Entwicklung. Ein Teil erweitert die MediaWiki API [Fou09b]. Sie erhält neue Funktionen um, automatisiert und maschinenlesbar, Zugriff auf die Informationen des Wikis zu erlangen. Anstelle von üblichen HTML Seiten die im Browser dargestellt werden können, werden der anfragenden Instanz via HTTP Protokoll, XML Daten gesendet. Mit der implementierten Erweiterung ist es daher möglich, Anfragen zur Informationsqualität eines Artikels an das Wiki zu stellen ohne eine Webseite nutzen zu müssen.

Ein zweiter Teil wurde zum Betrachten und Abgeben einer Bewertung erstellt, die dem Benutzer eine GUI bietet, welche in jeden beliebigen Wikiartikel eingebunden werden

kann. Diese GUI wurde in JavaFX ¹ entwickelt und nutzt zum Abfragen einer Bewertung eines Artikels, wie auch zum Abgeben einer Bewertung, die neu erstellten Funktionen der MediaWiki API. Da jede JavaFX Anwendung sich in einer Webseite wie ein Java-Applet verhält, zeigt diese Art der Einbindung gleichzeitig, dass jedes beliebige Programm auf die MediaWiki API zugreifen kann. Besonders wichtig ist das für einen möglichen späteren mobilen Einsatz, denn dafür ist JavaFX ebenfalls gerüstet.

Abbildung 1 stellt den Aufbau der gesamten Implementierung dar. Speziell sollte das Augenmerk darauf gelegt werden, dass die Benutzeroberfläche zur Bewertung sowohl im Wiki, also in einen Artikel eingebettet, betrieben werden kann, als auch theoretisch extern. Über die MediaWiki API können Bewertungen jeweils gesetzt oder abgefragt werden.

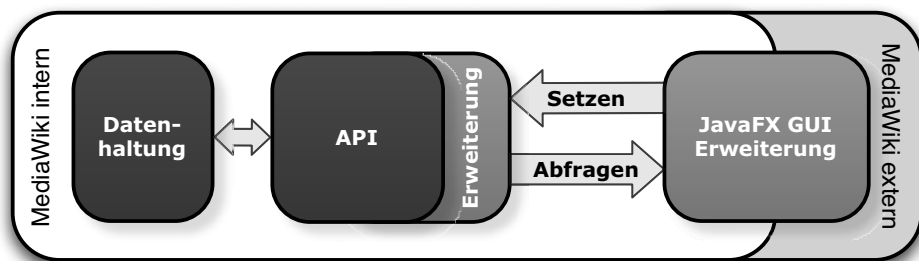


Abbildung 1: Aufbau der Erweiterung

4 Die Benutzeroberfläche

Die Oberfläche, die dem Benutzer präsentiert wird, teilt sich in verschiedene Ansichten. Die Hauptansicht präsentiert die abgegebenen Bewertungen zu den einzelnen Kriterien des Wiki-Set in Form einer Sternskala. Sie präsentiert sich dem Benutzer als erstes. Hinzu kommt, dass jedes Kriterium neben der absoluten, aktuellen Bewertung ebenfalls einen Trend anzeigt. Dieser Trend zeigt an, ob sich die Qualität verbessert oder verschlechtert. Auf dieser Skala kann jedes Kriterium allgemein bewertet werden. Möchte ein Benutzer die Bewertung eines Kriteriums genauer angeben, kann zu jedem Kriterium eine Ansicht aufgerufen werden, auf der die einzelnen Indikatoren bewertet und abgelesen werden können. Eine weitere Ansicht zeigt zu jedem Kriterium, und zur Gesamtbewertung, den zeitlichen Verlauf. Der zeitliche Verlauf wird dabei über die fortlaufenden Revisionen eines Artikels, als Liniendiagramm, abgebildet. Eine letzte Ansicht bietet dem Benutzer eine Übersicht, welche ihm die Artikel auflistet, die sich die Leser des Artikel vor und nach dem betreffenden Artikel angesehen haben. Diese Art der *social navigation* soll dem Leser helfen, weitere Informationen durch ähnliche Artikel zu bekommen, die ggf. artverwandt sein könnten.

Die Ansichten der Oberfläche, wie oben beschrieben, sind in der Abbildung 2 dargestellt.

¹<http://www.javaafx.com>

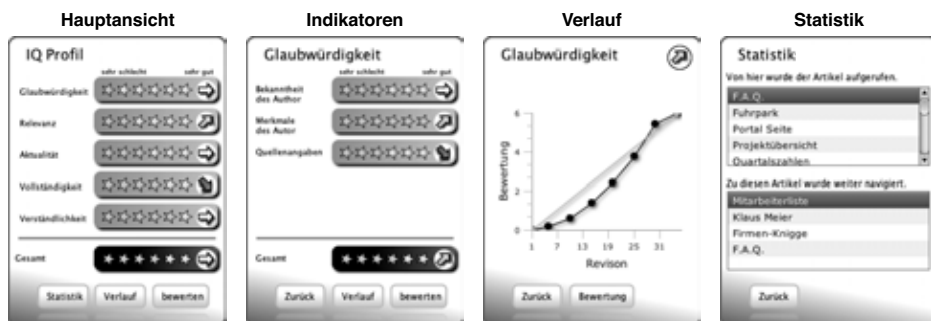


Abbildung 2: Aufbau der Erweiterung

5 Zusammenfassung

In dieser Arbeit entstand eine Implementierung des in [Fri09] erläuterten Wiki-Set. Es definiert Kriterien und Indikatoren zur Bewertung der Informationsqualität von Artikeln eines Unternehmenswiki. Es entstand eine Erweiterung für ein Wiki vom Typ MediaWiki. Die hinzugefügten Funktionen ergänzen die MediaWiki API um neue Funktionen, und stellen eine Benutzeroberfläche zur Verfügung, die in ausgewählte Wikiartikel eingebettet werden kann. Sie wurde in JavaFX entwickelt und nutzt die erweiterten API-Funktionen. Der gesamten Erweiterung steht als nächstes eine Evaluation im Unternehmenseinsatz bevor.

Literatur

- [Fou09a] Wikimedia Foundation. <http://www.mediawiki.org/wiki>, 2009.
- [Fou09b] Wikimedia Foundation. <http://www.mediawiki.org/wiki/API>, 2009.
- [FR09] Therese Friberg und Wolfgang Reinhardt. An Empirical Study on Criteria for Assessing Information Quality in Corporate Wikis. In Paul Bowen, Ahmed K. Elmagarmid, Hubert Österle und Kai-Uwe Sattler, Hrsg., *Proceedings of the 14th International Conference on Information Quality (ICIQ'09)*, Seiten 217–230, Potsdam, Germany, November 2009.
- [Fri09] Therese Friberg. Bewertung der Informationsqualität in Unternehmenswikis - empirische Studie zu Kriterien der Informationsqualität. Diplomarbeit Universität Paderborn, 2009.
- [MFG08] A Digital Lifestyle. Leben und Arbeiten mit Social Software. Innovationsprogramm Web 2.0 der MFG Baden-Württemberg, 2008.

Integration von MockUp-Konzepten in die Spezifikation grafischer Bedienoberflächen

Nuriye Ljaci

Universität Augsburg
Institut für Informatik
nuriye.ljaci@student.uni-augsburg.de

Abstract: Ob eine Software von seinen Benutzern verwendet wird, entscheidet sich oft am Aussehen und Workflow seiner grafischen Oberfläche (GUI). Deshalb müssen GUIs frühzeitig an die Wünsche des Kunden angepasst werden. Durch die Entwicklung eines GUI-MockUps (ein Wegwerf-Prototyp der GUI, ohne Implementierung von Funktionalitäten) kann der Kunde den Workflow der Anwendung besser nachvollziehen und Feedback zur entworfenen GUI geben. Aus diesem Grund werden in dieser Masterarbeit, an der Universität Augsburg in Zusammenarbeit mit Capgemini sd&m Research, MockUp-Konzepte in den Spezifikationsprozess von grafischen Oberflächen integriert.

1 Einleitung

Die Qualität von Software wird schnell mit dem Verständnis, dem Aussehen und dem Workflow ihrer grafischen Bedienoberfläche (GUI) gleichgesetzt. Der Benutzer sieht meist nur die GUI einer Software und schätzt deshalb die Qualität der Software umso besser ein, je benutzerfreundlicher die GUI ist. Die Gebrauchstauglichkeit (Usability) einer Software entwickelt sich zunehmend zu einer Qualitätseigenschaft der Softwareentwicklung (vgl. [Neb09]). Deshalb ist es wichtig, dass die Dialoge und der Arbeits-Workflow einer Anwendung schon zu Beginn eines Softwareprojektes nach den Wünschen des Kunden gestaltet werden. In der Spezifikationsphase wird das Layout der GUI meist auf Papier oder mittels eines Programmes gezeichnet und der Workflow nur textuell beschrieben. Doch wäre es sicherlich einfacher, schon beim Entwurf der GUI zu sehen, wie die Software später aussehen wird und diese gleich live ausprobieren kann. Der fachliche Designer könnte sich so das Feedback des Kunden einholen und schon zu Beginn eines Softwareprojektes gewünschte Änderungen an der GUI umsetzen. Durch die Erstellung von GUI-MockUps kann dieses Problem gelöst werden. GUI-MockUps werden oft als Wegwerfprototypen bezeichnet, welche jedoch keine Funktionalität der GUI enthalten sondern nur ihr Layout und ihr Verhalten simulieren. Die Capgemini sd&m AG benutzt zur Erzeugung der Spezifikation das Modellierungswerkzeug Enterprise Architect von Sparx Systems ([EA]). Deshalb wird in dieser Arbeit ein MockUp-Tool in dieses Modellierungswerkzeug integriert.

2 Auswahl eines MockUp-Tools

Um ein MockUp-Tool in den Spezifikationsprozess integrieren zu können, ist es wichtig, aus der Vielzahl von vorhandenen Tools, das richtige Tool auszuwählen. Zur Auswahl des MockUp-Tools wurde eine modifizierte Version der gewichteten Entscheidungsmatrix (vgl. [Sen]) verwendet. Nach dem Sammeln und Ausformulieren von bestimmten Kriterien, die das MockUp-Tool erfüllen muss, wurden diese in einer Tabelle den MockUp-Tools gegenübergestellt. In der Tabelle wurden für jedes MockUp-Tool Punkte zu den Kriterien vergeben. Aus diesen Punkten wurde die Gesamtpunktzahl der MockUp-Tools berechnet. Dabei gewinnt das MockUp-Tool mit der höchsten Gesamtpunktzahl.

Zur Definition der Kriterien wurden Interviews mit Software-Ingenieuren und Usability Experten durchgeführt. In dieser Arbeit wurden elf verschiedene MockUp-Tools bewertet. Bei den Interviews hat sich herausgestellt, dass die meisten fachlichen Designer, für das Zeichnen des GUI-Layouts, Microsoft Office Produkte (Excel, Powerpoint, Access und Visio) verwenden. Deshalb wurden diese Tools ebenfalls bewertet. Außerdem wurde noch das „Tool“ Papier bewertet, da viele fachliche Designer anfangs die GUI auf Papier entwerfen. Die MockUp-Tools wurden in die Entscheidungsmatrix eingetragen und anhand der Kriterien bewertet. Tabelle 1 zeigt die erzielten Endergebnisse der Tools.

MockUp-Tool	Kriterien alle erfüllt	Punkte
WireframeSketcher [Wir]	FALSCH	2,58
Screen Architect [Scr]	FALSCH	3,00
Pencil [Pen]	FALSCH	1,92
Balsamiq Mockups [Bal]	WAHR	3,44
Denim [Den]	FALSCH	2,79
Microsoft Expression Blend 3 + SketchFlow [Ble]	WAHR	4,27
Axure RP Pro [Axu]	WAHR	3,46
Lucid Spec [Luc]	FALSCH	3,00
iRise [iRi]	FALSCH	3,49
Adobe Flash Catalyst [Cat]	FALSCH	2,74
Microsoft Excel [MSO]	FALSCH	1,67
Microsoft Access [MSO]	FALSCH	2,41
Microsoft Visio [MSO]	FALSCH	2,29
Microsoft Powerpoint [MSO]	FALSCH	2,14
JFormDesigner [JFo]	FALSCH	3,14
Papier	FALSCH	1,60

Tabelle 1: Endergebnisse der Entscheidungsmatrix

Wie auf Tabelle 1 zu sehen, haben nur die MockUp-Tools Axure RP Pro, Microsoft Expression Blend3 + SketchFlow und Balsamiq Mockups die in dieser Arbeit geforderten Kriterien erfüllt. Um zu entscheiden, welches Tool in den Spezifikationsprozess integriert werden soll, wurde eine zusätzliche Evaluierung dieser Tools durchgeführt, bei der technische Aspekte getestet wurden. Balsamiq Mockups hat hier am besten abgeschnitten.

3 Architektur zur Integration des MockUp-Tools in das Modellierungswerkzeug

Die entworfene Architektur zur Integration des MockUp-Tools in das Modellierungswerkzeug ist in die drei große Komponenten *MockUp*, *Synchronisation* und *Modellierung* aufgeteilt. Die Komponente *MockUp* ist für den Zugriff auf die Daten des MockUp-Tools zuständig. Dabei kann sie sowohl lesend als auch schreibend auf die GUI-Daten zugreifen. Falls das MockUp-Tool keine Möglichkeit für einen schreibenden Zugriff bietet, greift sie nur lesend auf die GUI-Daten zu. Die Komponente *Modellierung* greift auf die Daten des Modellierungswerkzeugs zu. Damit die GUI-Daten aus dem MockUp-Tool importiert werden können, muss diese Komponente sowohl lesend als auch schreibend auf die Daten zugreifen. Mit der Komponente *Synchronisation* werden die Daten aus dem MockUp-Tool und dem Modellierungswerkzeug synchronisiert. Abbildung 1 zeigt die entworfene Architektur.

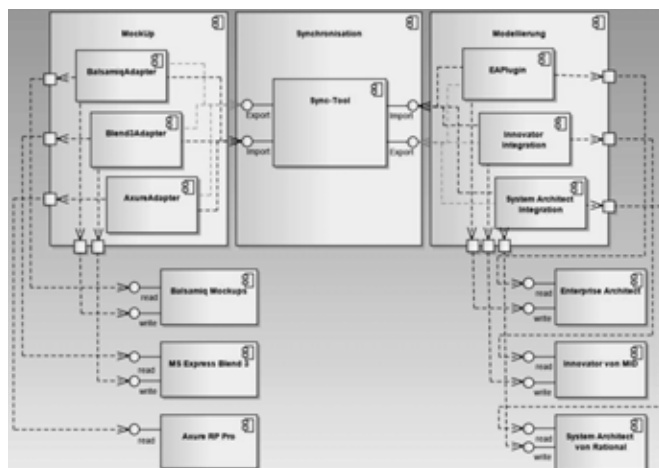


Abbildung 1: Entworfene Architektur

Jede dieser drei Komponenten ist unabhängig von den anderen Komponenten und kann auch als selbständig laufendes Programm umgesetzt werden. Die Modellierungskomponente ruft dabei die Synchronisationskomponente auf, während die Synchronisationskomponente die MockUp-Komponente aufruft. Die entworfene Architektur erlaubt, dass die Komponenten in verschiedenen Programmiersprachen und mit unterschiedlichen Technologien entwickelt werden können. Dies ist wichtig, da die von Tool-Hersteller angebotenen APIs auf unterschiedlichen Technologien basieren können. Durch diese Architektur ist es außerdem noch möglich, dass das MockUp-Tool ausgetauscht oder das System durch die Unterstützung eines weiteren MockUp-Tools erweitert werden kann. Genauso ist es möglich, das Modellierungswerkzeug auszutauschen. Als Modellierungswerkzeug sollte der Enterprise Architect benutzt werden und Balsamiq MockUps wurde als MockUp-Tool gewählt. In dieser Arbeit wurde ein Adapter für das MockUp-Tool Balsamiq Mockups, die Synchronisationskomponente und ein EA Plugin implementiert.

4 Zusammenfassung und Ausblick

Ziel dieser Arbeit war es, MockUp-Konzepte in den Spezifikationsprozess von grafischen Bedienoberflächen bei der Capgemini sd&m AG zu integrieren. Dazu wurden mit Hilfe einer eigenen Version der Entscheidungsmatrix verschiedene MockUp-Tools anhand von Kriterien bewertet. Diese Kriterien wurden aus Ergebnissen von Interviews mit Software-Ingenieuren und Usability Experten definiert und es konnte so ein geeignetes MockUp-Tool ausgewählt werden.

In dieser Arbeit wurde eine Architektur zur Integration eines MockUp-Tools in ein Modellierungswerkzeug entwickelt und diese für die beiden Tools Balsamiq Mockups und Enterprise Architect implementiert. Dabei werden mit dem entwickelten System die Daten der entworfenen GUI von Balsamiq Mockups in den Enterprise Architect importiert. Als Weiterführung dieser Arbeit könnte ein modellbasierter Ansatz zur automatischen Generierung der grafischen Oberfläche entwickelt werden, welcher die GUI-Daten aus dem Enterprise Architect liest.

Literatur

- [Axu] *Axure RP Pro*. Axure Software Solutions, Inc. <http://www.axure.com>.
- [Bal] *Balsamiq Mockups*. Balsamiq Studios, LLC. <http://www.balsamiq.com/products/mockups>.
- [Ble] *Microsoft Expression Blend3 + SketchFlow*. Microsoft. http://www.microsoft.com/expression/products/Blend_Overview.aspx.
- [Cat] *Adobe Flash Catalyst*. Adobe Systems Incorporated. <http://labs.adobe.com/technologies/flashcatalyst/>.
- [Den] *Denim*. University of Washington. <http://dub.washington.edu/denim/>.
- [EA] *Enterprise Architect*. Sparx Systems Ltd und SparxSystems Software GmbH. <http://www.sparxsystems.de/>.
- [iRi] *iRise*. iRise. <http://www.irise.com>.
- [JFo] *JFormDesigner*. FormDev Software GmbH. <http://www.formdev.com/>.
- [Luc] *Lucid Spec*. Elegance Technologies, Inc. <http://www.elegancetech.com>.
- [MSO] *Microsoft Office*. Microsoft. <http://office.microsoft.com/>.
- [Neb09] Karsten Nebe. *Integration von Software Engineering und Usability Engineering*. Shaker Verlag, 2009.
- [Pen] *Pencil*. Evolus. <http://www.evolus.vn/pencil/>.
- [Scr] *Screen Architect*. Catch Limited. <http://screenarchitect.com>.
- [Sen] Ralf Senftleben. *Methoden zur Entscheidungsfindung*. Online verfügbar auf <http://www.zeitzuleben.de/artikel/beruf/entscheidungsfindung-5.html>; zuletzt besucht am 15. Januar 2010.
- [Wir] *WireframeSketcher*. WireframeSketcher.com. <http://wireframesketcher.com>.

Testfallgenerierung aus UML-Modellen beschrieben in UML-Testing-Profile

Benjamin Niebuhr

Fachhochschule Brandenburg
Informatik und Medien
niebuhr@fh-brandenburg.de

Art der Arbeit: Bachelor-Arbeit

Betreuer/in der Arbeit:

Prof. Dr. Gabriele Schmidt, Fachhochschule Brandenburg

Dipl.-Math. Melanie Späth, Capgemini sd&m AG

Michael Mlynarski, M.Sc., Universität Paderborn

Abstract: Die Bachelor-Arbeit zeigt die technische Möglichkeit Testfälle aus einem mit dem UML-Testing-Profile beschriebenen Testmodell zu generieren und diese dann in verschiedene Formate zu exportieren. Der besondere Mehrwert liegt dabei auf der Anpassbarkeit der Schnittstellen, welcher mit den Technologien Java und Groovy entwickelt wurde.

1 Einleitung

Softwarefehler, die zu tragischen Unfällen führten, gab es viele. Ein tragischer Fehler ereignete sich in den 80er Jahren. Patienten wurden innerhalb einer Strahlentherapie der mehr als 75-fachen Strahlendosis ausgesetzt [Pf03]. Dieser und andere Fehler führten zu einer kontinuierlichen Verbesserung des Testprozesses.

Die bisherigen Testvorgehensweisen sind noch nicht perfekt, dass kann man gut an dem „Jahr-2010-Problem“ bei EC- und Kreditkarten erkennen [GBM10]. Eine Möglichkeit den Testprozess weiter zu optimieren bietet der Ansatz des modellbasierten Testen.

Dieser Ansatz wird bei Capgemini sd&m vorangetrieben und mit der Bachelor-Arbeit unterstützt. Diese beschreibt das Design eines Prototypen, der es ermöglicht, aus einem mit dem UML-Testing-Profile verfeinerten UML-Testmodell, Testfälle zu generieren.

2 Testfallgenerierung

Bei der Testfallgenerierung werden Testfälle erstellt, die die Funktionalität des Systems testen. Die Testfälle werden manuell aus der System spezifikation abgeleitet. Dabei entsteht ein hoher Aufwand, der größtenteils aus dem manuellen Kopieren und Einfügen von Informationen aus der Systemspezifikation in das jeweilige Testfallformat besteht. [SL09]

Um diesen Aufwand zu verringern und gleichzeitig die Nachvollziehbarkeit sowie Testabdeckung zu erhöhen, verfolgt die Bachelor-Arbeit den modellbasierten Testansatz [UL07]. Die Arbeit nutzt das von Capgemini entwickelt Testmodell. Dieses basiert auf der Unified Modelling Language (UML) und dem UML-Testing-Profile (UTP) [Ob05].

Das Testmodell ist eine der Anforderungen, die von Capgemini an eine automatisierte Testfallgenerierung stellen. Weiter ist es wichtig, dass der Entwicklungsprozess der Projektteams erhalten bleibt. Die Nachvollziehbarkeit der Generierung ist ebenso erforderlich um den Projektteams einen schnellen Einstieg in das Tool zu ermöglichen. Darüber hinaus sollen verschiedene Datenquellen angebunden werden. Die Evaluierung von der Markt vorhandenen Testfall-Generatoren ergab, dass kein Tool existiert, welches ein UML-UTP-Quellmodell und Aktivitätsdiagramme sowie eine flexible Anbindung von Testdaten unterstützt.

Der entwickelte Prototyp, dargestellt in Abbildung 1, transformiert die Testdaten sowie das UML-UTP-Testmodell in eine In-Memory-Representation. Im Verarbeitungsschritt werden die Testdaten selektiert und mit den bisher logischen Testfällen zu konkreten Testfällen geformt. Diese Testfälle werden dann exportiert, dabei ist das Format frei definierbar.

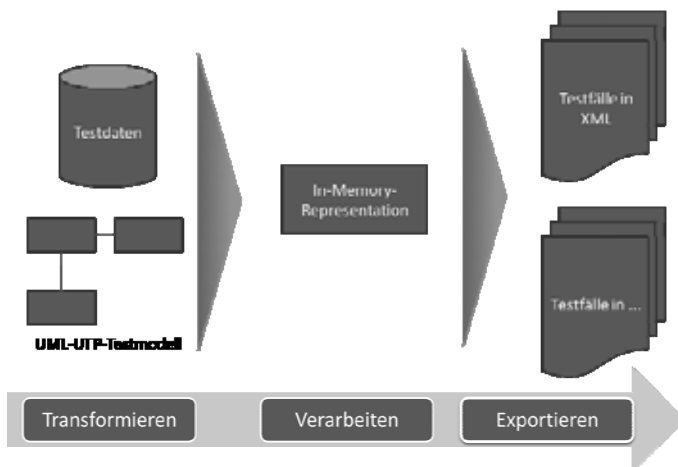


Abbildung 1: Übersicht des Prototypen

Auf die Anforderung von Capgemini sd&m wurde eingegangen, indem als Technologien Java und Groovy [G10] verwendet wurden. Da Groovy zur Laufzeit interpretiert wird und mit der Java Virtual Machine nutzbar ist, wird die Skriptsprache an Stellen verwendet, wo eine hohe Flexibilität gefordert ist.

Zum Beispiel wird Groovy zusammen mit dem Strategie-Entwurfsmuster beim Exportieren verwendet. Dies befähigt die Projektteams das Exportformat mit einem eigenen Skript selbst zu definieren.

Die Nachvollziehbarkeit wurde durch eine Logdatei und dem Exportieren von Zwischenergebnissen realisiert. Damit ist gewährleistet, dass die Projektteams den gesamten Testgenerierungsprozess verfolgen können.

3 Zusammenfassung und Ausblick

Der entwickelte Prototyp unterstützt Capgemini sd&m beim Transformationsprozess des Testmodells in Testfälle und berücksichtigt dabei die Projektumgebungen innerhalb des Unternehmens.

Die Schnittstellen ermöglichen es, neue umfangreiche und schnelle Erweiterungen, wie zum Beispiel den Export von Testfällen in verschiedene Formate oder die Implementierung unterschiedlicher Datenkombinationsmöglichkeiten.

Der aktuelle Stand des Prototypen befähigt abschließend zum Exportieren von Testfällen, die zur manuellen Ausführung bestimmt sind. Eine Erweiterung um aus dem Testmodell direkt automatisch ausführbare Testfälle zu generieren ist wünschenswert, aber ohne zusätzliche Informationen wie zum Beispiel über konkrete Definition von Methoden und Klassen nicht realisierbar.

Literaturverzeichnis

- [Pf03] Pfeiffer, Martin: Berühmt berüchtigte Softwarefehler Therac-25; 09.11.2003; verfügbar: <http://www.uni-koblenz.de/~beckert/Lehre/Seminar-Softwarefehler/Ausarbeitungen/pfeifer.pdf> (Zugriff 17.11.2009)
- [GBM10] Grass, Doris; Brügge, Mathias; Otto meier, Martin: Pleite des Plastikgeldes: Wie die Banken vom Jahr 2010 überrascht wurden; 08.01.2010; verfügbar: <http://www.ftd.de/unternehmen/finanzdienstleister/pleite-des-plastikgeldes-wie-die-banken-vom-jahr-2010-ueberrascht-wurden/50057097.html> (Zugriff 13.01.2010)
- [SL09] Spillner, Andreas; Linz, Tilo: Basiswissen Softwaretest; dpunkt.verlag; 2007
- [UL07] Utting, Mark; Legeard, Bruno: Practical Model-based Testing; Morgan Kaufmann; 2007
- [Ob05] Object Management Group, Inc.: UML Testing Profile 1.0; 07.07.2005; verfügbar: http://www.omg.org/technology/documents/formal/test_profile.htm (Zugriff 21.11.2009)
- [G10] Groovy; verfügbar: <http://groovy.codehaus.org/> (Zugriff 13.01.2010)

Systematisches Abwickeln und strukturiertes Ablegen von Qualitätszielen in der Individualsoftwareentwicklung

Thomas Noack, BTU Cottbus (noacktho@tu-cottbus.de)

Betreut durch

Prof. Dr. Claus Lewerentz, Lehrstuhl Software-Systemtechnik, BTU Cottbus

Abstract: Dieses Papier führt eine praktische Umsetzung eines zielorientierten Messansatzes für das konkrete industrielle Umfeld der Individualsoftwareentwicklung ein. Dazu wird der Ziel/Indikator/Maßnahme Ansatz formuliert, der die für eine Integration in die Unternehmensabläufe großer Unternehmen erforderliche Systematik und Struktur bereitstellt. Im Rahmen der diesem Papier zugrunde liegenden Studienarbeit wurden prozessunterstützende Formulare erarbeitet, die das Einführen und Betreiben des Ansatzes begleiten.

1 Problemstellung

In der Individualsoftwareentwicklung und der damit verbundenen Vielzahl an Projekten werden besondere Anforderungen an das zielorientierte Messen gestellt. Zwar existieren mit Goal Question Metric [BCR94], GQ(Indicator)M [PGF96] und GQM++ [GM97] bereits Ansätze, es kommt jedoch zu einem Konflikt, sobald sich konkrete Fragen in den unterschiedlichen Projekten stellen.

Wegen der Abbildung von Fragen auf Ziele und Metriken auf Fragen für jedes individuelle Projekt werden Analysen zwischen den Projekten erschwert und kontinuierliche Kosten verursacht. Außerdem existiert in GQM kein Modellelement, das die Erfahrungen im Umgang mit den Zielen betrachtet, um sie für zukünftige Projekte zur Verfügung zu stellen.

2 Ziel/Indikator/Maßnahme

2.1 Z/I/M Aktivitäten

Die folgenden Aktivitäten sind nötig, um das systematische Abwickeln und strukturierte Ablegen der Ziele in den Individualsoftwareentwicklungsprozess zu integrieren.

Ziel/ Indikator erheben. Es ist davon auszugehen, dass sich immer wieder neue Ziele und neue Möglichkeiten des Messens entwickeln. Deswegen werden Ziele und Indikatoren kontinuierlich aus dem Entwicklungsprozess erhoben und zur Verfügung gestellt.

Zielplan zusammenstellen. Die einzelnen Projekte unterliegen individuellen Anforderungen. In diesem Zusammenhang enthält ein Zielplan alle für das Projekt relevanten Ziele.

Maßnahme dokumentieren. Die Erfahrungen im Umgang mit den Zielen dürfen nicht in Vergessenheit geraten. Deshalb wird dokumentiert, welcher Indikator ein Problem aufgezeigt hat und welche Aktionen durchgeführt wurden, um ein Ziel zu erreichen.

Zielplan auswerten. Die Auswertung liefert die häufig gefährdeten Ziele und Maßnahmen mit der besten Wirkung auf konkrete Qualitätsprobleme.

Messverhalten analysieren. Damit Ziele den aktuellen Problemen und Indikatoren dem Stand der Messtechnik entsprechen, wird das Messverhalten analysiert. Dabei stellt sich die Frage, warum bestimmte Metriken betrachtet und wie sie interpretiert werden.

2.2 Z/I/M Modellelemente

Ziel. Ein Ziel beschreibt ein Übereinkommen zwischen zwei Parteien und dient der Qualitätszusicherung. Im Laufe der Zeit werden sich sowohl Indikatoren zum Beobachten, als auch Maßnahmen zum Erreichen eines Zieles bewähren.

Indikator. Ein Indikator beleuchtet einen bestimmten Aspekt eines Zieles und verfügt über einen festen Satz an Metriken. Er enthält eine Interpretationsvorlage, die es ermöglicht, ihn anhand der Metrikerwerte unternehmensweit einheitlich zu interpretieren.

Maßnahme. Eine Maßnahme wird durchgeführt, wenn die Metriken eines Indikators ein konkretes Problem aufzeigen und die Zielerreichung gefährdet ist. Sie beschreibt die Erfahrung im Umgang mit einem konkreten Ziel.

2.3 Systematisches Abwickeln und strukturiertes Ablegen

Die strukturierte Ablage der Ziele und Indikatoren bedarf eines unternehmensweit anerkannten Zielkataloges und projektspezifischer Zielpläne. Dies ist notwendig, da Ziele erfasst und kategorisiert sein müssen, damit sie in den Projekten verfolgt werden können. Abbildung 1 zeigt die strukturierte Ablage der Ziele.

Z/I/M Katalog. Im Z/I/M Katalog werden alle bisher erkannten Ziele und zur Verfügung stehenden Indikatoren erfasst. Ziele werden durch einen Pflicht-, Kann-, Neu- und Obsoletstatus kategorisiert.

Zielplan. Da niemals alle Ziele für ein konkretes Projekt relevant sein werden, wird der Zielplan als projektbegleitendes Controllingwerkzeug eingesetzt. Der Zielplan enthält somit nur eine Teilmenge der Ziele des Z/I/M Kataloges. Er bildet eine auf das Projekt und die Anforderungen des Kunden zugeschnittene Zielauswahl.

Kontinuierliche Prozessverbesserung. Eine der Hauptaufgaben des Qualitätsmanagements ist die kontinuierliche Verbesserung der Unternehmensabläufe. Die Daten für die dafür benötigte projektübergreifende Analyse liefern die Zielpläne.

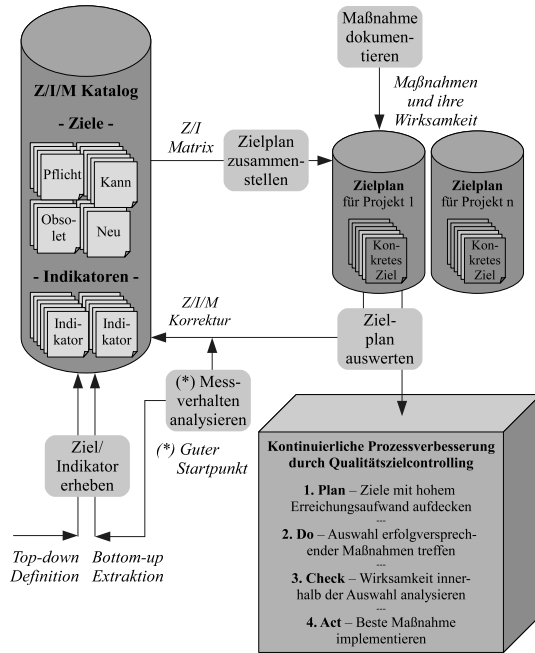


Abbildung 1: Strukturiertes Ablegen

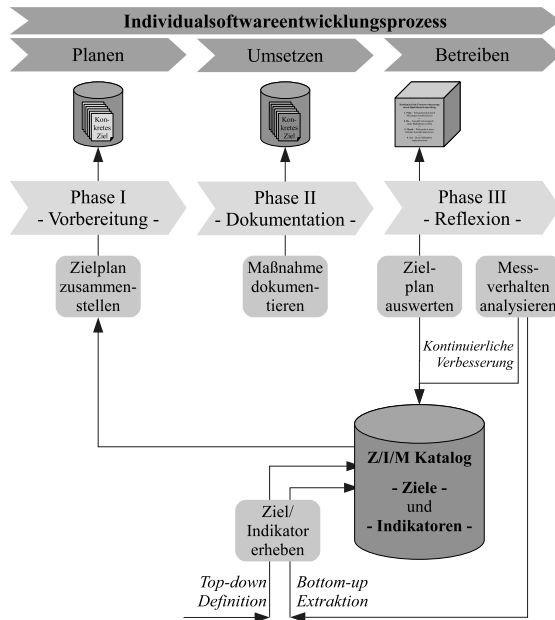


Abbildung 2: Systematisches Abwickeln

Um eine Systematik in der Abwicklung der Ziele zu erreichen, werden die Z/I/M Aktivitäten in den Entwicklungsprozess integriert. Abbildung 2 zeigt die prozessorientierte Anordnung der Aktivitäten.

Phase I - Vorbereitung. Da am Anfang eines Projektes die Anforderungen des Kunden stehen und der Kunde die Rolle des entscheidenden Qualitätstreibers einnehmen sollte, ist die Planungsphase prädestiniert, um Zielentscheidungen gemeinsam zu treffen. Deshalb wird zeitig erarbeitet, welche Ziele in den projektspezifischen Zielplan aufgenommen werden und welche Indikatoren am Besten geeignet sind, um Zielgefährdungen anzuzeigen.

Phase II - Dokumentation. Wenn ein Indikator eine Zielgefährdung anzeigt, werden Maßnahmen ergriffen, um der Gefährdung entgegenzuwirken. Sowohl diese Maßnahmen, als auch die Metrikwerte zum Zeitpunkt der Gefährdung werden dokumentiert.

Phase III - Reflexion. Eine Auswertung der Zielpläne aller Projekte deckt häufig gefährdete Qualitätsziele und Gefährdungsursachen auf. Die Erfahrungen im Umgang mit den Zielen werden gespeichert und für zukünftige Projekte aktualisiert zur Verfügung gestellt.

3 Ausblick

Durch das in dieser Arbeit beschriebene systematische Abwickeln und strukturierte Ablegen von Zielen wurde ein Ansatz zur Verwaltung von Risiken durch Qualitätszielcontrolling vorgestellt. Die Vorteile des Z/I/M Ansatzes liegen in der

- Integrierbarkeit, da die Unternehmensleitung durch ihren fachlichen Überblick Ziele weiterhin top-down definieren kann aber Indikatoren aus den technischen Fähigkeiten der Projektmitarbeiter bottom-up extrahiert werden können,
- Kommunizierbarkeit, da die Unternehmensleitung Ziele für alle Projekte verpflichtend vorgeben und die Projektleitung zusätzliche Ziele für ihr Projekt gemeinsam mit dem Kunden festlegen kann,
- Aktualität, da die Erfahrungen im Umgang mit den Zielen sowohl in die Verbesserung des Z/I/M Kataloges, als auch in die Verbesserung der Unternehmensabläufe einfließen.

Literatur

- [BCR94] Victor R. Basili, Gianluigi Caldiera und H. Dieter Rombach. The goal question metric approach. Encyclopedia of Software Engineering. Seiten 528 – 532. John Wiley & Sons, NY, 1994.
- [GM97] Andrew Gray und Stephen G. Macdonell. *GQM++ A Full Life Cycle Framework for the Development and Implementation of Software Metric Programs*. 1997.
- [PGF96] Robert E. Park, Wolfhart B. Goethert und William A. Florac. *Goal-Driven Software Measurement - A Guidebook*. 1996.

Testgetriebene Entwicklung eines Mehrphasen-Compilers

Michael Pasdziernik

FH Aachen

Fachbereich für Elektrotechnik und Informationstechnik

michael@pasdziernik.net

Art der Arbeit: Überarbeiteter Auszug aus der Bachelorarbeit „Entwicklung eines Mini-JavaOO-Compilers für JVM-Bytecode“

Betreuer: Prof. Dr. rer. nat. Heinrich Faßbender, Prof. Dr.-Ing. Martin Oßmann

Abstract: Die Anforderungen an einen Compiler sind durch die Spezifikation der Quell- und Zielsprache klar definiert. Es bietet sich an, einen Compiler mit Hilfe einer Test-First-Methodik zu entwickeln, indem die Sprachspezifikationen in Form von Testfällen formuliert werden. Im Rahmen des Bachelorprojektes ist ein solcher Ansatz zur Entwicklung eines Mehrphasen-Compilers zum Einsatz gekommen und ein rudimentäres Framework zur Durchführung von Eingabe-Ausgabe-Tests entstanden.

1 Methodik

Die testgetriebene Entwicklung (test-driven development, TDD) ermöglicht einen zielgerichteten iterativen Prozeß, der sowohl die funktionale als auch die strukturelle Qualität von Software sicherstellen kann. Neuer Code wird stets durch noch zu erfüllende Tests motiviert. Die Tests decken fehlende Funktionalität auf. Bereits vorhandene und erfüllte Tests geben die nötige Sicherheit, die Struktur des Codes durch fortlaufende Refactorings zu verbessern. [4]

Üblicherweise erfolgt die testgetriebene Entwicklung mit Hilfe von Unit-Tests, die sich auf die Klassen und Methoden des entwickelten Systems beziehen. Aufgrund der Struktur eines Mehrphasen-Compilers scheint ein anderer Ansatz geeigneter: Ein Compiler durchläuft zur Übersetzung eines Quellprogramms in ein Zielprogramm typischerweise mehrere sequentiell ausgeführte Phasen. Jede dieser Phasen ist gekennzeichnet durch eine Eingabe, einen Transformationsprozeß und eine Ausgabe. Die einzelnen Phasen können testgetrieben entwickelt werden, indem die einzelnen Transformationen in Form von Testfällen spezifiziert werden. Es werden also die einzelnen Phasen des Compilers als Black-box betrachtet und mit Hilfe von Eingabe-Ausgabe-Paaren getestet.

Im Rahmen des Bachelorprojektes wurde ein Compiler zur Übersetzung von MiniJavaOO, einer Teilmenge von Java [1], in Bytecode für die Java Virtual Machine (JVM) [2] entwickelt. Der Compiler ist in ein Frontend und ein Backend aufgeteilt [5]. In Abbildung 1 werden der Aufbau des Compilers und die Ausgaben der einzelnen Phasen

schematisch dargestellt. Die Ausgabe von Phase $n - 1$ ist die Eingabe für Phase n . Im Frontend wird das Quellprogramm in eine für die Weiterverarbeitung optimierte abstrakte Syntax übersetzt. Diese wird durch den abstrakten Syntaxbaum (AST) dargestellt. Der AST wird mit Symbolen und Symboltabellen annotiert. Im Backend wird aus den Informationen im annotierten AST Jasmin-Assemblercode¹ erzeugt, der dann in Bytecode für die JVM übersetzt wird. Um sicherzustellen, daß der generierte Bytecode die strukturellen und semantischen Bedingungen der JVM erfüllt, wird er schließlich durch einen Verifier überprüft.

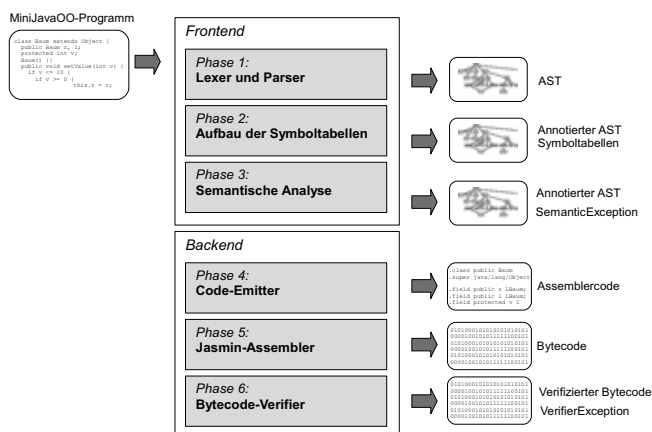


Abbildung 1: Aufbau des Compilers

Die Implementierung des Compilers kann horizontal oder vertikal erfolgen: Entweder wird die gesamte Sprachspezifikation Phase für Phase umgesetzt, oder es wird jeweils ein Teil der Sprache durch alle Phasen hindurch implementiert. Da die Sprache MiniJavaOO eine objektorientierte Erweiterung von MiniJavaFunProc² ist, bot es sich an, zunächst MiniJavaFunProc in allen Phasen zu implementieren und dann mit den Erweiterungen fortzufahren. Setzt man bei der Entwicklung von Phase n voraus, daß die Entwicklung von Phase $n - 1$ abgeschlossen ist, so kann als Test-Eingabe stets ein MiniJavaOO-Programm genutzt werden. Die eigentliche testgetriebene Entwicklung folgt dem Vorgehen im TDD:

1. Formulierung eines möglichst kleinen Teils der Sprachspezifikation als eine Reihe von MiniJavaOO-Programmen für die Eingabe des Testfalls. Formulierung der erwarteten Ausgabe der sich gerade in der Entwicklung befindlichen Compiler-Phase.
2. Durchführen des Tests – er muß scheitern. Beispielsweise bricht der Parser ab, da dessen EBNF-Regeln noch nicht angepaßt wurden. Oder die semantische Analyse wirft nicht die erwartete Exception.

¹Siehe <http://jasmin.sourceforge.net/>

²MiniJavaFunProc dient dazu, während des Compilerbau-Praktikums an der FH Aachen einen Einphasen-Compiler zu realisieren. Siehe <http://www.fassbender.fh-aachen.de/Download/Compilerbau/Folien/>

3. Implementierung der fehlenden Funktionalität in der Compiler-Phase, bis der Test erfüllt wird.
4. Durchführen von Refactorings zur Erhaltung und Verbesserung des Designs und der Code-Qualität. Durch das automatisierte Durchführen aller Tests wird sichergestellt, daß hierbei keine Funktionalität verloren geht.

Zur Spezifikation der erwarteten Test-Ausgabe muß jede Phase eine eindeutige textuelle Darstellung ihres Ergebnisses liefern. Diese ist jedoch nicht zwangsläufig gegeben, so sind die Zwischenergebnisse im Frontend des Compilers Baum- und Listenstrukturen von Java-Objekten. Die Datenstruktur des AST ist Bestandteil des eingesetzten Compiler-Frameworks ANTLR³. Um die Entwicklung von AST-Parsern zu ermöglichen, ist hier eine textuelle Darstellung bereits eingebaut [3]. Die textuelle Darstellung der Symbole und Symboltabellen muß zusätzlich implementiert werden.

2 Framework

Während der Entwicklung von Phase 1 und Phase 2 kommt gUnit⁴ zur Formulierung der Tests zum Einsatz. Es können zwei verschiedene Testarten geschrieben werden. Ein Test kann prüfen, ob das Eingabeprogramm im Sinne der Syntax korrekt ist. Weiterhin kann überprüft werden, ob die vom Parser erzeugte Ausgabe, z. B. ein AST, den Erwartungen entspricht.

Ab Phase 3 wird gUnit den Anforderungen nicht mehr gerecht. Weder kann gUnit mit mehr als zwei hintereinander geschalteten Parsern umgehen, noch kann die auf Java-Exceptions basierende Fehlerbehandlung des Compilers getestet werden. Weiterhin stellt sich heraus, daß die Formulierung eines Eingabe-Ausgabe-Paares sehr umfangreich werden kann und es daher hilfreich wäre, ineinander verschachtelte Textbausteine zu nutzen. Eine Kombination aus StringTemplate⁵, jUnit⁶ und einiger Hilfsklassen (siehe Abbildung 2) schafft Abhilfe.

Die Klasse `CompilerTestBase` bietet Methoden, die jeweils ein MiniJavaOO-Programm entgegennehmen und die Ausgabe der jeweiligen Compiler-Phase zurückliefern. Zur Steuerung des Compilers wird hierzu auf die Klasse `CompilerPhases` zurückgegriffen. Die Tests werden in StringTemplate-Dateien formuliert, welche die verschachtelten Textbausteine beinhalten. So können mehrere Testfälle, die sich nur in wenigen Codezeilen unterscheiden, effizient dargestellt werden. Zum Testen der Fehlerbehandlung kann als Ausgabe eine Exception mit spezifischem Error-Code erwartet werden. Die Klasse `CompilerTestTool` nimmt die Templates entgegen und führt dann die Tests durch. `CompilerTestTool2` ist eine Weiterentwicklung und flexibler in der Schachtelung der Templates.

³Siehe <http://antlr.org/>

⁴Siehe <http://www.antlr.org/wiki/display/ANTLR3/gUnit++Grammar+Unit+Testing>

⁵Siehe <http://www.stringtemplate.org/>

⁶Siehe <http://www.junit.org/>

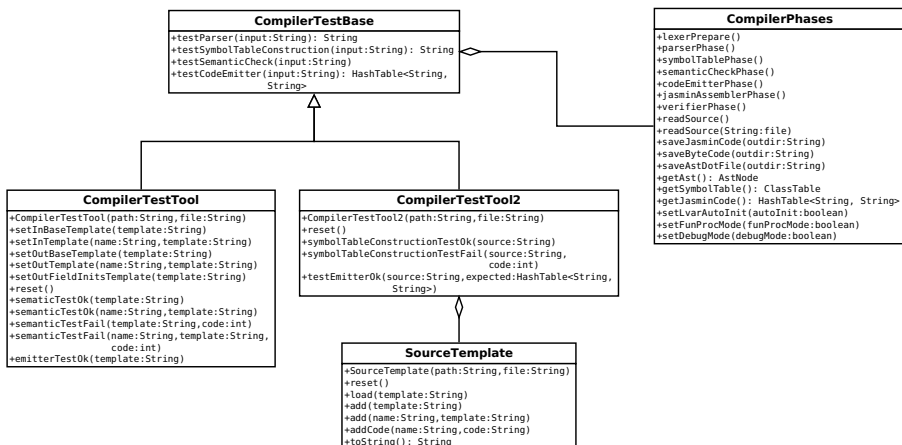


Abbildung 2: Aufbau des Test-Frameworks

3 Erfahrungen

Durch eine sorgfältige Formulierung der Testfälle kann bei der testgetriebenen Compilerentwicklung die Einhaltung der Sprachspezifikation sichergestellt werden und zusätzlich von den Vorteilen des TDD profitiert werden. Die umfangreichen Tests haben die nötige Sicherheit gegeben, auch größere Refactorings durchzuführen: Während der Entwicklung des Code-Emitters hat sich beispielsweise herausgestellt, daß die Struktur der Klassendefinitionen im AST zum Erzeugen des Assembler-Codes ungünstig ist. Um dies zu beheben, mußten alle vorausgehenden Phasen angepaßt werden.

Die oben beschriebenen Test-Klassen haben sich mit den Anforderungen des MiniJavaOO-Compilers entwickelt und können die Grundlage für ein generisches Framework zur Durchführung von Ein-Ausgabe-Tests auf der Basis von jUnit und StringTemplate liefern.

Literatur

- [1] J. Gosling, B. Joy, G. L. Steele: *The Java Language Specification*. Addison-Wesley Longman, 3., aktualisierte Auflage, 2005.
- [2] T. Lindholm, F. Yellin: *The Java Virtual Machine Specification*. Addison-Wesley Longman, 2., aktualisierte Auflage, 1999.
- [3] T. Parr: *The Definitive ANTLR Reference*. S. 149 ff., The Pragmatic Programmers, 2007.
- [4] F. Westphal: *Testgetriebene Entwicklung mit JUnit & FIT*. S. 2 ff., Dpunkt Verlag, 2005.
- [5] N. Wirth: *Grundlagen und Techniken des Compilerbaus*. S. 146 ff., Oldenbourg, 2., bearbeitete Auflage, 2008.

Ein Modellprüfungsansatz für die Erreichbarkeitsprüfung von Process Algebra Nets.

Stephan Prätsch

stephan.praetsch@informatik.uni-halle.de

Martin-Luther-Universität Halle-Wittenberg

Betreuer: Andreas Both, andreas.both@informatik.uni-halle.de

Abstract: Die Sicherstellung der Funktionalität von Service-orientierten Architekturen und Komponenten-Software erfordert neue Vorgehensweisen. In früheren Arbeiten wurde ein diesbezügliches Verfahren vorgestellt, welches auch unbeschränkte Parallelität und unbeschränkte Rekursion darstellen kann. Auf Basis dieser Darstellung ist die Verifikation von Benutzungsverträgen (Protokollen) möglich. Die Verifikation kann auf das Erreichbarkeitsproblem von Petri-Netzen zurückgeführt werden. Aktuell existiert ein Mangel an Werkzeugen, welche in der Lage sind mit unbeschränkten Petri-Netzen umzugehen. In dieser Arbeit wird ein Verfahren vorgestellt, welches diese Probleme adressiert.

1 Motivation

Heutzutage besteht der Softwareentwicklungsprozess zu großen Teilen aus der Wiederverwendung von bereits bestehenden Bausteinen. Diese sogenannten Komponenten (z. B. Webservices) werden so zusammengesetzt, dass sie das benötigte Verhalten zeigen (zuverlässige Komposition). Häufig wird in diesem Zusammenhang von Komponentensoftware oder Service-orientierten Architekturen (SOA) gesprochen. Diese Verfahren gewinnen immer weiter an Bedeutung insbesondere im Bereich EAI und Geschäftsprozessabbildung. Ein besonderes Merkmal ist in diesem Zusammenhang die Auslieferung der Komponenten in einer Blackbox-Variante, d. h. der Quellcode steht zur Auswertung nicht zu Verfügung, einzig die Schnittstellenbeschreibungen sind zugreifbar (z. B. als WSDL).

Die Komposition von Komponenten mit unbekanntem Verhalten ist schwierig und fehleranfällig. Dies ist insbesondere der Fall, falls zustandsbehaftete Komponenten oder dynamische Komposition (wie in SOA vorgesehen) erlaubt sind. Daraus leitet sich der Bedarf von neuen Technologien ab, welche eine robustere und zuverlässigere Komposition als Ziel haben, um somit die industriellen Einsatzgebiete zu erweitern. In [BZ08b, BZ08a] wurden Verfahren vorgestellt, die mittels Komponenten-Protokollen und deren Überprüfung (*protocol conformance*) die korrekte Komposition sicherstellen. Komponenten-Protokolle spezifizieren dabei die korrekte Benutzung einer Komponente unter Betrachtung der zur Verfügung stehenden Schnittstellenoperationen (Reihenfolge, Wiederholung). Zu diesem Zweck wird das abstrahierte Verhalten (enthält nur den relevanten Kontrollfluss) aller zur Verfügung stehenden Komponenten untersucht. Dieses wird als so-

genanntes *Process Algebra Net* [May97] dargestellt. Der Blackbox-Charakter bleibt dabei erhalten. Das Verfahren entscheidet, ob die so zusammengesetzten Komponenten auf jeden Fall die gewünschten Komponenten-Protokolle einhalten (konservative Überprüfung). Es ist bereits zur Übersetzungs- bzw. Bindungszeit anwendbar [BZ08b].

Der Verifikationsprozess wird zurückgeführt auf die Erreichbarkeitsprüfung von Petri-Netzen. Dies ist möglich, obwohl unbeschränkte Rekursion ebenfalls modelliert wird. Die zu betrachtende Modellprüfung ist sehr komplex. Zusätzlich stellte sich heraus, dass es einen Mangel an Werkzeugen gibt, die mit unbeschränkten Petri-Netzen umgehen können. In dieser Arbeit wird ein Verifikationsansatz vorgestellt, der die entdeckten Probleme (soweit möglich) eliminiert und die Berechnung einer Lösung sicherstellt. Besonderer Wert wurde dabei auf Praxistauglichkeit und Erweiterbarkeit gelegt.

2 Abstraktion von Quellcode in einem Prozess-Algebra-Netz

Da Komponenten als Blackbox ausgeliefert werden, muss ein abstraktes Verhalten (des Quellcodes) bekannt gegeben werden. Dieses wird als Process Algebra Net (PAN) abstrahiert (äquivalent zu Kaktusstack, nicht turing-mächtig). Es erfasst Parallelität, Rekursion sowie synchrone und asynchrone Kommunikation, durch Vereinigung von Petri-Netzen und Kellerautomaten [May00]. Das Protokoll von Komponente C_1 der Abbildung 1 könnte fordern, dass alle Aufrufe von a durchgeführt wurden, bevor d aufgerufen werden kann.

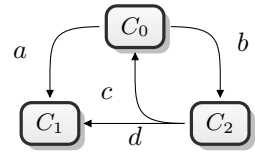


Figure 1: Komponentensoftware bestehend aus drei Komponenten.

3 Erreichbarkeitsprüfung von Process Algebra Nets

Jeder Pfad vom Start- zum Finalprozess $I \Rightarrow f$ beschreibt eine Aufrufreihenfolge, die zu einer unerlaubten Benutzung der betrachteten Komponente führen kann (da Überapproximation). Uns ist keine Implementierung bekannt, die das Erreichbarkeitsproblem für PANs löst. Das Erreichbarkeitsproblem ist reduzierbar auf Petri-Netze: Ein PAN wird in ein Petri-Netz überführt, indem alle sequentiellen Regeln $p_0 \rightarrow p_1.p_2$ durch $p_0 \rightarrow p_2$ ersetzt werden. Wenn in $I \Rightarrow f$ die Regel $p_0 \rightarrow p_2$ genutzt wurde, muss anschließend auch $p_1 \Rightarrow f$ geprüft werden [May97]. Am Ende werden alle Teillösungen zu einer PAN-Lösung zusammengefasst. Die Lösung des Erreichbarkeitsproblems für Petri-Netze benötigt mindestens exponentiellen Speicherplatzbedarf [CLM76, May81]. Jedoch konnten wir kein Werkzeug finden, dass die Erreichbarkeit in unbeschränkten Petri-Netzen lösen kann, ohne dabei abzubrechen (weil immer ein expliziter Überdeckungsbaum konstruiert wurde). Ziel muss es daher sein, ein Verfahren zu entwickeln, dass eine Zustandsexplosion weitgehend ausschließt. Als Nebenziel soll eine verbesserte Performanz angestrebt werden.

In [Prä09] wurde ein entsprechender Algorithmus entwickelt (Abb. 2): Über die Inzidenzmatrix M des Petri-Netzes und einen Zustandswechsel b , in den der Start- und Finalprozess einfließen, wird ein lineares Gleichungssystem (LGS) $M \cdot x = b$ erstellt. Die

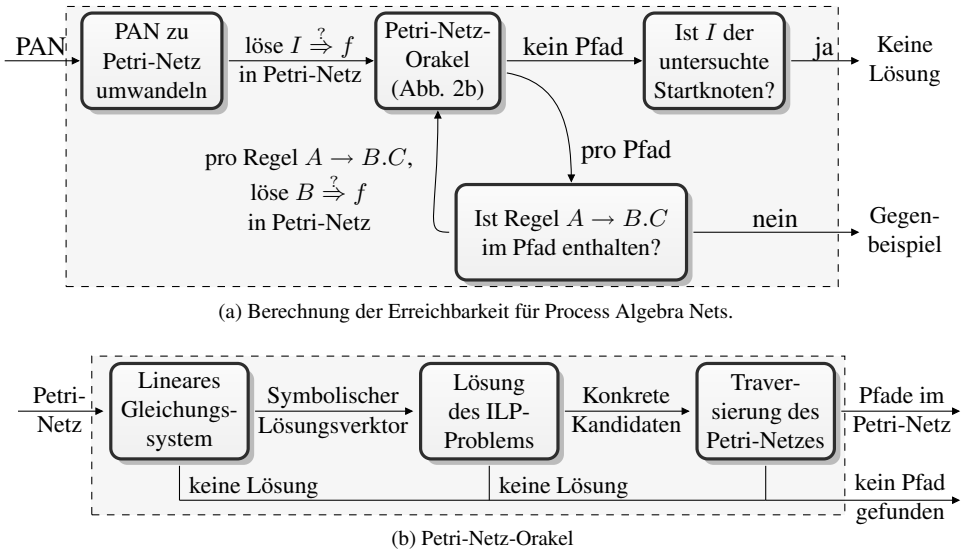


Figure 2: Funktionsweise des PAN-Lösers

Lösung dieses Systems ist ein symbolischer Lösungsvektor x , dessen ganzzahlig positiven Lösungen (im Weiteren als konkrete Kandidaten bezeichnet) einen Weg durch das Petri-Netz beschreiben können. Konkrete Kandidaten geben die Regeln und die Anzahl der Feuerungen vor, um durch das Petri-Netz zu gelangen, aber nicht die Reihenfolge. Um einen Pfad im Petri-Netz zu bestimmen, muss eine Reihenfolge der durch den konkreten Kandidaten gegebenen Regeln im Petri-Netz bestimmt bzw. traversiert werden. Da die Überführung des Petri-Netzes in ein LGS eine Überapproximation ist [Prä09], können unendlich viele konkrete Kandidaten existieren, die im Petri-Netz nicht feuerebar sind. Deshalb wurde die Antwort „weiß nicht“ erlaubt, um einen praktischen Einsatz weiterhin zu gewährleisten. Das Berechnen solcher konkreter Kandidaten ist ein Integer-Linear-Programming-Problem (ILP), das NP-vollständig [Sah74] ist.

In jedem Teilschritt des Petri-Netz-Orakels (LGS, ILP, Traversierung) ist ein Abbruch des Gesamtprozesses möglich, wenn festgestellt wird, dass keine Lösung existiert. Zudem kann jeder Teilschritt, unabhängig von den anderen, optimiert werden.

Der Vorteil der konkreten Kandidaten ist, dass die benutzbaren Ableitungsschritte (ohne Reihenfolge) angegeben sind. Die Zwischenlösung beschränkt also die benutzbaren Regeln, die zur Ableitung eines Gegenbeispiels erlaubt sind als auch auf die Anzahl der Benutzungen jeder Regel. Sackgassen lassen sich so leicht erkennen, da in jedem Schritt nur geprüft werden muss, ob mindestens eine der verbleibenden Regeln des betrachteten Kandidaten anwendbar ist. Andernfalls kann die Ableitung (Pfad) verworfen werden.

Es muss pro sequentieller Regel in einem Pfad, das Petri-Netz-Orakel erneut befragt werden, mit dem jeweiligen sequentiellen Prozess als Start. Letztendlich erhält man Petri-Netz-Pfade durch das PAN. Dieses Verfahren ist keine CEGAR-Schleife (counterexample-

guided abstraction refinement), da Ergebnisse iterativ berechnet werden können, ohne den gesamten Prozess mit einem geänderten Modell neu beginnen zu müssen.

4 Zusammenfassung

In dieser Arbeit wurde ein praxisorientierter Ansatz für die Modellprüfung von PANs (unbeschränkte Parallelität und unbeschränkte Rekursion darstellbar) bzw. der dafür benötigten Petri-Netze vorgestellt. Der Beitrag dieser Arbeit ist gezeigt zu haben, dass mit dem vorgestellten Ansatz eine praxistauglichere Modellprüfung zur Verfügung gestellt wird. Dies ist begründet in der schrittweisen Modellreduktion durch Überapproximationen. Außerdem ist es in jedem Schritt möglich die Modellprüfung abubrechen, wenn ein Gegenbeispiel ausgeschlossen werden kann. Nach unseren bisherigen Erfahrungen tritt dieser Fall häufig auf. Gerade das schrittweise Vorgehen sichert die Kontrolle der Berechnung, so dass unnötige Zustandsexplosionen vermieden werden können. Dies ist ein entscheidender Vorteil gegenüber bisherigen Ansätzen.

Die Architektur ähnelt einem „Pipes and Filters“-Muster. Die erwähnten Überapproximationen können unabhängig voneinander ausgeführt werden oder sind für sich genommen parallelisierbar. Darüber hinaus könnten Heuristiken für die Berechnung der Überapproximationen genutzt werden (soweit diese konservativ sind). Beides ermöglicht weitere Verbesserungen der tatsächlichen Laufzeit. Es wird die Aufgabe in zukünftigen Arbeiten sein, die Auswirkungen der verschiedenen Ansätze zu eruieren.

References

- [BZ08a] A. Both and W. Zimmermann. Automatic Protocol Conformance Checking of Recursive and Parallel BPEL Systems. In *IEEE ECOWS '08*, pages 81–91, Nov. 2008.
- [BZ08b] A. Both and W. Zimmermann. Automatic Protocol Conformance Checking of Recursive and Parallel Component-Based Systems. In Michel R.V. Chaudron and Clemens Szyperski, editors, *Component-based Software Engineering (CBSE2008)*, Oct. 2008.
- [CLM76] E. Cardoza, R. Lipton, and A. R. Meyer. Exponential space complete problems for Petri nets and commutative semigroups. In *STOC '76: Proceedings of the eighth annual ACM symposium on Theory of computing*, pages 50–54. ACM, 1976.
- [May81] E. W. Mayr. An algorithm for the general Petri net reachability problem. In *STOC '81: Proceedings of the thirteenth annual ACM symposium on Theory of computing*, pages 238–246. ACM, 1981.
- [May97] Richard Mayr. Combining Petri Nets and PA-Processes. In *TACS'97: Proc. of the Third Int. Symp. on Theo. Aspects of Computer Software*, pages 547–561. Springer, 1997.
- [May00] Richard Mayr. Process Rewrite Systems. *Information and Computation*, 156(1-2):264–286, 2000.
- [Prä09] S. Prätsch. Implementierung eines Modellprüfungsalgorithmus für Prozess-Algebra-Netze. Master's thesis, Institut für Informatik, MLU Halle-Wittenberg, Sept. 2009.
- [Sah74] S. Sahni. Computationally related problems. volume Vol. 3, pages 262–279, 1974.

Context- and Domain-Aware Semantic Web Services

Sascha Roth¹

`sascha.roth@computer.org`

Bernhard Humm¹

`b.humm@fbi.h-da.de`

Michael Rebstock²

`michael.rebstock@h-da.de`

Hochschule Darmstadt - University of Applied Sciences

— ¹Faculty of Computer Science —

— ²Faculty of Economics and Business Administration —

Abstract: This research proposal introduces an approach for Semantic Web Services that includes context information when querying a service. Thereby, a mapping between context information and domain knowledge helps to reduce the service candidates and ideally enables to compute a subset of formerly undecidable queries without any user intervention.

1 Introduction

“Service composition is today largely a static affair. All service interactions are anticipated in advance and there is a perfect match between output and input signatures and functionality” [PvdH07]. Obviously, service compositions, commonly implemented as executable business processes, lack agility beginning with the way they are built. Semantic Web Services are a promising technology to overcome this issue. It is widely agreed that there does not exist a global ontological model for all possible services, so that specific domain knowledge exists. Queries for requesting a service often do not include the context information when making a request, or, at least do not provide a mapping between the context information and the domain of a service consumer.

This proposal starts with an example using the set theory. First, a naïve example illustrating the core idea is described, such that the major advantage of the proposed solution is clarified and resulting challenges are explained considering related work. Then, expected research outcomes are presented outlining the problem, thesis and impact of a solution. Finally, a brief research agenda gives information on how the proposed approach will be refined and evaluated.

2 Context and Domain Knowledge in Semantic Web Services

As aforementioned, it is widely agreed that no global ontology model for all possible services exists, so that a specific domain knowledge D must exist (see

Figure 1(a)). For a published Semantic Web Service S (see Figure 1(b)), a request is given by query q (see Figure 2(a)). Further, an intuitive processing of the ontology

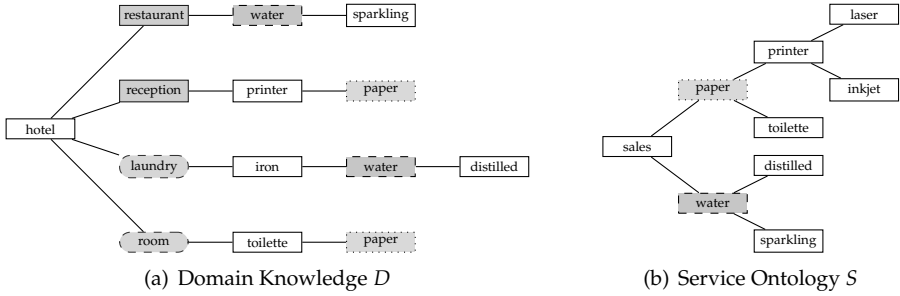


Figure 1: Service and Domain Knowledge

results with an undecidable question having a set of results R (see Figure 3(a)), also known as *homonyms*, since there are several meanings for one term which lead to semantic ambiguity (cf. [RFP08, p. 110]). Hence, it is not computable whether the intention of q is to buy distilled or to buy sparkling water, nor is it decidable to buy printer or toilette paper (cf. Figure 3(a)). Semantic ambiguity is a major problem challenging today's Semantic Web Service community. The additional context information $C|C \subseteq D$ (see Figure 2(b)) enables the computation of $R^* = R \setminus (D \setminus C)$ (see Figure 3(b)). Obviously, there is a more efficient way for calculating the set of results R^* without considering all elements of S and computing all elements of R . However, sending C with q let a mediation engine conclude that $R' = R \setminus R^*$ are services, parameters, or service compositions that are not candidates for answering q , i.e. wrong services, parameters, or service compositions for q . With

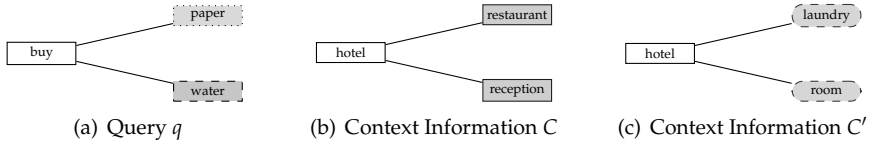


Figure 2: Service Query

the context information C' (see Figure 2(c)), the possible candidates are reduced so that the formerly undecidable query q becomes computable avoiding any user interaction (cf. Figure 3(c)). This approach helps to minimize the amount of user intervention, as required [RFP08, p. 7]. In this example, C and C' could be gained from positioning devices similar as location-based services do. However, context information for employees could also be matched against the domain model, such that more plausible results can be calculated.

Context consideration already has been used (e.g. [RFP08, sec. 8.3]). However, the aforementioned approach is considered as novel since it takes the already existing

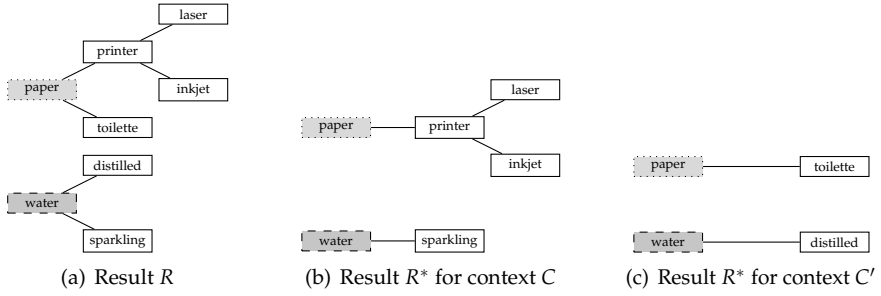


Figure 3: Results

domain knowledge [RFP08, p. 5] as an additional dimension into account when making a decision. Ideally, the context information is an excerpt of the domain knowledge enabling more precise results. Moreover, imprecise context information could query an user intervention such that a context \Leftrightarrow domain mapping knowledge base evolves over time. The introduced approach is a naïve illustration, which avoids facing real world domain knowledge and queries. A major challenge is to close the gap between the introduced approach and real world services. Context information and domain knowledge mapping mechanisms have to be found, combined and evaluated with existing approaches using context information (e.g. [RFP08, sec. 8.3], [YSPL07]). In particular rating mechanisms (see e.g. [RFP08, p. 162]) have to be re-evaluated, since more precise results are expected due to the additional mapping of context information to specific domain knowledge.

3 Final Problem Statement

Service ambiguity needs human user intervention. Obviously, this constitutes a problem, since services ideally are autonomously chosen and invoked. Complete autonomy is an unreachable target. Therefore, approaches have to try to eliminate unreasonable choices, and maximize the amount of computable situations. It is to prove whether or not sending context information with a service request significantly reduces complexity/candidates in terms that less paths have to be analyzed/considered for an invocation. The vision is that a set or subset of formerly not computable queries becomes computable or reduced to a set of more reasonable choices for user interventions, or, in the worst case, remains unchanged. A solution would have multiple impacts. First, better performance is expected due to a reduced lookup effort, i.e. reduced search tree, for services. Second, more precise service recommendations are possible, i.e. automatic service compositions may enable more precise service mediation having analyzed the context within services are used for with respect to the underlying domain. Third, not only service lookups, but rather parameter mapping and service compositions are influenced by this approach. Consequently, it is to expect that all components within a semantic

infrastructure benefit from the additional context information when querying a service. After discussing the impact, the limitations of the introduced approach is outlined. Given query q and context information C , it is not expected to entirely eliminate non-computational cases, but rather to reduce the initial tree of choices. Applied to the example containing useless choices in R , this means reducing to a set of choices that are more reasonable in the given context, i.e. finding R^* . This approach positively influences the learning behavior of an artificial intelligence (AI) component commonly used for service mediation. In particular manually chosen services within a given context could be used to develop and refine the domain model more precisely and to provide a domain/context mapping with learning strategies commonly used in AI engineering.

4 Research Agenda

This section gives a cross-grained overview on how to proceed with the research. A more detailed project plan has to be found in later stages. First, extensive literature review covering, but not limited to, the following topics: *Semantic Web Services, ontology engineering, AI, in particular relevant lazy learning strategies*. Second, evaluation of real world ontology models for a specific domain concerning the possibility to provide context information for service consumers and find an appropriate degree of context granularity. Third, developing concepts of how to provide context information with a query in order to gain on precision in the service mediation process. Fourth, find an existing, ideally mature, code base in order to implement the developed concepts in terms of an exploratory prototype enabling to demonstrate the benefit of the found concepts on a real world example. Fifth, development of the aforementioned exploratory prototype demonstrating practical relevance and feasibility. During research, publishing results and concepts early, e.g. as work in progress, is of great relevance, to gain feedback.

References

- [PvdH07] Mike P. Papazoglou and Willem-Jan van den Heuvel. Service Oriented Architectures: Approaches, Technologies and Research Issues. *VLDB Journal*, 16(3):389–415, 2007.
- [RFP08] Michael Rebstock, Janina Fengel, and Heiko Paulheim. *Ontologies-Based Business Integration*. Springer, Berlin, 2008.
- [YSPL07] Qi Yong, Qi Saiyu, Zhu Pu, and Shen Linfeng. Context-Aware Semantic Web Service Discovery. *Semantics, Knowledge and Grid, International Conference on*, pages 499–502, 2007.

Der Physik-Engine Editor *ped*

Fabian Schönfeld

fabian.schoenfeld@gmx.net

Abstract: Dieser Artikel beschreibt *ped*, den Editor der am Erlanger Lehrstuhl für Systemsimulation entwickelten Physik-Engine *pe*. Die bestehende *pe* C++ API für Starrkörper-Simulationen wird hierbei um ein graphisches Interface erweitert, welches eine in Echtzeit frei modifizierbare *pe* Welt zur Verfügung stellt, in der *pe* Simulationen beliebig aufgebaut, modifiziert und untersucht werden können. Zusätzlich ermöglicht *ped* die freie Manipulation der internen Zeit der *pe*-Welt, wodurch Simulationen nicht länger zu de facto statischen Animationen degradiert werden sondern dynamische Prozesse bleiben, welche zu jedem Zeitpunkt weiterhin Modifizierbar sind. Abgerundet wird das *ped* Paket schließlich durch zahlreiche ergänzende Funktionen die den Überblick über die gesamte Situation erleichtern, wie die Verfolgung einzelner Körper über den Simulationsablauf hinweg, eine zusammenfassende Visualisierung der aktuellen Zeitlinie oder kontextsensitive Zustandsmarkierungen einzelner Körper.

1 Motivation und Überblick

Die Physik-Engine *pe* wird seit 2006 am Lehrstuhl für Systemsimulation der Universität Erlangen von Klaus Iglberger entwickelt [IR09b] und ermöglicht das flexible Aufsetzen großangelegter Rigid Body Simulationen [IR09a]. Der Zugang zum eigentlichen *pe* Framework ist jedoch auf die Nutzung einer C++ API beschränkt, wodurch das Erzeugen neuer Simulationen nur über handgeschriebene Programme möglich ist. Als Konsequenz sind zeitlich dynamische Simulationen mit viel Aufwand verbunden und in der Regel wird sich darauf beschränkt gewünschte Anfangszustände zu erstellen, deren physikalische Interaktionen im Anschluss berechnet und eingeschränkt visualisiert¹ werden können. Modifizierungen des späteren Simulationsverlaufs sind dabei zwar möglich, jedoch sind gezielte Änderungen bestimmter Zeitschritte hierbei mit viel Aufwand verbunden um die jeweiligen Schritte zu identifizieren und das darunterliegende Program entsprechend anzupassen, neu zu kompilieren und auf den gewünschten Effekt hin erneut zu untersuchen.

Als sinnvolle Erweiterung des *pe* Projekts wurde daher die Entwicklung des *pe* Editors *ped* im Rahmen einer Diplomarbeit in Angriff genommen, um das bestehende Framework in ein graphisches Interface einzubetten. Dieses soll den bequemen Zugriff auf die existierende *pe* Funktionalität erlauben und zusätzliche Werkzeuge bieten um dynamische *pe* Simulationen auf einfache Art und Weise handzuhaben. *ped* setzt sich dabei aus drei Kernkomponenten zusammen: dem *pe* Framework zum Berechnen der Interaktionen zwischen

¹Einzelne Zeitschritte können als POV-Ray Bilder ausgegeben werden, während Simulationen als schwarz/weiß Animationen betrachtet werden können

den beteiligten Körpern, der Open Source Irrlicht Grafik-Engine [Eng] um die Simulation in Echtzeit auf dem Bildschirm auszugeben, und dem eigentlichen *ped* Programm, welches *pe* um seine Funktionen erweitert und die drei Komponenten zu einer geschlossenen *ped*-Simulation zusammenführt.

2 *ped* Simulation

Im Gegensatz zu einer traditionellen *pe*-Simulation besteht die Aufgabe einer *ped*-Simulation nicht mehr im Berechnen von zeitlich lokalen Interaktionen zwischen Körpern, sondern um das Erhalten der Konsistenz eines globalen *Zeitrahmens*. Zu diesem Zweck wurde ein duales Protokoll implementiert, welches zum einen über alle beteiligten Körper der Simulation Buch führt (der sog. *Body Pool*), und zum anderen sämtliche Ereignisse des internen Zeitrahmens der Simulation registriert (die sog. *Timeline*). Zusammen mit dem aktuellen Zustand der *pe* Welt und der Darstellung selbiger via Irrlicht Visualisierung, beschreiben diese beiden Module den gesamten Zustand der *ped*-Simulation. Die Aufgabe von *ped* ist es nun diesen Zustand konsistent und frei modifizierbar zu halten.

Das einfachere der beiden *ped*-exklusiven Module ist dabei der Body Pool, welcher als schlichte Liste fungiert und einen Eintrag pro simuliertem Körper speichert. Zu beachten ist hierbei lediglich, dass dies nicht nur die aktuell sichtbaren Körper umfasst, sondern auch jeden anderen Körper der erst im Verlauf der folgenden Simulation oder nur während bereits vergangener Zeitschritte existiert.

Aufwändiger gestaltet sich die Timeline, welche alle Interaktionen des Nutzers als Modifikationen der Simulation protokolliert sowie ausgesuchte Zustände (konfigurierbar: jeder *n*-te Zeitschritt) der Simulation speichert. Die primäre Funktion dieser Datensammlung ist das Ermöglichen freier Zeitsprünge innerhalb des bestehenden Zeitrahmens der Simulation. Sowohl bei Vorwärts- als auch Rückwärtssprüngen wird hierbei auf Timeline-Daten zugegriffen um bereits besuchte Zustände nach Möglichkeit zu rekonstruieren um die Kosten eines Zeitsprunges möglichst gering zu halten.

ped Zeitsprünge verlaufen dabei nach einem einfachen Muster: zunächst wird der gespeicherte Zustand *s* ermittelt welcher dem gewünschten Zeitschritt *z* am nächsten liegt, und dieser rekonstruiert. Anschließend werden alle Schritte bis zum Zielschritt via *pe* simuliert, während gleichzeitig alle protokollierten Events zwischen den Schritten *s* und *z* wiederholt werden.

3 User Interface

Das *ped* User Interface baut primär auf einem bereits vorhandenen GUI Framework der Irrlicht Engine auf, welches bequem Klassen für diverse Fenster, Buttons, etc. zur Verfügung stellt. Das Abfangen und behandeln von Nutzereingaben wird dabei über einen zentralen *Event Receiver* gehandhabt, der alle Eingaben direkt an eine interne *ped* Schicht weiter-

gibt. Diese fungiert als zusätzliche Sicherheitsgrenze und hat alleinigen Zugriff auf die Funktionen der *ped* Simulation.

Das *ped* Interface bietet dem Nutzer einfachen Zugriff auf alle relevanten Funktionen und ist in vier Bereiche unterteilt:

- **Editing Tools:** Hier finden sich alle gängigen Editierfunktionen, d.h. alle Buttons zum Hinzufügen (oder Löschen) beliebiger Körper. Zur Auswahl stehen dabei Primitive (Kugeln, Kapseln, Quader, Ebenen sowie zusammengesetzte Körper, sog. *Unions*), sog. *Presets* (frei konfigurierbare Wände- und Brunnen-Konstruktionen), Zugriff auf den integrierten Union Editor, sowie das Einfügen verschiedenster Generatoren.
- **Warn-System:** Das *GUI Warning System* dient der direkten Kommunikation mit dem Nutzer, markiert Fehler innerhalb der Simulation und gibt Warnungen aus.
- **System Bereich:** Hier kann die aktuelle Simulation gespeichert, und deren Darstellung angepasst werden. Dabei stehen mehrere visuelle Hilfen zur Auswahl, inklusive das Rendern der Simulation als vereinfachte Skizze.
- **Zeitkontrolle:** Dieser Bereich beinhaltet alle Elemente um die interne Zeit der Simulation zu beeinflussen. Dazu gehört ein einfaches Panel, eine intuitiver Zeitschieber und eine knappe Visualisierung der aktuellen Timeline.

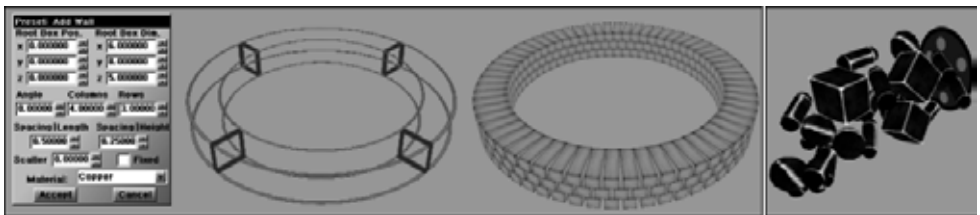


Abbildung 1: Schrittweise erstellen eines Brunnens per Preset Funktion (links) und Zufalls Körper-Generator (rechts)

4 Fazit

Das Einbetten der vorhandenen *pe* Funktionalität in die *ped* Editorumgebung resultiert in zwei wesentlichen Vorteilen: zum einen stellt es weiteren Entwicklungen der *pe* ein komfortables Tool zur Verfügung *pe* Simulationen darzustellen und interaktiv zu modifizieren, was die Implementierung zusätzlicher Neuerungen, sowie das Eingrenzen möglicher Fehlerquellen deutlich erleichtert. Zum anderen öffnet sich die *pe* via ihrem neuen Editor gänzlich neuen Nutzerkreisen, da mittels *ped* neue Simulationen auch gänzlich ohne weitere Programmierkenntnisse aufgesetzt werden können und mit diesen experimentiert und gespielt werden kann.

Literatur

- [Eng] Irrlicht 3D Real Time Visualization Engine. Homepage of the Irrlicht framework: <http://irrlicht.sourceforge.net/>.
- [IR09a] K. Iglberger und U. Rde. Massively Parallel Rigid Body Dynamics Simulation. *Computer Science - Research and Development*, 23(3):159, 2009.
- [IR09b] Klaus Iglberger und Ulrich Rde. The pe Rigid Multibody Physics Engine. Bericht, Chair for System Simulation, University ErlangenNuremberg, 2009.

Prototypische Realisierung einer Integration von Agenten und Workflows

Thomas Wagner
wagner@informatik.uni-hamburg.de

Art der Arbeit: Diplomarbeit im Fach Informatik an der Universität Hamburg
Betreuer: Dr. Daniel Moldt und Prof. Dr. Rüdiger Valk

Abstract: In diesem Bericht werden verschiedene Prototypen vorgestellt, in denen die Workflow- und Agententechnologie so miteinander verbunden werden, dass Vorteile beider Technologien genutzt werden können. Diese Prototypen orientieren sich dabei an einer theoretischen Gesamtarchitektur, die in fünf Stufen schrittweise die Integration der Technologien erweitert.

1 Einleitung und Motivation

Zwei bewährte Technologien für die Entwicklung verteilter Systeme sind die Agenten- und Workflowtechnologie. Diese beiden Technologien bieten die Möglichkeiten große und komplexe Systeme auf eine natürliche Weise zu gestalten. Beide Technologien fokussieren dabei unterschiedliche Aspekte. In der *Agententechnologie* wird die interne Struktur eines Systems klar beschrieben. Soll jedoch das Verhalten des Gesamtsystems dargestellt werden, wird dies mit zunehmender Komplexität immer weniger möglich. Die *Workflowtechnologie* dagegen kann das Verhalten eines Systems gut darstellen. Dafür fällt es hier schwer, die interne Struktur eines Systems zu beschreiben. Beide Technologien besitzen also die gegenseitige Schwäche als Stärke. Die Motivation ist nun, beide Technologien ineinander zu integrieren und so die Vorteile beider für die Anwendungsentwicklung zugänglich zu machen, sowie die Nachteile möglichst auszugleichen.

In [Ree09] wird eine theoretische Architektur vorgestellt, die in fünf Stufen eine solche Integration durchführt. Den *Ausgangspunkt* stellen dabei normale Agenten Management Systeme (AgMS) und Workflow Management Systeme (WFMS) dar, in denen es keine Integration zwischen den Konzepten gibt. In der *zweiten Stufe* wird eine der beiden Technologien dazu verwendet die andere zu realisieren. Dadurch wird es möglich, eine gewisse Integration in die Architektur einzubringen, die sich aber nur hintergründig auswirkt. Die *dritte Stufe* der Gesamtarchitektur erweitert die Möglichkeiten zur Anwendungsentwicklung, indem gleichzeitig der Zugriff auf ein AgMS und ein WFMS erlaubt wird. Daher ist die Integration von Agenten und Workflows zwar bereits beliebig umfangreich möglich, jedoch fehlt ihr eine fest definierte und klare Struktur. Diese wird in der *vierten Stufe* der Gesamtarchi-

tektur eingeführt, indem die Funktionen der beiden Management Systeme in eine gemeinsame Schnittstelle integriert werden. Somit stehen zwar nur wieder entweder Agenten oder Workflows zur Verfügung, jedoch werden intern beide Technologien miteinander verwendet. Die beiden Varianten der vierten Stufe werden für die *fünfte Stufe* miteinander kombiniert, so dass eine beidseitige, vollständige Integration der beiden Technologien entsteht. In dieser Stufe ist es wieder möglich, Agenten und Workflows beliebig miteinander zu verwenden. Im Gegensatz zu Stufe III geschieht dies aber auf eine strukturierte und klar definierte Weise. Im Rahmen der hier zusammengefassten Diplomarbeit [Wag09] wurden Prototypen entwickelt, die sich in die zweite bis vierte Stufe der theoretischen Gesamtarchitektur einordnen lassen und wichtige Schritte in Richtung der fünften Stufe leisten. Diese sollen im Folgenden beschrieben werden.

Zuerst soll aber ein kurzer Überblick über den technischen Hintergrund der vorgestellten Systeme gegeben werden. Alle Prototypen basieren auf den MULAN- und CAPA-Agentenarchitekturen (siehe [Röl04] und [Duv02]). Diese Agentenarchitekturen basieren auf Referenznetzen, einem höheren Petrinetzformalismus. Als Entwicklungs- und Laufzeitumgebung für die Systeme wird der Petrinetzeditor und -simulator RENEW verwendet, welcher zusammen mit den Referenznetzen in [Kum02] vorgestellt wird. Für die Workflows werden auf den Prinzipien der Workflow-Petrinetze (siehe [vdA97]) aufbauende Referenznetze mit Tasktransitionen verwendet (siehe [Jac02]). Durch die jeweilige Fundierung, konzeptionell mittels höherer Petrinetze und technisch mittels Java, ist es möglich, beide Perspektiven zu integrieren.

2 Vorstellung der Prototypen

Der erste Prototyp ist ein *agentenbasiertes WFMS* (AgWFMS). Darin wird die Workflowfunktionalität von einer Menge verschiedener Agenten bereitgestellt, weshalb sich das AgWFMS in die zweite Stufe der Gesamtarchitektur einordnen lässt. Gegenüber der ersten Stufe der Gesamtarchitektur, ist in dieser Stufe bereits eine leichte Integration der Technologien zu beobachten. Durch die Verwendung von Agenten wird z.B. die Verteilung des Systems durch die Autonomie der einzelnen Agenten unterstützt. Den Kern des AgWFMS machen dabei drei Agenten aus. Diese Agenten sind dafür verantwortlich die Workflow-Petrinetze zu erstellen, zu verwalten und auszuführen. Sie schalten auch die Tasktransitionen und verteilen die damit verbundenen Aufgaben an die Benutzer des Systems. Weiterhin gibt es für jeden Benutzer einen Benutzeragenten. Dieser verwaltet die Benutzungsoberfläche des Benutzers und repräsentiert ihn im System. Zusätzlich gibt es Agenten für die Datenbankverwaltung und einen Agenten, der die angemeldeten Benutzer verwaltet. Es ist im AgWFMS möglich, Workflows zu definieren und auszuführen. Die berechtigten Benutzer können verfügbare Aufgaben anfordern, bearbeiten und entweder abschließen oder abbrechen. Dadurch ist die gesamte Funktionalität eines WFMS durch die Interaktionen zwischen den Agenten modelliert.

Die weiteren zwei Prototypen, lassen sich in die vierte Stufe der Gesamtarchitektur einordnen. Sie bauen beide auf dem AgWFMS auf, welches dazu zu einem System der dritten Stufe erweitert wird, indem der Zugriff auf die Funktionen des durch MULAN und CAPA gegebenen AgMS im Hintergrund erlaubt wird. Dadurch wird die Kombination der Agenten- und Workflowanteile vereinfacht, da das AgWFMS auch auf den Agenten aufbaut. Es muss an dieser Stelle angemerkt werden, dass die vorgestellten Prototypen nur einzelne Aspekte dessen realisieren, was prinzipiell durch das Konzept der vierten Stufe der Gesamtarchitektur beschrieben wird.

Zuerst werden die Funktionen des AgMS und AgWFMS der dritten Stufe zu einem WFMS integriert. Prinzipiell erweitert dieses als *Struktur-Agentenworkflow* (S-AgWf) bezeichnete System das AgWFMS um die Möglichkeit, verteilte, hierarchische Workflows auszuführen. Dafür werden intern die Eigenschaften der Agenten ausgenutzt. Im S-AgWf System wird je ein Workflow durch genau einen Agenten gekapselt. Dieser so genannte Strukturagent besitzt dafür einen internen Workflow. Jeder Task in diesem Workflow veranlasst den Strukturagenten einen Subworkflow in einer anderen Instanz des AgWFMS zu starten. Der Subworkflow wird normal in dem anderen AgWFMS ausgeführt. Ist die Ausführung abgeschlossen, sendet das AgWFMS die Ergebnisse zurück an den Strukturagenten, der intern seinen Task abschließt. Auf diese Weise wird der gesamte interne Workflow des Strukturagenten ausgeführt. Die Integration von Workflow- und Agententechnologie ist in diesem System explizit und fest definiert in der Anwendung. Agenten werden automatisch und strukturiert verwendet, um die nach außen hin angebotenen Workflows zu unterstützen. Die Autonomie des Strukturagenten wird direkt dafür verwendet, die Verteilung der verschiedenen Subworkflows zu verwalten. Weiterhin können in zukünftigen Erweiterungen weitere Agenteneigenschaften, wie z.B. Mobilität oder Intelligenz, für den S-AgWf ausgenutzt werden.

Als Nächstes soll das *ProDi-System* vorgestellt werden. Dieses System integriert das AgMS und AgWFMS der dritten Stufe zu einem erweiterten AgMS. Dazu werden Komponenten des AgWFMS so modifiziert, dass es möglich wird, Interaktionen zwischen den verschiedenen Agenten als Workflows zu modellieren. Jeder Task in einem Interaktionsworkflow entspricht einem Protokoll¹, welches von den Agenten ausgeführt wird. Somit kann das ProDi-System als ein WFMS angesehen werden, in dem die Agenten die Tasks ausführen. Die Protokolle werden von dem so genannten *Protocol Dispatcher* (ProDi) Agenten verteilt. Immer wenn ein Task aktiviert ist, sucht der ProDi Agent einen geeigneten Agenten für das entsprechende Protokoll aus. Diesem werden die Startnachricht des Protokolls und alle nötigen Informationen zugesandt. Am Ende des Protokolls sendet der Agent alle relevanten Ergebnisse zurück an den ProDi Agenten, der diese für den Workflow übernimmt und die Ausführung fortsetzt. Auch in diesem System findet die Verbindung von Agenten und Workflows explizit und fest definiert in der Anwendung statt. Es wird der Aspekt des externen Verhaltens der Agenten durch Workflows modelliert. Die Vorteile, die aus dieser Integration entstehen, sind die erhöhte Flexibilität und Fehlertoleranz der Interaktionsausführung. Im Gegensatz zur regulären Ausführung

¹Als Protokolle werden in MULAN und CAPA interne Verhaltensweisen von Agenten bezeichnet.

von Interaktionen in MULAN und CAPA sind die Agenten nicht mehr fest einer Interaktion zugeordnet. Stattdessen können ausfallende Agenten dynamisch durch andere Agenten in der Interaktionsausführung ersetzt werden.

3 Ausblick und Fazit

Die vorgestellten Prototypen vermischen und integrieren Agenten- und Workflow-technologie. Die Ergebnisse können dafür verwendet werden, ein System der fünften Stufe der Gesamtarchitektur zu entwickeln. Insbesondere die Prototypen der vierten Stufe liefern bereits Ansätze und erste Implementierungen für wichtige Aspekte eines solchen Systems. Diese können dann erweitert und miteinander verbunden werden, um ein System der fünften Stufe zu erhalten, welches die Möglichkeiten der dritten Stufe mit den Strukturen der vierten Stufe kombiniert.

Zusammengefasst konnte mit den vorgestellten Prototypen gezeigt werden, dass eine Integration von Workflows und Agenten so möglich ist, dass Vorteile beider Technologien benutzt werden können. Der S-AgWf profitiert als WFMS von den verwendeten Agenten und das ProDi-System profitiert als AgMS von den verwendeten Workflows. Somit können diese Ergebnisse dazu verwendet werden, weitere Arbeiten durchzuführen und die Integration noch weiter voran zu treiben. Die formale Integration auf Basis höherer Petrinetze verspricht zudem eine entsprechende Unterstützung bei der Verifikation gewünschter Systemeigenschaften der WFMS und AgMS.

Literatur

- [Duv02] Michael Duvigneau. Bereitstellung einer Agentenplattform für petrinetzbasierte Agenten. Diplomarbeit, Universität Hamburg, Fachbereich Informatik, 2002.
- [Jac02] Thomas Jacob. Implementierung einer sicheren und rollenbasierten Workflowmanagement-Komponente für ein Petrinetzwerkzeug. Diplomarbeit, Universität Hamburg, Fachbereich Informatik, 2002.
- [Kum02] Olaf Kummer. *Referenznetze*. Logos Verlag, Berlin, 2002.
- [Ree09] Christine Reese. *Prozess-Infrastruktur für Agentenanwendungen*. Dissertation, Universität Hamburg, Department Informatik, 2009.
- [Röl04] Heiko Rölke. *Modellierung von Agenten und Multiagentensystemen – Grundlagen und Anwendungen*, Jgg. 2 of *Agent Technology – Theory and Applications*. Logos Verlag, Berlin, 2004.
- [vdA97] Wil M.P. van der Aalst. Verification of Workflow Nets. In *ICATPN '97: Proceedings of the 18th International Conference on Application and Theory of Petri Nets*, Seiten 407–426, Berlin Heidelberg New York, 1997. Springer-Verlag.
- [Wag09] Thomas Wagner. Prototypische Realisierung einer Integration von Agenten und Workflows. Diplomarbeit, Universität Hamburg, Department Informatik, 2009.

Ellipsis - Containerbasiertes Agentensystem in einer Hochverfügbarkeitsstruktur

Sebastian Apel

Friedrich-Schiller-Universität Jena
Fakultät für Mathematik und Informatik
Lehrstuhl für Softwaretechnik
seb@42-net.de

Art der Arbeit: Diplomarbeit

Betreuer/in der Arbeit: Dipl.-Inf. Volkmar Schau, Prof. Dr. Wilhelm R. Rossak

Abstract: Agenten als Technologie für den Mainstream: Um dem Agentenansatz gerecht zu werden, wurden umfangreiche MAS konzipiert, entwickelt und veröffentlicht. Im Gegensatz zu einer Separierung des Systems von bestehenden Techniken, verfolgt Ellipsis den Ansatz der Integration des MAS in Hochverfügbarkeitsstrukturen. Ellipsis als prototypische Implementierung eines JMX und FIPA basierten Agentencontainers für Java EE Applikationsserver verfolgt die gezielte Integration der Agententechnologie in bestehende und neu zu entwickelnde Businessprozesse, Berücksichtigung anerkannter Standards, sowie die optimale Interaktion zw. Agenten und Java EE Komponenten wie EJBs und Servlets. Weitergeführt als OpenSource Projekt und eine mögliche Aufnahme als JCP, soll die agentenorientierte Entwicklung einer breiten Masse an Entwickler zugänglicher machen.

1 Einleitung

„Agententechnologie wird als ein vielversprechender Ansatz gesehen, die Probleme der Konstruktion komplexer Software-Systeme, wie heterogene Systemumgebungen und dezentrale Strukturen besser beherrschbar zu machen.“ [PBL05] Um dieser Komplexität gerecht zu werden, wurden einige umfangreiche Multiagentensysteme (MAS) konzipiert, entwickelt und veröffentlicht. Diese außerhalb von bestehenden Businessprozessen ausgeführten MAS müssen unter zusätzlichen Arbeitsaufwand mit den jeweiligen Ressourcen gekoppelt werden, für eine optimale Lösung der dem Agenten zugetragenen Aufgaben. Ziel ist die Entkopplung und ein unkomplizierter Einsatz - Agenten als zusätzliches Werkzeug zur Erweiterung existierender oder neuer Businessprozesse.

Der Lehrstuhl für Softwaretechnik, der Fakultät für Mathematik und Informatik, an der Friedrich Schiller Universität Jena forscht seit 1998 im Bereich Agenten [S06]. Ellipsis versteht sich als Nachfolger des 2003 fertiggestellten und seitdem kontinuierlich weiterentwickelten Agententoolkits Tracy2. Im Rahmen der Diplomarbeit wurden Ansätze von Tracy untersucht und mit Spezifikationen, vor allem aus den Bereich der Foundation for Intelligent Physical Agents (FIPA), kombiniert. Unter Einsatz von Spezifikationen aus dem Bereich der Java Enterprise Edition (Java EE) soll die vollständige Integration

des MAS in eine Hochverfügbarkeitsstruktur, wie beispielsweise dem JBoss Applikationsserver, gelingen.

2 Ellipsis

Ellipsis ist ein containerbasiertes MAS innerhalb einer Hochverfügbarkeitsstruktur. Mit diesem Container ist es mobilen Agenten [BR05] möglich, in einem javabasierten Applikationsservers Aufgaben zu lösen, miteinander zu interagieren und über verschiedenen Agency's hinweg zu migrieren. Darüber hinaus können bestehende Businessobjekte in die Problemlösung mit einbezogen und auf verfügbare Skalierungsmechanismen des Applikationsservers zurückgegriffen werden.

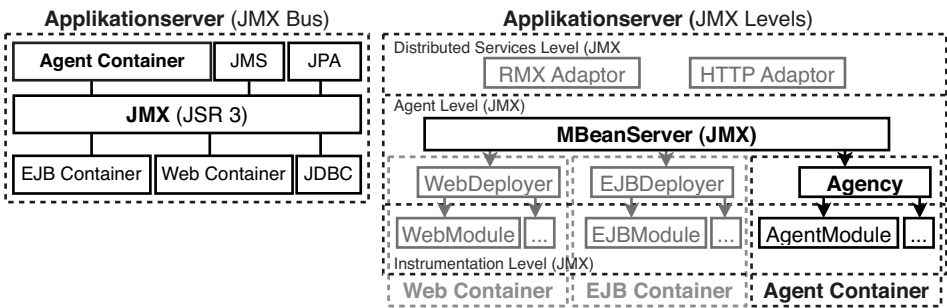


Abbildung 1: Abstrakte Busdarstellung eines JMX basierten Applikationsservers (links); Darstellung der drei JMX MBean Level und die Einordnung der in Ellipsis verwendeten MBeans (rechts)

Der Container liegt im Rahmen der Diplomarbeit als Referenzimplementierung vor und setzt die spezifizierten Ansätze für eine rekonfigurierbare Punkt-zu-Punkt Kommunikation zwischen Agenten, der JMX Service basierten Applikationsserver-Integration (siehe Abb. 1), sowie der Interaktion von Agenten und Java EE Komponenten um. Die Spezifikation beschreibt die Bereitstellung von Services innerhalb des Agenten Containers und stellt eine Ontologie zum Hinzufügen und Abfragen dieser Serviceangebote bereit. Der Einsatz standardisierter Java und FIPA Spezifikationen grenzt den Container nicht für einen Einsatz in speziellen Applikationsservern ein und garantiert einen minimaler Einarbeitungsaufwand. Eine Weiterführung der Entwicklung als OpenSource Projekt und die Aufnahme eines Standardisierungsprozesses im Rahmen eines Java Specification Request (JSR) soll die agentenorientierte Entwicklung einem breiteren Publikum zugänglicher machen.

3 System Architektur

Kernelement von Ellipsis ist eine auf Basis von JMX entwickelte Agency. Diese kann in Java-, bzw. JMX basierte Applikationsserver integriert werden und in Kombination mit anderen Container verwendet werden (siehe Abb. 1). Die Agency wurde mit einem minimalen Funktionsumfang ausgestattet. Sie verwaltet die im System existierenden Agen-

ten, eine Liste an angebotenen Services, ein Verzeichnis zugehöriger Knoten für einen reibungslosen Clusterbetrieb, sowie den Lebenszyklus und die Kommunikation von Agenten. Mithilfe verschiedenen Sicherheitsmechanismen, wie getrennte Klassenlader und Zertifizierung von Code [O04], kann sichergestellt werden, dass Agenten nur innerhalb den ihnen zustehenden Freiraum operieren.

3.1 Kommunikation

Die Kommunikation zwischen Agenten läuft über eine sichere, nachrichtenbasierte, von dem Agentenmodul initialisierte Punkt-zu-Punkt Verbindung. Die mit der FIPA Spezifikation [F02] umgesetzten Nachrichten können über eine zur Laufzeit rekonfigurierbare, Pipeline basierte Kommunikationsleitung an den gewünschten Empfänger übermittelt werden, unabhängig davon, ob der Agent auf dem selben Knoten, im selben Cluster oder weit entfernt ausgeführt wird.

Kommuniziert wird auf Basis von frei definierbaren Ontologien und einer frei wählbaren Sprache zum Nachrichtenaustausch. Kernkomponenten verwenden eine im Rahmen von Ellipsis spezifizierte Serviceontology zur Abfrage und Registrierung von existierenden und neu zu erstellenden Serviceangeboten. Die mit Ellipsis entwickelten Systemagenten verwenden für den Informationsaustausch RDF mit dem Repräsentationsformat RDF/XML und zur Aktionssteuerung das von FIPA vorgeschlagene FIPA RDF-0 Schema [F01]. Aufbauend auf der durch Ellipsis definierten Serviceontology wurde das FIPA RDF-0 Schema mit Rahmenbedingungen erweitert, um eine bessere Schließbarkeit und Zugehörigkeit der einzelnen Properties für die Aktionssteuerung innerhalb von Ellipsis zu ermöglichen.

3.2 Java EE Integration

Mit der nahtlosen Integration in Applikationsserver ist eine für den Agent einfache Interaktion zwischen Java EE Komponenten, wie Servlets, Datenbankverbindungen und Enterprise Java Beans (EJBs), möglich. Parallel können EJBs oder Servlets neue Agenten erzeugen, agentenorientierte Prozesse starten oder Nachrichten an Agenten senden. Der Agentencontainer dient nicht allein als weiterer Service innerhalb einer Hochverfügbarkeitsstruktur, er ist als weiteres Kernelement konstruiert und soll eine einfache agentenorientierte Entwicklung in Applikationsservern ermöglichen.

4 Zusammenfassung und Ausblick

Die prototypische Implementierung des containerbasierten Multiagentensystems im Rahmen der Diplomarbeit stellt einen Anfang dar. Eine frei zugängliche Veröffentlichung als OpenSource Anwendung und Spezifizierung im Rahmen eines Java Community Process (JCP), bzw. weiterführend in einem JSR und die damit verbundene Integration in existierende Hochverfügbarkeitsstrukturen, kann die Entwicklung von Agenten einfacher gestalten und damit die agentenorientierte Entwicklung einer breiten Masse an Entwicklern zugänglicher gemacht werden.

Literaturverzeichnis

- [PBL05] Alexander Pokahr; Lars Braubach; Winfried Lamersdorf: Universität Hamburg, Agenten: Technologie für den Mainstream?, Mai 2005
- [V06] Volkmar Schau: Programming the Plugin-oriented Software Architecture of the Tracy2 Agent Toolkit, 05. Mai 2006, Seite 2
- [F01] FIPA: FIPA RDF Content Language Specification, 10.08.2001, <http://www.fipa.org/specs/fipa00011/XC00011B.html> (besucht Januar 2010)
- [F02] FIPA: FIPA ACL Message Structure Spezifikation, 06.12.2002, <http://www.fipa.org/specs/fipa00061/SC00061G.html> (besucht Januar 2010)
- [J08] JBoss: JBoss Application Server 4.2.2, Januar 2008, http://www.jboss.org/file-access/default/members/jbossas/freezone/docs/Server_Configuration_Guide/beta422/html/index.html, Kapitel 3.1., Abbildung 3.1.
- [BR05] Peter Braun; Wilhelm R. Rossak: Mobile Agents. Concepts, Mobility Models, and the Tracy Toolkit, Februar 2005
- [O04] Scott Oaks: Java Security, 2nd Edition, May 2001, O'Reilly

Weiterführende Literatur

- [S06] Sun Microsystems: Java Management Extensions (JMX) Specification, Version 1.4, November 2006, <http://java.sun.com/javase/technologies/> (besucht April 2009)
- [S06] Scott Stark: How to Run JBoss with a Java Security Manager, 14.03.2006, <http://www.jboss.org/community/wiki/ConfiguringAJavaSecurityManager> (besucht August 2009)
- [RN04] S. J. Russell; P. Norvig: Künstliche Intelligenz. Ein moderner Ansatz, Pearson Education Deutschland, Pearson Studium, 2004. (Seite 66 gesichtet)
- [WT00] Andrew B. Williams; Costas Tsatsoulis: DOGGIE: A Multiagent System for Learning Diverse Web Ontologies, AAAI-2000, Paper, August 2000
- [S69] J. R. Searle: Speech Acts: An Essay on the Philosophy of Language. Cambridge University Press, Cambridge, UK, 1969
- [GFC04] A. Gómez-Pérez; M. Fernández-López; Oscar Corcho: Ontological Engineering: with examples from the areas of Knowledge Management, e-Commerce and the Semantic Web, Springer Verlag, 2. Auflage, 2004
- [W95] M. Wooldridge: Intelligent agents: Theory and practice. Knowledge Engineering Review, 10(2):115–152, 1995

Social Software auf PVR-Systemen

Khadim Diouf

Universität Paderborn - Fakultät für Elektrotechnik, Informatik und Mathematik
Fachgebiet: Didaktik der Informatik
bamba@uni-paderborn.de

Art der Arbeit: Diplomarbeit

Betreuer/in der Arbeit: Wolfgang Reinhardt

Abstract:

Soziale Netzwerke (SN) werden heute weltweit angeboten und sind ein beliebtes Mittel der Kommunikation über alle Kontinente und Kulturen hinweg. Nach allgemeiner Ansicht aller Marktteilnehmer werden alle Systeme der Unterhaltungselektronik und hier insbesondere digitale Set-Top-Box Systeme zur Darstellung von TV-Inhalten in spätestens 5 Jahren über einen Internet - Anschluss verfügen [O08].

Aus diesen beiden Fakten ergibt sich einerseits die Frage, in welcher Art und Weise das große Potential der Sozialen Netzwerke für die Kommunikation innerhalb einer Familie, innerhalb einer Freundesgruppe usw. auf TV-Plattformen effizient genutzt werden kann. Andererseits ergibt sich die Frage wie das Prinzip der sozialen Netzwerke überhaupt auf einer TV-Plattform genutzt bzw. übertragen werden kann, um hier auch die TV-Plattform an das allgemeine Kommunikationsnetz anzuschließen.

1 Social Software im Internet

Als Social Software werden Internet-Anwendungen bezeichnet, die der menschlichen Kommunikation und dem Austausch von Nachrichten oder Daten dienen. Coates [C01] bezeichnet Social Software als „*Software that supports, extends, or derives added value from human social behavior*“.

Der Begriff Social Software umfasst ein breites Spektrum von Anwendungen, die alle als Ziel haben die Verwaltung von Informationen, Identitäten und Beziehungen. Dazu zählen Wiki, Foren, Blogs, RSS, calendar sharing, soziale Netzwerke und Instant Messaging. In dieser Arbeit werden wir uns auf die zwei zuletzt genannten Anwendungssysteme konzentrieren.

Im Internet zählen soziale Netzwerke, wie MySpace, Facebook, StudiVZ und weitere, heute zu den beliebtesten und am meisten frequentierten Internet Seiten. Die Wachstumsraten derartiger Dienste sind enorm und übertreffen bei weitem alle bisher bekannten Dienste. Vor allem jüngere Leute nutzen diese Dienste zur vollständigen Organisation ihres sozialen Umfelds. Aber auch im Geschäftsleben werden diese Netzwerke zur Anbahnung von Kontakten usw.

Andererseits haben wir die Instant-Messaging Software, die auch zur zwischenmenschlichen Kommunikation dient, jedoch in einer anderen Art und Weise. Denn mit z.B. Skype oder ICQ können direkte Verbindungen zwischen zwei oder mehreren Benutzer ermöglicht werden.

2 Ansätze für Social Software auf PVR-Systemen

Die Firma DiscVision¹ in Paderborn stellt einen Server unter dem Namen connectTV zur Verfügung. Dieser Server bildet einen wesentlichen Baustein zur P2P Verbindung zwischen STB² / PVR³ Systemen und anderen IT-Systemen (Computer, Mobiltelefon). Durch die Anbindung aller STB / PVR Systeme an eine zentrale (skalierbare) Datenbank können Informationen zwischen den Systemen ausgetauscht werden. connectTV ermöglicht die Fernsteuerung einer STB / PVR von jedem anderen IT System, sowie den Zugriff auf breitbandige und schmalbandige Inhalte eines PVR Systems.

Der Einsatz von sozialen Netzwerken in PVR-Systemen muss konzeptionell der neuen Umgebung angepasst werden. Dafür müssen neue Konzepte zum Erweitern der bisher bekannten Funktionalitäten überlegt werden.

2.1 Unterschiedliche Szenarien

Im Folgenden werden einige Beispiele für Szenarien sozialer Netzwerke und Konzepte derartiger Netzwerke vorgestellt, wie sie auf TV Plattformen übertragbar werden können. In all diesen Szenarien wird die Nutzung des sozialen Netzwerkes im Wesentlichen auf der TV Plattform durchgeführt.

- **Online-Communication-on-TV:** Über Textmessaging, VoIP oder Videotelefonie kann eine direkte Kommunikation über die TV Plattform zwischen Freunden etabliert werden.
- **Share-the-Moment:** Das Konzept dieses Dienstes erlaubt es den aktuellen Sender, den ein Nutzer gerade sieht, auch auf der STB / PVR seines Freundes einzuschalten. Damit kann man den Effekt simulieren, den man oft hat, wenn man in einer Wohnung seine Mitbewohner auffordert schnell mal zum TV zu kommen, da dort eine interessante Szene läuft.

¹ <http://www.discvision.de>

² STB = Set Top Box

³ PVR = Personal Video Recorder, ein PVR ist ein STB mit integrierter oder über USB2.0 angeschlossener Festplatte, die TV-Inhalte auf dieser Festplatte aufnimmt und für einen Zeitunabhängigen zugriff bereitstellt.

- Push-content-to-friends: Hat jemand einen interessanten Inhalt, z.B. eine TV Aufnahme oder eine Bildliste oder eine Audio Präsentation, so kann er diese Präsentation auf seinem eigenen TV laufen lassen und diese auch noch einige seiner Freunde schicken, um einen zeitsynchronen Ablauf zu simulieren. Somit können alle gemeinsam die Inhalte auf ihrem TV betrachten und parallel darüber sprechen usw.

Durch die Nutzung von Web Services Schnittstellen (Google Open Social⁴ API) sind hier prinzipiell alle Möglichkeiten gegeben, jedes soziale Netzwerk auf der TV-Plattform zu importieren und lauffähig zu machen.

Diese vorgestellten Szenarien erheben keinen Anspruch auf Vollständigkeit, sondern sollen nur exemplarisch zeigen, welche Szenarien sozialer Netzwerke sich auf TV-Plattformen umsetzen lassen. Im Vergleich zu sozialen Netzwerken im Internet werden hier vermehrt visuelle Medien zwischen den Freunden bereitgestellt. Der Anteil an Textnachrichten und textuellen Inhalten ist dagegen sehr viel geringer.

In Abbildung 1 wird prototypisch eine Anwendungsmaske angezeigt, die die Benutzerin „Claudia“ sehen würde, nachdem sie sich beim sozialen Netzwerk angemeldet hat.

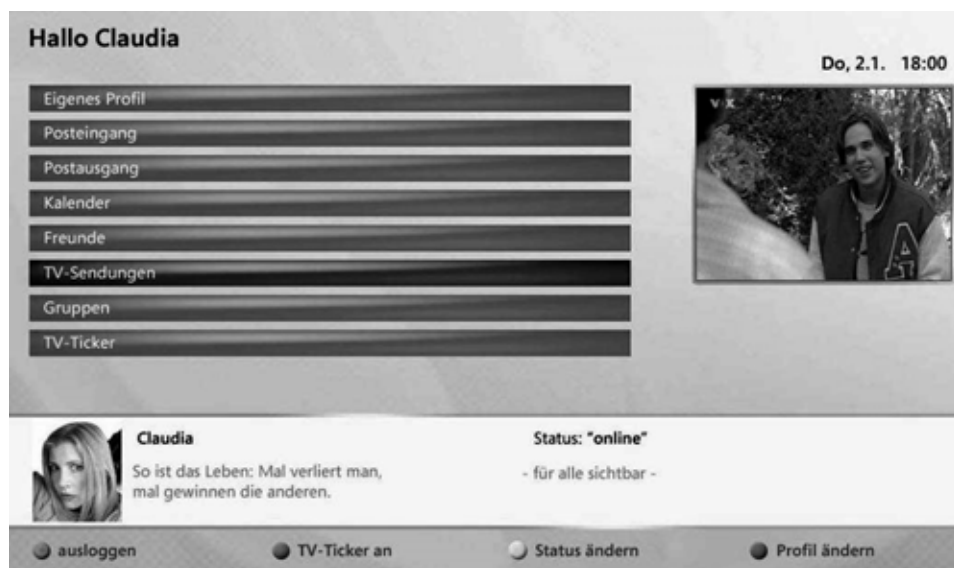


Abbildung 1: Die Menü-Seite von „Claudia“ nach der Anmeldung

⁴ Google Open Social ist eine offene Schnittstelle, die dem Entwickler ermöglicht soziale Anwendungen, auf unterschiedlichen sozialen Netzwerken, ohne große Aufwand zu integrieren.

2.2 Einschränkungen gegenüber webbasierte soziale Netzwerken

Im Gegensatz zu Computern, die über eine Tastatur und eine Maus verfügen, haben wir bei PVR-Systemen standardmäßig nur die Fernbedienung als Eingabemedium. Diese Einschränkung ist natürlich sehr ausschlaggebend in dieser Arbeit, denn dadurch müssen neue Wege erfunden werden für eine benutzerfreundliche Bedienung von social Software basierten Anwendungen.

Die Eingabe von Texten geschieht über die Navigation mit der Fernbedienung auf einer softwarebasierten Tastatur (Speller). Die Bedienung des Spellers beansprucht mehr Zeit als eine reelle Tastatur. Um dieses Problem zu lösen und die Sachen einfacher zu machen plant die Firma DiscVision mittelfristig, eine Tastatur an den PVR anzuschließen oder als zweite Variante den T9-Algorithmus zu integrieren.

Da PVR-Systeme über keine Maus verfügen, kann die Navigation der Elemente so gestaltet werden, dass man mit der Fernbedienung alle Bewegungsmöglichkeiten abdecken kann. Mit den Tasten „links“, „rechts“, „oben“, „unten“ kann man von einem Element zum anderen navigieren, und zum Auswählen bzw. Navigieren von Seite zu Seite werden die Tasten „home“, „exit“ und die „softkeys“ benutzt.

3 Zusammenfassung / Ausblick

In dieser Arbeit wurde beschrieben, wie soziale Netzwerke und Instant Messaging auf PVR-Systemen umgesetzt werden können. Wir konnten zeigen, dass man mit einem PVR-System, soziale Netzwerke und Chat-Software, wie sie bisher am Computer bekannt sind, auf einfache Art und Weise bedienen kann.

Trotz einigen Einschränkungen, wie fehlenden Eingabegeräte (Tastatur, Maus), konnten wir eine benutzerfreundliche Software erstellen, die die meisten Funktionen anbietet, wie es schon in vorhandenen computerbasierten social-software Anwendungen existiert. Um die Bedienbarkeit des Softwares benutzerfreundlicher zu gestalten, sollen die implementierten Anwendungen von textuellen Inhalten sich distanzieren und mehr bildliche und videobasierte Inhalte anbieten, d.h. die Möglichkeiten noch mehr ausnutzen, die eine TV Plattform schon anbietet.

Literaturverzeichnis

- [RK07] A. Richter und M. Koch: Social Software – Status quo und Zukunft, Technischer Bericht Nr. 2007-01, Fakultät für Informatik, Universität der Bundeswehr München, Feb. 2007. http://www.unibw.de/wow5_3/forschung/social_software/
- [C01] T. Coates, (2005): An addendum to a definition of Social Software. http://www.plasticbag.org/archives/2005/01/an_addendum_to_a_definition_of_social_software (letzter Zugriff am 8.1.2009).
- [O01] P. Otellini: CEO at Intel, <http://www.tssf.biz/public-business-strategy-area/business-strategy-news-area/business-news-item-single-view/period/1230598191/article/it-industry-prepares-for-new-growth-segments/17.html>.

Automatically Tagging Online Community Content using Probabilistic Topic Models

M. Döhring^{1,2}, A. Berlea¹, U. Störl²

doehring@stud.fbi.h-da.de,
u.stoerl@fbi.h-da.de, alexandru.berlea@sap.com

¹ SAP Research, ² University of Applied Sciences Darmstadt

1 Motivation and Scope

The continuous growth of information available on the WWW leads to an increase in information overflow in current times. If a search query is not very precisely specified, one can expect hundreds or thousands of potentially relevant search result. The assignment and use of keyphrases, which are relatively small sets of terms reflecting the main topics treated within a document is known of being suitable to mitigate this problem. In collaborative online platforms, people assign their own keyphrases to a variety of items, predominantly web pages, but also to images, video clips, restaurant locations and many more. A user is thereby able to see what other keyphrases users already have assigned to the items. This collaborative keyphrase indexing process is called *tagging* [Med09].

Companies currently carefully inch their way forward concerning the provision of Web 2.0 features to their customers and a majority of online communities in this sector is still in its formation phase. When launching a community, it is wise to provide some initial content and meta-data in the community to encourage user participation and to attract new users. More precisely, when providing a set of documents within the community, it is desirable to also provide some initial tags that can be used to find information and be reused for new documents created by the community members. The crucial problem at this point is, that the manual creation of a controlled vocabulary is usually complex and expensive [Qui05]. This work therefore sketches a completely unsupervised approach for assigning tags to documents from a collection, like e.g. software documentation pages. The tags should be chosen by their ability to best reflect the topic mixture a document deals with.

Table 1: Sample topics as detected in SAP NetWeaver documentation

Topic	1	2	3	4	5	6	7	8
A	database	sql	oracle	maxdb	db	connection	table	jdbc
B	lock	transaction	enqueue	exclusive	release	entry	rollback	owner

2 Outline of Topic-Based Auto-Tagging

The main problems with existing even state-of-the-art unsupervised approaches for keyphrase extraction or auto-tagging are that they either only observe single documents (e.g. [MT04], [GGL09]) or take only single word tags into account (e.g. [DWPP08]). Consequently, it is sometimes hard to detect multi-word terms like `CLIENT_CERTIFICATE` which have a low occurrence count but a high relevancy for a document dealing with a topic mixture of *portal landscapes* and *user authentication* as tags for the document.

For finding an explicit representation of topics corresponding to human intuition in a document collection, [SSRZG04] have proposed the probabilistic topic model (PTM) approach. PTM assume an inherent latent semantic structure in a document corpus which the approach tries to reveal by examining the co-occurrences of words in documents. One can imagine a topic as pattern of word co-occurrences that can be found usually in several documents within the corpus. By dint of the occurrence probabilities inferred for PTM, one can assign relevancy figures for each word-topic and each topic-document relation. Two example topics detected in SAP NetWeaver Composition Environment 7.1 technical document set together with their top 8 words in order of decreasing relevance are presented in Table 1. Just by examining the two word sets, the reader can conclude that topic A deals with *databases*, while topic B is about *synchronization*.

The general idea of looking on topics as probability distributions over single words has been extended by [WMW07] for multi-word terms. One main contribution of our work is the sound formulation of a topic-based relevancy measure for (multi-word) tags within document collection and the proper definition of its alignment within a pre- and postprocessing pipeline as depicted in Figure 1. As a proof of concept, this pipeline has been implemented using LinguistX Platform¹ for tokenization, POS-tagging and lemmatization and Mallet [McC02] for model inference.

The pipeline takes arbitrary text or HTML documents and first removes irrelevant content sections e.g. by dint of predefined XPath patterns. Next it applies standard natural language preprocessing steps removing tendentially irrelevant word tokens and collapsing similar word forms to their common morphological root. It then executes the inference procedure described in [WMW07], which is a kind of Markov-Chain Monte-Carlo. All proper-noun single words, multi-words which are considered to belong together in most of the states of the Markov-Chain as well as additional terms (e.g. a whitelist of product names) are extracted as tag candidates. Based on the topic model and an optionally available thesaurus, our approach is also capable of deciding e.g. whether to expand abbreviations like UME to `unplanned_material_exchange` or `user_management_engine`. After

¹<http://www.inxightfedsys.com/products/sdks/lx>

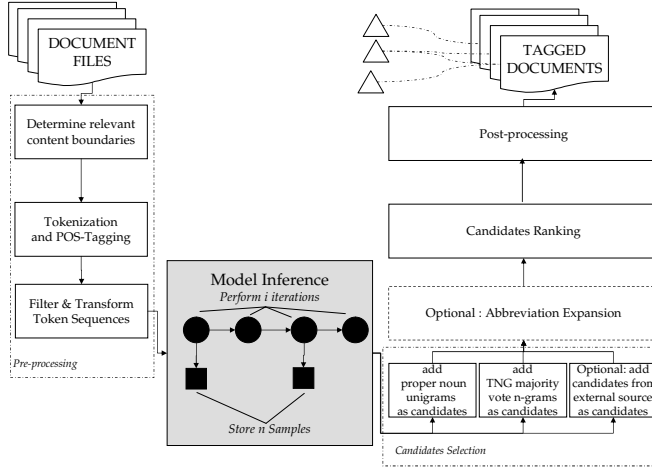


Figure 1: Steps for Topic-Based Tagging

the top-ranked candidates according to our proposed measure have been determined and have gone through a post-processing step (e.g. filtering tags not assigned to a minimum of other documents in the collection), the tags are attached to their corresponding documents.

3 Results of the User Evaluation

The implementation described in the previous section has been applied to 12801 online documentation pages of SAP SAP NetWeaver Composition Environment 7.1. The 10 randomly drawn representative documents together with their automatically assigned tags were presented to two user groups via a web interface. One group consisted of 30 SAP employees, one of 11 students in computer science. All participants were presented the automatically assigned tags for rating as well as eventually tags from other users. Moreover, missing tags could be added per document. With the obtained indexer-consistency and tag suitability measures obtained from the ratings, the following findings can be supported:

- Tagging in technical documents is a quite subjective issue, as users partly tend to disagree on what is the optimal tag set for a document.
- The tag sets delivered by our approach are at least comparably accepted by the participants than those tags added by other participants.
- Our approach tendentially produces tags that are better accepted by participants when rated in isolation than those tags added by other participants (e.g. highly important for tag-clouds).

Moreover, on public gold-standards ([NK07], [Hul04]) for keyphrase extraction, we achieved comparable or better performance than [ZG09] reported for state-of-the-art procedures.

4 Future Work

Our approach could benefit from enhancing the relevancy measure by including reference corpus statistics e.g. from Wikipedia. Furthermore, we did not fully exploit the richness of HTML information (like bold tags). Regarding the growth degree of online content, also the scalability of the topic model inference algorithm should be studied further.

References

- [DWPP08] Mark Dredze, Hanna M. Wallach, Danny Puller und Fernando Pereira. Generating summary keywords for emails using topics. In *IUI '08*, Seiten 199–206, New York, USA, 2008.
- [GGL09] Maria Grineva, Maxim Grinev und Dmitry Lizorkin. Extracting key terms from noisy and multitheme documents. In *WWW '09*, Seiten 661–670, New York, NY, USA, 2009. ACM.
- [Hul04] Anette Hulth. Combining Machine Learning and Natural Language Processing for Automatic Keyword Extraction. Department of Computer and Systems Sciences, Stockholm University, 2004. PhD Thesis.
- [McC02] Andrew Kachites McCallum. MALLET: A Machine Learning for Language Toolkit, 2002. <http://mallet.cs.umass.edu>, Last Accessed January 11th 2010.
- [Med09] Olena Medelyan. What do subject indexing, keyphrase extraction and autotagging have in common? Terminology clarification, 2009. <http://maui-indexer.blogspot.com/2009/07/what-do-subject-indexing-keyphrase.html>, Last Accessed January 11th 2010.
- [MT04] R. Mihalcea und P. Tarau. TextRank: Bringing Order into Texts. In *Proceedings of EMNLP-04 and the 2004 Conference on Empirical Methods in Natural Language Processing*, July 2004.
- [NK07] Thuy Nguyen und Min-Yen Kan. Keyphrase Extraction in Scientific Publications. In *Lecture Notes in Computer Science*, Jgg. 4822, Seiten 317–326. Springer, 2007.
- [Qui05] Emanuele Quintarelli. Folksonomies: power to the people. 2005. <http://www.iskoi.org/doc/folksonomies.htm>, Last Accessed January 11th 2010.
- [SSRZG04] Mark Steyvers, Padhraic Smyth, Michal Rosen-Zvi und Thomas Griffiths. Probabilistic author-topic models for information discovery. In *KDD '04*, Seiten 306–315, New York, NY, USA, 2004. ACM.
- [WMW07] Xuerui Wang, Andrew McCallum und Xing Wei. Topical N-grams: Phrase and Topic Discovery, with an Application to Information Retrieval. In *Proceedings of the 7th IEEE International Conference on Data Mining*, Seiten 697–702, 2007.
- [ZG09] Torsten Zesch und Iryna Gurevych. Approximate Matching for Evaluating Keyphrase Extraction. In *Proceedings of the 7th International Conference on Recent Advances in Natural Language Processing*, Borovets, Bulgaria, Sep 2009.

Image-based Tracking

Jens Maiero

Hochschule Bonn-Rhein-Sieg

jens.maiero@hochschule-bonn-rhein-sieg.de

Abstract: In Mixed Reality (MR)-Umgebungen wird die Realität des Benutzers mit virtuellen, computergenerierten Objekten erweitert. Um virtuelle Objekte in einer MR-Umgebung korrekt zu visualisieren, wird die Position und Orientierung des Benutzers bzw. der Kamera benötigt. Das Bestimmen und Verfolgen von Blickpunkt und -winkel, das sogenannte Tracking, stellt einen wesentlichen Bereich von MR-Applikationen dar. In aktuellen Systemen erweist sich die Initialisierungsphase oft als zeitintensiv und komplex. Deshalb wurde in dieser Arbeit ein markerloses Trackingsystem entwickelt, das im Gegensatz zu anderen Trackingverfahren [RD06, LF05] eine vergleichsweise schnelle und unkomplizierte Initialisierungsphase bereitstellt.

1 Einleitung

Das hier vorgestellte merkmalsbasierte Trackingsystem verfolgt strukturierte planare Objekte. Der Fokus dieser Arbeit lag auf einer schnellen, unkomplizierten Initialisierungsphase. Diese setzt sich aus einer Objekterkennung und einer anschließenden relativen Positionsbestimmung zwischen realem Objekt und Blickpunkt/-winkel des Benutzers zusammen.

In der schnellen Initialisierungsphase muss lediglich ein planares, strukturiertes Objekt im aktuellen Frame markiert werden. Hierbei unterscheidet der Tracker zwei Fälle:

1. **Initialisierung mit Marker:** In diesem Fall wird das reale Objekt mit Hilfe eines Markers *vermessen*, demnach ist eine Positionsberechnung zwischen Kamera und Objekt exakt möglich.
2. **Initialisierung ohne Marker:** Ist kein Marker vorhanden oder das Anbringen eines Markers nicht möglich, kann der Tracker dennoch initialisiert werden. Dabei wird ein künstlich generierter Marker in das markierte Objekt gezeichnet. Die Positionsberechnung ist in diesem Falle nur bis auf einen Skalierungsfaktor möglich. Trotzdem kann das Objekt problemlos augmentiert werden.

In der Online-Phase wird mit Hilfe eines Merkmalsdetektors die Abbildung (Homographie) zwischen Referenzobjekt und dem aktuellen Kamerabild berechnet. Anschließend wird das Referenzobjekt transformiert und einem Markertrackingsystem übergeben, welches die Position und Orientierung berechnet. Mit Hilfe der berechneten Position wird im letzten Schritt die Umgebung mit einem dreidimensionalen Objekt erweitert (siehe Abbildung 1).

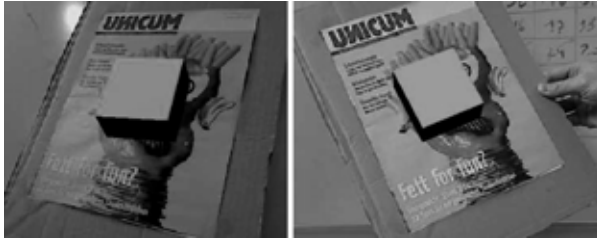


Abbildung 1: Image-based Tracking: dreidimensionale Ansicht einer augmentierten Zeitschrift

2 Verwandte Arbeiten

Computer Vision-Algorithmen zur Objekt- bzw. Kameraverfolgung können in markerbasiertes und markerloses Tracking gegliedert werden. Dabei wird primär zwischen den verwendeten Algorithmen unterschieden. Darüber hinaus werden die meisten Trackingsysteme auch in Bezug auf die verwendeten Endgeräte untersucht [WLS08].

Ein **markerbasiertes** System berechnet die Pose der Kamera anhand von vordefinierten bzw. bekannten Markern. Die herkömmlichen Marker können mit einem normalem Drucker ausgedruckt werden und besitzen in der Regel eine bekannte Größe und Form, die zur Posebestimmung dient. Die Identifikation des Markers erfolgt über ein Muster im Inneren [Kat04, WLS08]. Ein Nachteil des markerbasierten Ansatzes ist, dass ein Anbringen solcher Marker in manchen Umgebungen nicht ohne weiteres möglich ist bzw. sie aus ästhetischen Gründen nicht erwünscht sind. Außerdem ist ein markerbasiertes Tracking in Outdoor-Szenarien schwierig realisierbar, da der Bewegungsradius eingeschränkt ist.

Das sogenannte **markerlose** Tracking, kommt ohne künstliche Referenz-Marken aus. Die Kamerapose wird aus natürlichen Features der Umgebung berechnet. Bei einem monokularen Kamerabild werden die Posedaten anhand von wiedererkennbaren Features bzw. Objekten von Frame zu Frame oder über ein natürliches Referenzobjekt berechnet. Für ein stabiles Tracking sind die Methoden der Extraktion und Wiedererkennung von Merkmalen oder auch das Matching von Konturen von großer Bedeutung. Die Tatsache, dass keine bzw. nur wenige Informationen über die Umgebung vorhanden sind, bringt einen Nachteil mit sich: die Initialisierungsphase, also das Bestimmen der absoluten Kamerapose ist ohne Referenzobjekte oder einen manuellen Schritt nicht möglich. Deshalb wird in markerlosen Verfahren, die eine absolute Position berechnen, auf natürliche Referenzobjekte zurückgegriffen [SFZ00, LF05, RD06, ED06].

3 Implementierung

Die Realisierung kann in zwei Teilgebiete unterteilt werden, einmal die Merkmalsextraktion und darauf folgend der Matching-Algorithmus. Bevor ein definiertes Objekt getrackt werden kann, muss ein Referenzbild des zu verfolgenden Objektes in der Offline-Phase er-

stellt werden. Hierzu werden sogenannte Speeded Up Robust Features (SURF) [BTV06] aus dem Kamerabild extrahiert und mit dem dazugehörigem Frame abgespeichert.

In der Online-Phase werden in jedem Frame ebenfalls Merkmale extrahiert, um diese mit dem erstellten Referenzbild zu matchen. Ist eine hinreichend gute Korrespondenz gefunden wird eine 2-dimensionale Abbildung zwischen Referenz und Kamerabild berechnet.

3.1 Merkmalsextraktion

Die Wahl eines geeigneten Merkmalsdetektor hat erheblichen Einfluss auf den Tracker in Bezug auf Performanz, Genauigkeit und Robustheit und ist somit einer der essentiellen Teile der Arbeit [CXH07]. Die eingesetzte SURF-Bibliothek, OpenSURF, hat in den Tests die besten Ergebnisse für die genannten Eigenschaften erzielt. Dabei ist zu beachten, dass die Performanz abhängig von der Kamera-Auflösung ist.

Ein gutes Merkmal unterliegt mehreren Kriterien. So sollte z.B. die Wiedererkennungsrates eines Features hoch sein. Dies gewährleistet, dass dieselben Merkmale in unterschiedlichen Bildern robust wiedergefunden werden. Die Merkmale sollen sowohl in rotierten als auch skalierten Bildern wieder auffindbar sein. In Hinblick auf die mathematischen Eigenschaften bedeutet dies, dass Merkmale Skalierungs- und Rotationsinvariant sein müssen.

3.2 Korrespondenz-Suche

Bei Korrespondenz-Suche werden in jeden Frame die verwendeten SURF-Merkmale extrahiert und mit dem gespeicherten Keyframe/Referenzobjekt verglichen. Hierfür wurde der Random Sample Consensus Algorithmus (RANSAC) implementiert. Der Algorithmus zeichnet sich dadurch aus, dass er eine große Datenmenge mit einer verhältnismäßig hohen Anzahl an Ausreißern an ein vorgegebenes Modell anpassen kann [FB87].

Aus den korrespondierenden Punktpaaren wird eine zweidimensionale Abbildung berechnet. Diese Transformation, auch Homographie, bildet Punkte einer Ebene auf eine andere Ebene durch eine lineare Transformation ab, wobei die Homographie-Matrix H nicht singulär, das heißt invertierbar sein muss [HZ04].

$$x'_i = H \cdot x_i \tag{1}$$

Wobei x_i und x'_i die homogenen Koordinaten der korrespondierenden Punkte sind und H die Homographie. Die Lösung der Gleichung 1 wird über die direkte lineare Transformation (DLT) bestimmt.

Das transformierte Referenzbild, welches ein Marker enthält, wird einem Markertrackingsystem übergeben, um so die relative Position zwischen Objekt und Benutzer zu bestimmen.

4 Ergebnisse und Ausblick

Bei der Evaluation wurden unterschiedliche Messungen durchgeführt, so wurde das System in Bezug auf Genauigkeit, Performanz und variierende Umgebungsparameter untersucht. Das System zeigt zufriedenstellende Ergebnisse in Hinblick auf Genauigkeit und variierende Umgebungsparameter. Der Flaschenhals, die Performanz, liefert mit 5 Frames pro Sekunde noch keine echtzeitfähige Daten. Durch die Auslagerung der Merkmalsextraktion auf die GPU wurden bis zu 20 Frames pro Sekunde erreicht.

Das Anwendungsgebiet des Trackers ist vielseitig, so könnte dieser beispielsweise als Interaktionsgerät für dreidimensionale Umgebungen oder in augmentierten Umgebungen zum Einsatz kommen. Des Weiteren könnte durch eine komplexere Transformation nicht nur Ebenen sondern auch dreidimensionale Objekte getrackt werden.

Literatur

- [BTVO6] Herbert Bay, Tinne Tuytelaars, , and L. Van Gool. SURF: Speeded Up Robust Features. In *9th European Conference on Computer Vision*, Graz Austria, 2006.
- [CXH07] D. Cheng, Shane Xie, and E. Himmerle. Comparison of local descriptors for image registration of geometrically-complex 3D scenes. In *Mechatronics and Machine Vision in Practice, 2007. M2VIP 2007. 14th International Conference on*, pages 140–145, 2007.
- [ED06] E. D. Eade and T. W. Drummond. Edge landmarks in monocular SLAM. In *British Machine Vision Conference (BMVC)*, volume 1, pages 469–476, 2006.
- [FB87] Martin A. Fischler and Robert C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. pages 726–740, 1987.
- [HZ04] Richard Hartley and Andrew Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, March 2004.
- [Kat04] Hirokazu Kato. Developing AR Applications with ARToolKit. In *Mixed and Augmented Reality, 2004. ISMAR 2004. Third IEEE and ACM International Symposium on*, pages 305–305, Nov. 2004.
- [LF05] Vincent Lepetit and Pascal Fua. Monocular model-based 3D tracking of rigid objects. *Found. Trends. Comput. Graph. Vis.*, pages 1–89, 2005.
- [RD06] G. Reitmayr and T.W. Drummond. Going out: robust model-based tracking for outdoor augmented reality. In *Mixed and Augmented Reality, 2006. ISMAR 2006. IEEE/ACM International Symposium on*, pages 109–118, Oct. 2006.
- [SFZ00] G. Simon, A.W. Fitzgibbon, and A. Zisserman. Markerless tracking using planar structures in the scene. In *Augmented Reality, 2000. (ISAR 2000). Proceedings. IEEE and ACM International Symposium on*, pages 120–128, 2000.
- [WLS08] Daniel Wagner, Tobias Langlotz, and Dieter Schmalstieg. Robust and unobtrusive marker tracking on mobile phones. In *ISMAR '08: Proceedings of the 7th IEEE/ACM International Symposium on Mixed and Augmented Reality*, pages 121–124. IEEE Computer Society, 2008.

CARD-PAL – Ein kartenbasierter Ansatz zur Visualisierung und Formalisierung statischer Aspekte während der Anforderungsanalyse

Sarah Möckel
sarah.moeckel@uni-jena.de

Betreuung:
Prof. Dr. W. Rossak, Dipl.-Inf. Ivonne Erfurth
Lehrstuhl für Softwaretechnik
Friedrich-Schiller-Universität Jena

1 Stand der Forschung

In der objektorientierten Softwareentwicklung hat sich die UML [OMG07] als Modellierungs- und Beschreibungssprache etabliert. In den frühen Entwicklungsphasen werden den Auftraggebern und Kunden UML-Diagramme, insbesondere das Klassendiagramm, jedoch kaum zur Besprechung vorgelegt [ER07]. Diese Notation wirkt in ihrer Vielfalt der Beschreibungsmöglichkeiten häufig erdrückend und bietet einen beachtlichen Detailreichtum [Oes04]. Damit ist die UML für externe Stakeholder nicht (intuitiv) verständlich.

Chaos Reports der Standish Group [Gro95, Gro01] betonen, dass die frühzeitige Einbeziehung der Auftraggeber und Kunden ein wichtiger Faktor für den Erfolg eines Softwareprojekts ist. Diverse Techniken der partizipativen Softwareentwicklung [FMR⁺89] widmen sich nicht nur der frühen Einbeziehung externer Stakeholder, sondern ziehen sie als mitgestaltende Experten im Entwicklungsprozess zu Rate. Unter diesen Techniken sind die Kartenspiele CARDS [TMDR93, Mul01], CUTA [Laf96] und TOD [DMK98] zu finden. Allesamt sind sie vorwiegend Techniken des GUI-Designs oder stellen dynamische Systemaspekte dar. Für die Analyse und Spezifikation eines Softwaresystems müssen auch statische Aspekte erfasst werden. Mit dem CRC Rollenspiel [BC89] können zwar statische Aspekte analysiert werden; doch ist CRC keine partizipative Technik und als Dokumentation für den Kunden ungeeignet.

Am Lehrstuhl für Softwaretechnik der Friedrich-Schiller-Universität wird am Referenzmodell UPEX [ER05] zur kundenorientierten Entwicklung von Softwaresystemen gearbeitet. In diesem kommen verschiedene partizipative Techniken zum Einsatz. Neben einem CUTA-ähnlichem Kartenspiel zur Analyse von Arbeitsprozessen [ER07], wird an einer kundenfreundlichen Notationsform zur Erfassung statischer Systemanforderungen gearbeitet. Diese Notation soll in ein UML-Klassendiagramm formalisierbar sein.

2 CARD-PAL

CARD-PAL¹ ist das Konzept einer kartenbasierten Technik, welche der Erkundung und Dokumentation statischer Anforderungen dient. Karten und Szenarien sind die beiden Notationselemente von CARD-PAL, welche bei der Kommunikation mit den Kunden zum Einsatz kommen. Karten repräsentieren Objekte und Begriffe aus der Welt des Kunden. In (statischen) Szenarien werden strukturelle Beziehungen und das Zusammenwirken dieser Objekte visualisiert.

Für den Einsatz der beiden Notationselemente ist folgendes Vorgehen ausgearbeitet worden: Ausgehend von einem konkreten Arbeitsprozess werden Geschäftsobjekte detektiert und auf CARD-PAL-Karten dokumentiert. Diese Karten finden sich auch in den danach zu erstellenden Szenarien wieder. Anschließend werden die Szenarien in Objektdiagramme überführt, deren Informationen wiederum in einem UML-Klassendiagramm abstrahiert werden.

2.1 Die Karten

Bei CARD-PAL gibt es vier Kartentypen: Personenkarten, Hilfsmittelkarten, Ortskarten und Karten für Fachbegriffe. Bei der Besprechung eines Arbeitsablaufs dokumentieren Kunden und Softwareentwickler gemeinsam auf diesen vorgefertigten Kartenvorlagen Geschäftsobjekte.

Aus Sicht eines Informatikers repräsentieren diese Karten UML-Objekte. Bereits durch die Wahl eines geeigneten Kartentyps können diese Objekte vorklassifiziert werden. Im Wesentlichen entsprechen die Kartentypen den UML-Stereotypen `entity`, `boundary` und `controller`. Bei der Auswahl des Kartentyps ist ein Fragenkatalog behilflich.

2.2 Die Szenarien

CARD-PAL-Szenarien sind kleine, bebilderte Diagramme, in welchen Objekte – repräsentiert durch Karten – positioniert und in Beziehung gesetzt werden. Ein (statisches) Szenario ermöglicht einen spezifischen Blickwinkel auf die Beziehungen der Geschäftsobjekte der Kundenwelt.

Die CARD-PAL-Szenarien begründen sich u.a. auf der allgemeinen (linguistischen) Semantik von Objektbeziehungen [WCH87] und auf kognitive Kategorisierungsprozesse [Eck91]. Damit orientieren sie sich an der universellen Wahrnehmung des Menschen und nicht an der systemtheoretischen Betrachtungsweise eines Softwareentwicklers.

Im Rahmen der Diplomarbeit werden sechs Szenarien definiert und detailliert beschrieben. Diese sind:

¹cards for domain specific terms, persons, auxiliary means and locations

Basisszenario:

zur Detektion von Assoziationen zwischen Objekten.

Teil-Ganzes-Szenario:

zur Beschreibung von Teil-Ganzes-Beziehungen zwischen Objekten.

Organisationsszenario:

zur Beschreibung organisatorischer Strukturen.

Kartographieszenario:

zur Visualisierung von Beziehungen zwischen Orten und der Lokalisierung von Geschäftsobjekten.

Wechselwirkungsszenario:

zur Visualisierung von Kollaborationen und Abhängigkeiten zwischen Objekten.

Karteiszenario:

zur Klassifizierung der Objekte und zur Generalisierung von Klassen.

In den Szenarien werden die UML-Konstrukte Objekt, Klasse, is-a Beziehung, Assoziation, Aggregation und Komposition [OMG07] betrachtet und mit der kartenbasierten Darstellung dem Kunden verständlich näher gebracht. Jedes Szenario wird in ein Objektdiagramm transformiert. Durch Kombination mehrerer Szenarien gelingt es, aus deren Objektdiagrammen ein fast vollständiges UML-Klassendiagramm zu abstrahieren, welches nunmehr dem Entwickler als formales Datenmodell für weitere Prozesse der Softwareentwicklung zur Verfügung steht.

3 Weitere Forschungsdirektiven

Anhand eines erdachten Fallbeispiels aus dem Bibliothekswesen ist das theoretisch entwickelte Konzept einer ersten Erprobung unterzogen worden. Dieser Test kann kein Ersatz für praktische Erfahrungswerte sein. Es ist vorgesehen, CARD-PAL im praxisnahen Umfeld bezüglich seiner Verständlichkeit für UML-Laien zu evaluieren.

Bereits an Hand des kleinen Fallbeispiels dieser Diplomarbeit zeigen sich weitere Entwicklungsperspektiven für CARD-PAL. Mit einem detaillierten Katalog von Transformationsregeln, kann CARD-PAL nicht nur in einem automatisierten Werkzeug umgesetzt werden; es scheint außerdem möglich, weitere UML-Konstrukte wie Attribute oder Methoden aus den Szenarien zu gewinnen.

CARD-PAL ist derart konzipiert, dass es durch projektspezifische Kartentypen und Szenarien erweitert werden kann. Neben der Überführung in UML-Objekt- und Klassendiagramme können zudem andere objektorientierte Diagramme (z.B. Verteilungs-, Kollaborations- oder Kompositionsdiagramme) für die Formalisierung in Betracht gezogen werden. Insbesondere vor dem Hintergrund der automatisierten Transformation einer kundenfreundlichen Darstellung in formale Modelle ist CARD-PAL nicht nur für die Partizipative Analyse sondern auch für die modellbasierte Softwaretechnik bei der Entwicklung kundenindividueller Software von Interesse.

Literatur

- [BC89] Kent Beck and Ward Cunningham. A Laboratory for Teaching Object Oriented Thinking. In *OOPSLA '89: Conference Proceedings on Object-Oriented Programming Systems, Languages and Applications*, pages 1–6, New York, NY, USA, 1989. ACM.
- [DMK98] Tom Dayton, Al McFarland, and Joseph Kramer. *User Interface Design, Bridging the Gap from User Requirements to Design*, chapter Bridging User Needs to Object Oriented GUI Prototype via Task Object Design, pages 15–56. CRC Press LLC, Boca Raton, FL, USA [u.a], 1998.
- [Eck91] Thomas Eckes. *Psychologie der Begriffe: Strukturen des Wissens und Prozesse der Kategorisierung*. Habilitationsschrift (1990), Universität des Saarlandes, 1991.
- [ER05] Ivonne Erfurth and Wilhelm R. Rossak. UPEX: User Participation by EXample. In *ESEC/FSE-13: Proceedings of the 10th European Software Engineering Conference held jointly with 13th ACM SIGSOFT International Symposium on Foundations of Software Engineering*, pages 374–376, New York, NY, USA, 2005. ACM.
- [ER07] Ivonne Erfurth and Wilhelm R. Rossak. A Look at Typical Difficulties in Practical Software Development from the Developer Perspective – A Field Study and a First Solution Proposal with UPEX. In *ECBS '07: Proceedings of the 14th Annual IEEE International Conference and Workshops on the Engineering of Computer-Based Systems*, pages 241–248, Tucson, AZ, USA, März 2007. IEEE.
- [FMR⁺89] Christiane Floyd, Wolf-Michael Mehl, Fanny-Michaela Reisin, Gerhard Schmidt, and Gregor Wolf. Out of Scandinavia: Alternative Approaches to Software Design and System Development. *Human-Computer Interaction*, 4(4):253–350, 1989.
- [Gro95] The Standish Group. Chaos Report. Technical report, The Standish Group, Inc., 1995.
- [Gro01] The Standish Group. Extreme Chaos. Technical report, The Standish Group, Inc., 2001.
- [Laf96] Daniel Lafrenière. CUTA: A Simple, Low-Cost Approach to Task Analysis. *Interactions*, 3(5):35–39, September/Oktober 1996.
- [Mul01] Michael J. Muller. Layered Participatory Analysis: New Developments in the CARD Technique. In *Conference on Human Factors in Computing Systems. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '01, pages 90–97, Seattle, WA, USA, März/April 2001. ACM Special Interest Group on Computer-Human Interaction, ACM, New York, NY, USA.
- [Oes04] Bernd Oestereich. *Objektorientierte Softwareentwicklung Analyse und Design mit der UML 2.0*. Oldenbourg Verlag, 6. edition, 2004.
- [OMG07] OMG. OMG Unified Modeling Language (OMG UML), Infrastructure, V2.1.2. <http://www.omg.org/cgi-bin/apps/doc?formal/07-11-02.pdf>, November 2007.
- [TMDR93] Leslie Gayle Tudór, Michael J. Muller, Tom Dayton, and Robert W. Root. A Participatory Design Technique for High-Level Task Analysis, Critique, and Redesign: The CARD Method. In *The Human Factors and Ergonomics Society 1993 Meeting*, pages 295–299, 1993.
- [WCH87] Morton E. Winston, Roger Chaffin, and Douglas Herrmann. A Taxonomy of Part-Whole Relations. *Cognitive Science*, 11(4):417–444, 1987.

Ein Konzept und Entwicklung eines Prototyps zur Auswertung von Volleyballspielen auf Grundlage der Videoaufzeichnungen

Alexander Neidhart

Hochschule Bremerhaven
Abteilung Informatik
Alexander.Neidhart@gmx.de

Art der Arbeit: Diplomarbeit
Betreuer der Arbeit: Prof. Dr. Günter Matthiessen und
Dipl.-Phys. Ulrich Brosig

Abstract: Eine Video eines Volleyballpunktspiel wird mit dem Softwareprogramm ausgewertet. Durch Wiederholungs- und Pausierungsmöglichkeiten können Schlagrichtungen, die Positionen der Spieler und Bewertungen der Ballberührungen gespeichert werden. Danach können z.B. alle verschlagenen Aufschläge eines Spielers hintereinander angeschaut werden - ohne das Video spulen zu müssen. Aus allen Sätzen können die Schlagrichtungen eines Spielers grafisch übereinander angezeigt werden.

1 Motivation

Leistungsorientierte Volleyballtrainer und -spieler nutzen Auswertungen eines Volleyballpunktspiels. Dies kann handschriftlich während des Spiels geschehen oder mit einem Softwareprogramm.

Mit einem handschriftlichen und sehr detaillierten System habe ich ein 1,5 Stunden dauerndes Spiel in 12 Stunden auf 50 Papierseiten erfasst. Jedoch war die Auswertung schwierig, da die Schlagrichtungen und die Spielerpositionen auf vielen Blättern verteilt waren.

Mit dem Softwareprogramm des Marktführers DataVolley <Da10> können während des Spiels nahezu die gleichen Daten eingegeben werden. Ein einstündiges Spiel ist einer Stunde erfasst. Dadurch können Trainer während des Spiels Auswertungen erhalten. Erst nach dem Spiel werden die Daten mit dem Video verknüpft, so dass z.B. alle verschlagenen Aufschläge eines Spielers hintereinander angezeigt werden, ohne dass sonstige Ballberührungen ablenken. Diese Videoverknüpfung dauert eine weitere Stunde. Insgesamt braucht der Eingebende statt meiner 12 Stunden auf dem Papier 2 Stunden. Zusätzlich sind die Spielzüge mit dem Video verknüpft und dadurch aus der Datenbank ansteuer- und auswertbar.

Allerdings benötigt der Eingebende ein halbes Jahr, um mit DataVolley umgehen zu können <Pr09>.

2 Der Prototyp

Die Volleyballvereine der Bundesliga sind mit DataVolley zufrieden. Es gibt keinen Verbesserungsbedarf <Pr09>.

Das Ziel ist es daher eine Software für nicht professionelle Vereine zu erstellen von der Bezirksliga bis möglichst zur Regionalliga. Die Einarbeitungszeit in das neue Programm soll nur Stunden und keine Monate betragen.

Die Priorität ist nicht das Erfassen eines Live-Spiels wie bei DataVolley, sondern die Datenerfassung erfolgt mit Hilfe von Videomaterial. Dadurch ist das Spielgeschehen während der Eingabe pausierbar, wodurch viel Einarbeitungszeit entfällt.

Mit einem Fragebogens werden die Bedürfnisse von Trainern an einer solchen Software ermittelt. Um nicht erst nach vielen Entwicklungsmonaten feststellen zu müssen, dass die Benutzer nicht mit der Datenerfassung zurecht kommen, wird ein Prototyp erstellt. Mit diesem werden Benutzer getestet, ob sie in gewünschter Zeit das Programm bedienen können und wie lange die Dateneingabe für ein Punktspiel dauert.

Bei der Programmierung werden Java und folgende Werkzeuge eingesetzt: eine Video API zum Abspielen und Schneiden der Videos, Test Driven Development <Li05>, Design Patterns <GHJV96>, eine DVD-Brenn API¹, eine objektorientierte Datenbank, eine Hashfunktion zur automatisierten Erkennung der Videos und Reflection <Su10>, um private Methoden und Property- und VetoChangeListener testen zu können. Dabei wurde darauf geachtet, dass die Lizenzen der Werkzeuge kostenlos oder kostengünstig sind.

Zu jeder Ballberührung werden nicht sofort alle Daten zu einer Ballberührung eingegeben, sondern das Programm verlangt eine Arbeitsteilung. Im Gegensatz zu DataVolley werden zuerst die Positionen der Spieler und die Zeitstempel im Video eingegeben (Abbildung 1). Dieses System ist innovativ, denn kein anderes Volleyballauswertungsprogramm nutzt dieses System. Die Benutzeroberfläche ist dadurch überschaubar, selbsterklärend und die Einarbeitungszeit ist kurz.

¹ Spieler erhalten eine DVD zur Fehlerauswertung, auf der nur sie bei Spielaktionen zu sehen sind. Erfahrungsgemäß langweilt die Spieler das Ansehen von Spielaktionen, an denen sie nicht beteiligt sind.

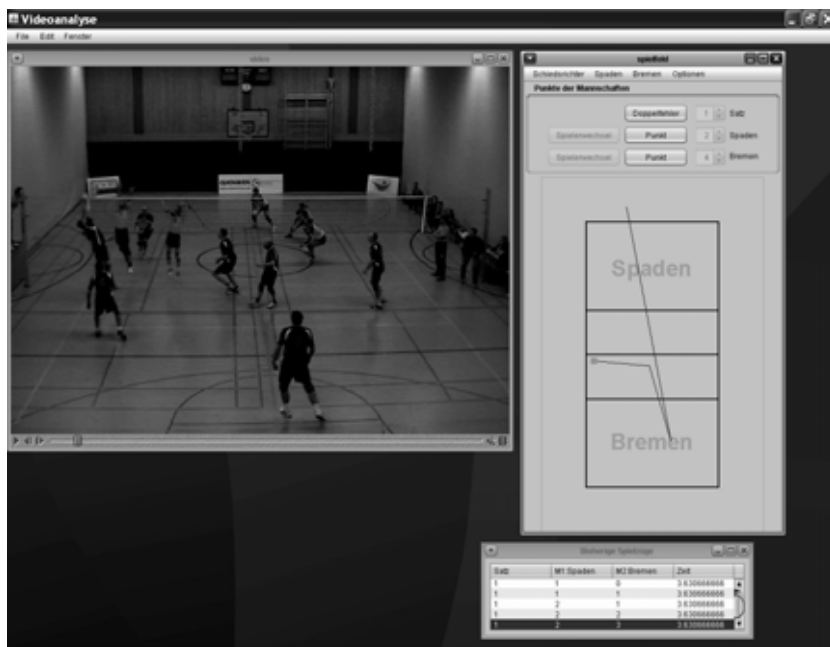


Abbildung 1: Eingabe der Schlagrichtung einer Ballberührung

Nach Eingabe aller Spielzüge wird die zu bewertende Ballberührung in einer Endlosschleife angezeigt. Der Auswerter kann sich den Erfolg einer Berührung anschauen und sie bewerten (Abbildung 2) – ohne zurückspulen zu müssen, wenn das Geschehen zu schnell ablief.

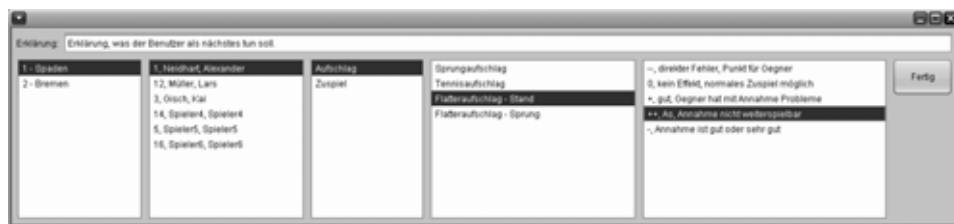


Abbildung 2: Eingabe der Bewertung einer Ballberührung

Literaturverzeichnis

- [Da10] DataProject: Volleyball Auswertungssoftware DataVolley, 09.01.2010,
http://www.dataproject.com/datalist_listino.aspx
- [GHJV96] Gamma, E; Helm, R; Johnson, R.; Vlissides, J.: Entwurfsmuster, Addison-Wesley GmbH, Bonn, 1996, S. 4 – 7
- [Li05] Link, J.: Softwaretests mit JUNIT, 2.überarbeitete Auflage, dpunkt.verlag GmbH, Heidelberg, 2005, S. 33 – 62
- [Su10] Sun Microsystems: Java Reflection API Tutorial, 09.01.2010,
<http://java.sun.com/docs/books/tutorial/reflect/index.html>
- [Pr09] Daniel Prade, der Scout der 1. Bundesliga Volleyball Damenmannschaft „NA Hamburg“, gab mir diese Informationen am 17.02.09

Virtuelle interaktive Prospekte mit Hilfe der Nintendo Wiimote

Timm Richlick

Fachhochschule Aachen
Fachbereich Elektro- & Informationstechnik
richlick@fh-aachen.de

Art der Arbeit:

Bachelorarbeit

Betreuer/in der Arbeit:

Prof. Dr. Thomas Ritz

Abstract: Trotz fortschrittlicher Informationsmedien, allen voran das Internet, spielen traditionelle Printmedien in unserem Alltag nach wie vor eine bedeutende Rolle. Diese Arbeit beschreibt die Entwicklung eines Prototypen für ein Informations-Terminal, das die Vorteile eines Printmediums mit den Vorteilen eines Internet-PCs verbinden soll. Hierzu wird mittels der Infrarotkamera eines Nintendo-Wiimote-Spielecontrollers, eines handelsüblichen Beamers und einer beweglichen Projektionsfläche eine Art virtuelles Prospekt entwickelt, das die Haptik und Benutzerfreundlichkeit eines Papierprospektes bieten soll. Das virtuelle Prospekt lässt sich über intuitive Bewegungsgesten steuern und ist in der Lage dynamische Inhalte per HTML darzustellen.

1 Einleitung

Bücher, Prospekte, Zeitungen und Zeitschriften sind seit über 5000 Jahren fester Bestandteil unserer Kultur. Es bedarf keiner technischen Einweisung, um ein Printmedium zu bedienen. Demgegenüber bietet das Internet als modernes Informationsmedium eine enorme Flexibilität, beispielsweise durch dynamische und multimediale Inhalte oder durch einen hohen Grad an Interaktionsmöglichkeiten. Bezüglich der Bedienung bieten beide Medien bestimmte Vor- und Nachteile.

Ziel dieser Arbeit ist es einen Prototyp für ein stationäres Unterhaltungs-Terminal zu entwickeln, das sich ähnlich wie ein gedrucktes Prospekt handhaben lässt, zusätzlich aber die Vorteile eines elektronischen Mediums bietet. Dabei sollen alternative und intuitive Navigationsmöglichkeiten durch das „virtuelle Prospekt“ eruiert sowie beispielhaft in den Prototyp implementiert werden. Jegliche Interaktion zwischen Benutzer und System soll durch gewohnte, einfache Bewegungsgesten geschehen und somit auch für weniger technikaffine Menschen zugänglich und leicht verständlich sein. Auf Tasten, Knöpfe o.ä. wird dabei vollständig verzichtet.

Als Anregung hierzu dienen die von Johnny Lee veröffentlichten Internetvideos über die Verwendung des Wiimote-Controllers als PC-Eingabegerät [Jo07][Jo08][Co09]. Ursprünglich wurde der Controller zur Bedienung einer Spielekonsole entwickelt. Mit Hilfe des Controllers ist es möglich die Position eines Objektes im zweidimensionalen Raum zu verfolgen und dadurch beispielsweise bestimmte Bewegungsgesten zu erkennen.

2 Board-Tracking und Projektion

Zur Realisierung der Zielsetzung wird zunächst ein kabelloses, flaches Plastikgehäuse gebaut, auf dem insgesamt 8 Infrarot-LEDs inklusive Stromversorgung montiert werden. Dieses „Board“ dient als Display für das virtuelle Prospekt. Der Benutzer kann das Board in die Hand nehmen, die Prospektseiten betrachten und mittels Bewegungsgesten durch die Prospektseiten blättern.

Um die Position des Boards zu erkennen, wird die Infrarotkamera der Wiimote verwendet. Diese ist in der Lage die Infrarot-LEDs zu orten und die Positionswerte per Bluetooth an den Rechner zu übermitteln (hierzu wird eine von Hand erweiterte Version der WiiYourself!-Bibliothek verwendet [Gl09]). Zur Auswertung der Daten dient ein C++-Programm, das mittels OpenGL ([Kr10]) ein Bild berechnet, das wiederum an den Beamer weitergegeben wird. Der Beamer projiziert dann die Prospektseite genau an die Stelle, an der sich momentan das Board im Raum befindet. Als Informationsquelle dienen HTML-Webseiten, die von dem System eingelesen und dann auf das Board projiziert werden.

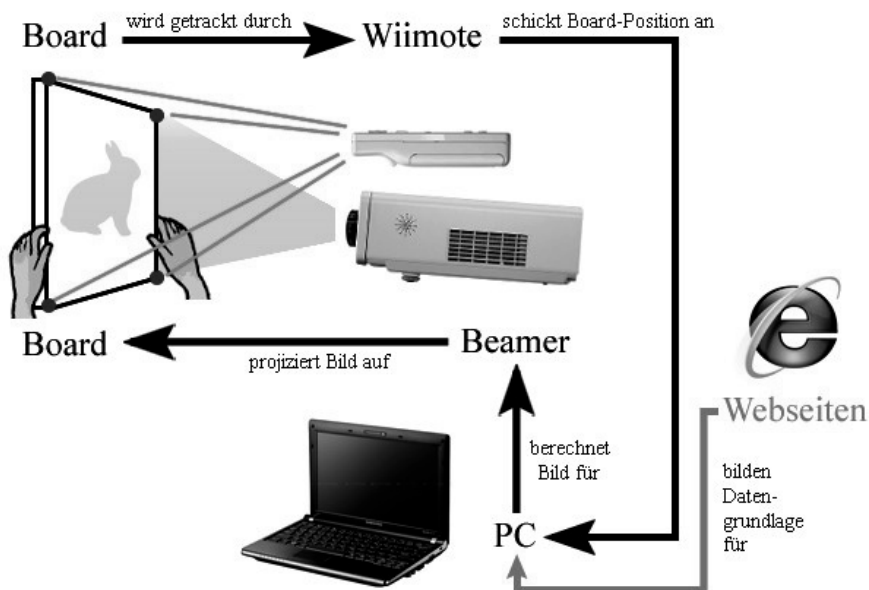


Abbildung 1: Systemaufbau

3 Navigation durch Bewegungsgesten

Auf jeder Seite des Boards sind vier Infrarot-LEDs montiert; somit lassen sich sowohl Vorder- als auch Rückseite als Projektionsfläche nutzen. Zur Navigation durch mehrere Prospektseiten kann das Board wie ein Printmedium umgeblättert werden, allerdings in vier statt zwei verschiedene Richtungen (oben, unten, rechts und links). Somit wird eine zweidimensionale Navigationsstruktur ermöglicht. Zur Erkennung der Bewegungsgesten werden mehrere, eigens entwickelte Algorithmen eingesetzt. Diese approximieren die dreidimensionale Bewegungsgeste durch die gegebenen, zweidimensionalen LED-Koordinaten (z.B. anhand der perspektivischen Verzerrung oder der Geschwindigkeit der Punktbewegungen relativ zueinander).

Durch leichtes Hoch- und Herunterbewegen des Boards kann der Benutzer, wie bei einer Webseite, durch überlange Prospektseiten scrollen. Zusätzlich verfügt das System über einen „BigPage-Modus“. Hier kann der Benutzer mit Hilfe des Boards eine Art Landkarte absuchen und durch Navigationspunkte zu anderen Seiten springen (ähnlich wie Hyperlinks auf einer Webseite).



Abbildung 2: Fotos des fertigen Systems

4 Ergebnisse

Der folgende Link zeigt ein Video des fertigen Prototyps:

http://www.youtube.com/watch?v=OFvfh5_PdAI

Das System wurde durch einige Probanden getestet und von diesen mehrheitlich als unterhaltsam und leicht verständlich eingestuft. Die Ergebnisse dieser Arbeit sprechen dafür, dass bestimmte Teilbereiche, darunter die Steuerung technischer Systeme mittels Bewegungsgesten und die Kombination von Technik mit gewohnten Informationsquellen, in Zukunft durchaus Aussicht auf Erfolg haben können.

Literaturverzeichnis

- [Jo07] Lee, J. C.: Foldable Displays (tracked with the Wiimote). Online verfügbar unter http://www.youtube.com/watch?v=nhSR_6-Y5Kg, zuletzt aktualisiert am 27.09.2007, zuletzt geprüft am 28.12.2009.
- [Jo08] Lee, J. C.: Wii-Projects. Online verfügbar unter <http://johnnylee.net/projects/wii/>, zuletzt aktualisiert am 01.06.2008, zuletzt geprüft am 11.01.2010.
- [Co09] MSDN Blog Coding4Fun: Managed Library for Nintendo's Wiimote. Online verfügbar unter <http://blogs.msdn.com/coding4fun/archive/2007/03/14/1879033.aspx>, zuletzt aktualisiert am 19.01.2009, zuletzt geprüft am 11.01.2010.
- [Gl09] gl.tter: WiiYourself!. Online verfügbar unter <http://wiiyourself.gl.tter.org/>, zuletzt aktualisiert am 27.07.2009, zuletzt geprüft am 28.12.2009.
- [Kr10] Khronos Group: OpenGL. Online verfügbar unter <http://www.opengl.org/>, zuletzt aktualisiert am 08.01.2010, zuletzt geprüft am 11.01.2010.

Integration von syntaxbasierter Nutzerunterstützung in Sketching-Editoren

Christian Rutetzki

christian.rutetzki@unibw.de

Universität der Bundeswehr München

Abstract: Sketching-Editoren erlauben das freihändige Zeichnen von Diagrammen mit einem Stift als Eingabegerät. Dank der anschließenden Erkennung kann das Diagramm analysiert und weiter verarbeitet werden, wie es die Nutzer von herkömmlichen Diagrammeditoren gewohnt sind. Diese Arbeit geht nun einen Schritt weiter und integriert *Nutzerunterstützung* in Sketching-Editoren, d.h. der Nutzer wird über mögliche Fehler nicht nur informiert, sondern bekommt konkrete Vorschläge, wie das Problem gelöst werden kann. So können etwa fehlende Komponenten berechnet und in die Skizze integriert werden. Als Beispielanwendung wurde ein Sketching-Editor für Geschäftsprozessmodelle entwickelt, der neuartige Unterstützungsfunktionen bereit stellt.

1 Einleitung

Aufgrund der Freiheit beim Zeichnen erfreuen sich Sketching-Editoren für Diagramme zunehmender Beliebtheit. Dabei wird ein Stift als Eingabegerät verwendet, was für eine sehr natürliche Interaktion sorgt. Da bei der Erkennung der freihändig gezeichneten Diagramme Komponenten nicht oder falsch erkannt werden können, ist eine entsprechende Hilfestellung erstrebenswert. So existieren zwar Sketching-Editoren, die dem Nutzer Fehler in Diagrammen deutlich machen, allerdings bieten sie bisher keine Korrekturvorschläge und unterstützen den Nutzer auch generell nicht bei der Erstellung korrekter Diagramme.

Der vorgestellte Ansatz zeigt, wie Sketching-Editoren um eine Funktionalität zur Korrektur von Diagrammen ergänzt werden können. Ein erstellter Sketching-Editor für eine Teilmenge strukturierter Geschäftsprozessmodelle [Obj09] zeigt die Realisierbarkeit.

Diese Arbeit ist im Rahmen einer Bachelorarbeit an der Universität der Bundeswehr München entstanden [Rut10].

2 Der erweiterte Sketching-Prozess

In einem herkömmlichen Sketching-Editor skizziert der Nutzer zunächst mit mehreren Strichen sein Diagramm. Im Zuge der Erkennung gehen daraus konkrete Diagrammkom-

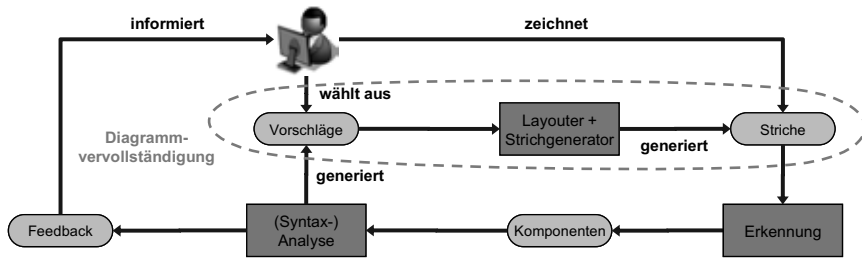


Abbildung 1: Konzept der Integration von Nutzerunterstützung in Sketching-Editoren

ponenten wie etwa Pfeile hervor. Die Analyse überprüft dann, ob die erkannten Diagrammkomponenten zusammen ein syntaktisch korrektes Diagramm ergeben.

Das Konzept (siehe Abb. 1) des hier vorgestellten Ansatzes sieht für das Analysemodul außerdem vor, bei Bedarf Verbesserungsvorschläge zu generieren. Der Nutzer kann im Anschluss zwischen ihnen wählen und sie im Diagramm in entsprechende Striche umsetzen lassen.

Ein Vorschlag beschreibt im Wesentlichen, welche Komponenten hinzugefügt werden müssen, um das Diagramm zu vervollständigen. Der Layouter weist diesen Komponenten dann Positionen im Diagramm zu und der Strichgenerator erzeugt die entsprechenden Striche.

Von einem bestehenden Sketching-Editor ausgehend, muss also die Analyse ergänzt, ein Layouter und Strokegenerator erstellt sowie eine Auswahlmöglichkeit für den Nutzer geschaffen werden. Für die konkrete Umsetzung wurde DSketch [Bri09, BM08] als Basis verwendet. DSketch ermöglicht das Erstellen von Sketching-Editoren auf Grundlage einer Sprachspezifikation. Das zugehörige Framework wurde für die Realisierung des Konzepts um eine Bibliothek erweitert, die aus dem Ansatz [MM09, MMM08] hervorgeht. Diese ergänzt die Analyse um die Fähigkeit, Vervollständigungen zu generieren. Dies erfordert keine zusätzliche Spezifikation, sondern die bestehenden Syntaxregeln werden ausgenutzt. Eine Erweiterung der Benutzerschnittstelle erlaubt dem Nutzer eines DSketch-Editors schließlich, die hinzugewonnene Funktionalität auch zu bedienen. Letztendlich ist so eine Erweiterung des DSketch-Frameworks entstanden, welche das generische Erstellen von Sketching-Editoren mit einer Diagrammvervollständigungsfunktionalität ermöglicht.

3 Sketching-Editor für Geschäftsprozessmodelle als Beispiel

An dieser Stelle wird die neu hinzugewonnene Funktionalität an einem erstellten Sketching-Editor für eine Teilmenge strukturierter Geschäftsprozessmodelle (ohne *Pools* und *Intermediate Events*) vorgestellt.

Über eine Toolbar kann der Nutzer Vorschläge für die Vervollständigung von Diagrammen generieren lassen. Diese werden in Form einer Vorschau auf der Zeichenfläche ein-

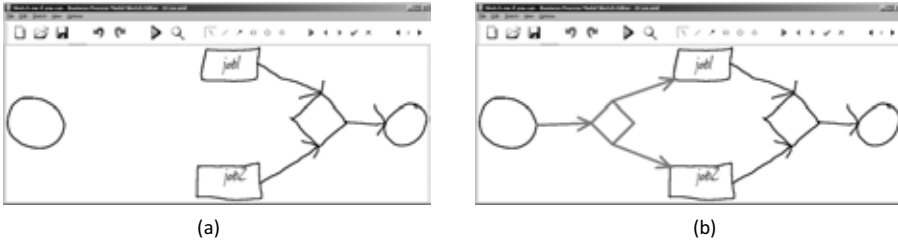


Abbildung 2: Vervollständigung: (a) Unvollständiges Diagramm, (b) Vorschau der Vervollständigung

geblendet. Der Nutzer kann zwischen allen Vorschlägen „blättern“, einen davon direkt übernehmen oder mit dem Stift nachzeichnen. Anders als im Falle des Nachzeichnens, erzeugt das direkte Übernehmen allerdings perfekte Formen und schafft somit ein etwas unnatürliches Aussehen.

Der Nutzer kann vor der Berechnung der Vervollständigung die Anzahl hinzuzufügender Komponenten festlegen. Abbildung 2 zeigt ein unvollständiges Diagramm, sowie die Vorschau der berechneten Vervollständigung, bei der ein Gateway sowie mehrere Pfeile hinzugefügt werden.

4 Zusammenfassung und Ausblick

Hier wurde ein Konzept vorgestellt, mit dem bestehende Sketching-Editoren um eine syntax-basierte Nutzerunterstützung erweitert werden können. Unvollständige Diagramme können komplettiert werden, und der Nutzer kann sich Beispieldiagramme (also Vervollständigungen des leeren Diagramms) generieren lassen.

Nicht zuletzt kann das Konzept des Sketchings selber verbessert werden, da der Nutzer auf diese Weise auch falsch oder nicht erkannte Komponenten identifizieren und daraufhin ersetzen kann.

Die Umsetzung des Konzeptes konnte an DSketch demonstriert werden. Es ist also tatsächlich möglich, Sketching-Editoren um Diagrammvervollständigung zu erweitern. Ein entsprechender Sketching-Editor für eine Teilmenge strukturierter Geschäftsprozessmodelle ist entstanden, der auf der Homepage www.unibw.de/inf2/DiaGen/assistance/sketching neben einer Screencast-Serie zur Verfügung steht.

Als Folgearbeiten sollten Editoren für weitere Sprachen realisiert sowie mit Nutzerstudien evaluiert werden.

Literatur

- [BM08] Florian Brieler und Mark Minas. Recognition and processing of hand-drawn diagrams using syntactic and semantic analysis. In *AVI '08: Proceedings of the working conference on Advanced visual interfaces*, Seiten 181–188, New York, NY, USA, 2008. ACM.
- [Bri09] Florian Brieler. *A Generic Approach to the Recognition and Analysis of Sketched Diagrams Using Context Information*. Dissertation, Universität der Bundeswehr München, 2009.
- [MM09] Steffen Mazanek und Mark Minas. Business Process Models as a Showcase for Syntax-based Assistance in Diagram Editors. In Andy Schürr und Bran Selic, Hrsg., *Model Driven Engineering Languages and Systems, Proc. ACM/IEEE 12th International Conference, MODELS 2009, Denver, CO, USA, October 2009*, Jgg. 5795 of *Lecture Notes in Computer Science*, Seiten 322–336. Springer-Verlag, 2009.
- [MMM08] Steffen Mazanek, Sonja Maier und Mark Minas. Auto-completion for diagram editors based on graph grammars. In *VLHCC '08: Proceedings of the 2008 IEEE Symposium on Visual Languages and Human-Centric Computing*, Seiten 242–245, Washington, DC, USA, 2008. IEEE Computer Society.
- [Obj09] Object Management Group. Business Process Model and Notation (BPMN). <http://www.omg.org/spec/BPMN>, 2009.
- [Rut10] Christian Rutetzki. *Integration von syntaxbasierter Nutzerunterstützung in Sketching-Editoren*. Bachelorarbeit, Universität der Bundeswehr München, 2010.

I'm Home

Smartphone-enabled Gestural Interaction with Multi-Modal Smart-Home Systems

Tilo Westermann

mail@tilowestermann.eu

Abstract: In this work, the problem of an inadequate usability for controlling home appliances is addressed. Previous approaches seemed promising, but may not measure up to users' expectations of a user-friendly control modality. This thesis investigates the use of a smartphone-based gestural interface for smart-home control in order to cope with this problem. A series of studies for the creation of a set of applicable gestures is introduced. Furthermore, details on the implementation of a gesture recognition system for mobile devices using the Fast Dynamic Time Warping (FastDTW) algorithm are presented. The results show a recognition rate of 96.87% for an experienced user, which sounds promising for future applications.

1 Introduction

Today's home environments are equipped with an increasing number of devices. However, interaction techniques, commonly occurring in the form of remote controls, remain the same, hence a clutter on the coffee table is the inevitable. An existing approach to this problem is the combination of multiple remote controls to a single 'universal' remote control. Although eliminating the issue of a multitude of remote controls, these devices confront the user with a plethora of buttons and modes.

There are also approaches to the design of smart environments with alternative input techniques. Due to their requirements, only a few are eligible for the use in a smart-home environment. Voice-based interfaces seem to be well suited, but are still subject to high error rates, which complicates the practical use. Besides voice, gestures seem to be a natural way of interaction for human beings. Research on the transfer of gestures into this new domain has shown promising approaches, as provided by [PNKC08, SPHB08]. Some of these approaches utilize mobile devices equipped with motion sensors. Moving mobile devices into the context of stationary home environments seems worthwhile, as mobile devices have become a pervasive part of everyday life [SGV05].

This thesis investigates the use of a smartphone-based gestural interface for smart-home control to encounter the emerging problem of an abundance of remote controls that are not necessarily user-friendly, or are replaced through a universal remote control.

2 Method

To obtain a first impression of whether or not a gesture-based control modality has any appeal to possible end users, a survey concerning this question was conducted. Seventeen students from different disciplines were asked for their attitude towards a gestural interface and which home appliances they would like to control. In order to be able to examine the applicability of gestural interaction, a method for the development of gestures is introduced, where users were involved in the design process. Following the promising approach of Wobbrock et al. [WMW09], a first study considered the search for a gesture vocabulary. In this study, 18 users were asked to conceive gestures for predetermined tasks (e.g. ‘turn on the TV’). Based on the results of this study, 26 out of 174 gestures were selected to compose a first gesture vocabulary. In the following study, 22 users who had not previously participated were asked to reassign commands to given gestures, which were shown as short video clips. As a result, the gesture vocabulary was further reduced. A third study, inspired by the work of Nielsen et al. [NSMG04], should examine the memorability of the elected gestures.

To find a suitable smartphone, a set of requirements was established. This set included the need for motion sensors, an appropriate input modality, connectivity capabilities and a certain degree of popularity. The *Apple iPhone* was chosen, as it fulfilled the criteria. On the assumption that remote control-like functions may be well suited for specific tasks, e.g. selecting a list entry, the application was designed to allow for both, gestural interaction and control via the graphical user interface (GUI). In addition, gestures are trained by the user, allowing for custom gestures. As the probability of incidental recognitions for continuous gesture recognition seemed too high, recognition was triggered by the user by a button press.

For the recognition of gestures Dynamic Time Warping (DTW) [SC78] has been used. DTW was introduced in the field of speech recognition and is able to find the optimal alignment of two given data sequences, where one of them may be stretched or shrunk along the time axis. However, since working with a mobile device with limited processing power and the amount of data could quickly take on a large scale, this approach came up against its limits. To overcome this problem, an improvement to DTW was applied. Fast Dynamic Time Warping (FastDTW) [SC07] amends the existing algorithm with the addition of data abstraction and constraints. While DTW has a quadratic time and space complexity, FastDTW has linear time and space complexity. In addition, the *iPhone 3G* was replaced by an *iPhone 3GS*, which provided a further decrease of processing time.

A final usability study should evaluate the approach of this thesis. Away from the technical side, it is of prime importance whether potential users of such a system would ever use this modality and how they would cope with it. The study was conducted in a fully furnished living room, where electronic devices are connected to a network. 27 participants evaluated the usability of the implemented approach by comparing speech-based input with gestural control and a combination of both modalities.

3 Results

The initial survey indicated that users are generally interested in controlling home appliances via gestures. The following set of studies concerned with finding appropriate gestures for smart-home control yielded a vocabulary of 13 gestures. Figure 1 shows three gestures of the final vocabulary: Figure 1a is assigned to three functions: ‘blinds down’, ‘volume down (radio, TV)’ and ‘dim light’. Functions assigned to Figure 1b are ‘next channel (radio, TV)’ and ‘next message (answering machine)’. Figure 1c shows a gesture that is used for turning the light on and off.

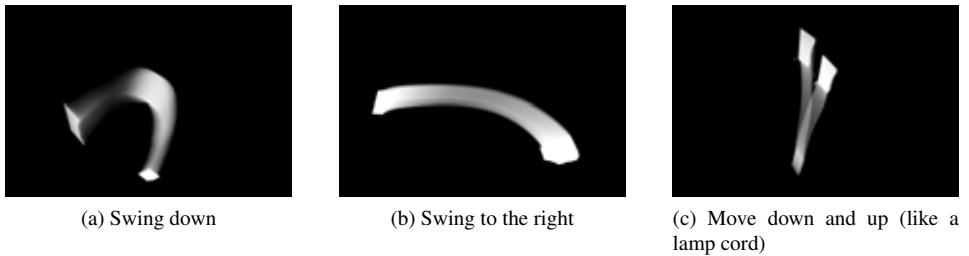


Figure 1: Gestures, represented by long exposure photos

As completion of the gesture recognition implementation on the mobile device, a verification test was run. Seven gestures, which were later used in the usability study, were each performed 50 times. This test delivered a recognition rate of 96.87% (2.28% not recognized, 0.85% incorrectly recognized).

The results of the usability study were divided into gesture recognition rate and analysis of questionnaires. The recognition rate in this study is significantly lower than in the previous test. In total, 18.82% of the gestures could apparently not be assigned to any of the previously trained gestures and were not considered for the evaluation of the recognition rate. While recognition rates varied widely among participants, the average rate of correctly recognized gestures is 64.84% (21.83% incorrectly recognized, 13.33% not recognized). As realized in hindsight, there was a bug in the gesture recognition algorithm, which lead to suboptimal results. Considering the false and not recognized gestures, a remarkable number may be explained by this programming error. As indicated by the participants' comments after completion of the study, the reason for the high number of unrecognizable gestures may be that the training period was too short to remember all the gestures. While the participants went through a relatively short training period, the test described above was conducted by a user who was able to rehearse the gestures over a longer period of time (this test was conducted after fixing the programming error). Another criticism leveled at the smartphone-based interaction was that the error handling was not sufficient - this will be remedied by a more detailed feedback in the future.

4 Conclusion

In this thesis, the problem of an inadequate usability for controlling home appliances is addressed. The presented approach makes use of a commercially available smartphone with built-in motion sensors to allow for gestural and GUI-based interaction. This approach seems worthwhile, as the mobile phone has become a pervasive part of everyday life. A series of studies for the creation of a set of applicable gestures is introduced and details on the implementation of a gesture recognition system for mobile devices are presented. As an example platform, the Apple iPhone was chosen for evaluating the implementation in a final usability study. The results from this study show that gestures may be used for controlling devices in a smart-home environment, but there are still points for improvement, such as error handling and feedback from the system.

References

- [NSMG04] M. Nielsen, M. Storrang, T.B. Moeslund, and E. Granum. A procedure for developing intuitive and ergonomic gesture interfaces for HCI. *Lecture notes in computer science*, pages 409–420, 2004.
- [PNKC08] J.W. Park, Y. Nam, D.H. Kim, and W.D. Cho. SWINGREMOCON: A GESTURE-BASED REMOTE CONTROLLER FOR HOME APPLIANCES. *International Conference Interfaces and Human Computer Interaction*, pages 268–271, 2008.
- [SC78] H. Sakoe and S. Chiba. Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 26(1):43–49, 1978.
- [SC07] Stan Salvador and Philip Chan. Toward accurate dynamic time warping in linear time and space. *Intell. Data Anal.*, 11(5):561–580, 2007.
- [SGV05] K. Stordahl, I. G. Gjerde, and R. Venturin. Long-Term Forecasts for the Mobile Market in Western Europe. In *Proceedings of 16th Regional European ITS Conference*, pages 76–77, 2005.
- [SPHB08] Thomas Schlömer, Benjamin Poppinga, Niels Henze, and Susanne Boll. Gesture recognition with a Wii controller. In *Proceedings of the 2nd international conference on Tangible and embedded interaction - TEI '08*, page 11, New York, New York, USA, 2008. ACM Press.
- [WMW09] J.O. Wobbrock, M.R. Morris, and A.D. Wilson. User-defined gestures for surface computing. *Proceedings of the 27th international conference on Human factors in computing systems*, pages 1083–1092, 2009.

Synthese von Subspace Clustering und Subgraph Mining zur Analyse von Graphen mit Knotenbeschriftungen

Brigitte Boden

brigitte.boden@rwth-aachen.de

RWTH Aachen University

Lehrstuhl für Informatik 9, Prof. Dr. rer. nat. Thomas Seidl

Art der Arbeit: Diplomarbeit

Betreuer der Arbeit: Dipl.-Inform. Stephan Günnemann

Abstract: In dieser Diplomarbeit wurde ein neuer Clusteringansatz entwickelt, der nicht nur die Attribute der Objekte, sondern auch Informationen über Verbindungen zwischen den Objekten berücksichtigt. Diese Informationen können als Graph mit beschrifteten Knoten dargestellt werden. Die Knoten innerhalb eines Clusters sollen im Graphen dicht verbunden sein und sich in einer Teilmenge der Attribute ähnlich sein. Zum Finden dieser Cluster werden Ansätze aus dem Subspace Clustering und dem Subgraph Mining kombiniert.

1 Einleitung

Data-Mining-Verfahren dienen dazu, aus gespeicherten Datenmengen neues Wissen zu extrahieren. Clustering ist ein spezielles Data-Mining-Verfahren, das zur Gruppierung von Objekten dient. Dabei werden die Objekte in sogenannte Cluster eingeteilt, so dass Objekte mit ähnlichen Eigenschaften im selben Cluster und Objekte mit unähnlichen Eigenschaften in unterschiedlichen Clustern liegen. In hochdimensionalen Daten ist es jedoch unwahrscheinlich, dass eine Gruppe von Knoten sich in *allen* Eigenschaften ähnlich ist. Aus diesem Grund wurde das Subspace Clustering entwickelt. Dieses gruppiert die Objekte so, dass die Objekte innerhalb eines Clusters sich in einer *Teilmenge* der Eigenschaften ähnlich sind.

Bei vielen Anwendungen sind neben den Eigenschaften der Objekte auch Informationen über die Beziehungen zwischen den Objekten bekannt. Diese Informationen können als Graph dargestellt werden. Beide Arten von Informationen können nützlich sein, um Gruppen ähnlicher Objekte zu finden. Die Objekte innerhalb einer Gruppe sollen einerseits *ähnliche Attribute* haben und andererseits im Graphen *dicht verbunden* sein. Für viele Anwendungen ist es außerdem sinnvoll, die Überlappung von Clustern zu erlauben.

Verwandte Arbeiten. Es existieren bereits einige Ansätze, die sowohl Verbindungen zwischen den Objekten als auch Attribute der Objekte betrachten. Bei einigen Ansätzen wie ComTector [DWP⁺07], Research Community Mining [ITM06] und K-groups [KMS03] werden die Attribute nur als Post-Processing-Schritt betrachtet. Ansätze wie Co-Clustering

[HZZL02], Spectral Clustering (z.B. [STM07], [NAJ04]) und Attributed Graph Clustering [ZCY09] verwenden kombinierte Distanzfunktionen, die sowohl die Attribute als auch die Verbindungen zwischen den Objekten berücksichtigen. Die Einteilung der Objekte in die Cluster erfolgt bei diesen Ansätzen durch einen Fullspace-Clustering-Algorithmus (z.B. k-Medoids [RK90] in [ZCY09]). Subspace Clustering ist somit nicht möglich und die Überlappung von Clustern ist bei keinem der Ansätze erlaubt. Der Algorithmus Co-PaM [MCRE09] findet auch Cluster von Knoten, die sich in einer Teilmenge der Attribute ähnlich sind und im Graphen dicht verbunden sind. Auch die Überlappung von Clustern ist erlaubt. Ein Nachteil dieser Methode ist, dass das Ergebnis redundante Cluster enthält. Dadurch wird oft eine sehr große Zahl von Clustern ausgegeben, die sich untereinander stark ähneln.

2 Das neue Clusteringmodell

Die gesuchten Cluster werden anhand von Abbildung 1 veranschaulicht. Zum Vergleich sind ebenfalls die Knotenmengen markiert, die durch reines Subspace Clustering oder Subgraph Mining gefunden werden könnten. Die blau hinterlegten Knoten zeigen das gewünschte Ergebnis. Diese Knoten sind durch viele Kanten verbunden und sind sich sogar in allen Attributen ähnlich. Die übrigen Knoten sollen nicht zu dieser Gruppe hinzugefügt werden, da sie die Dichte oder die Zahl der gemeinsamen Attribute verkleinern würden. Der Benutzer sollte dem Algorithmus vorgeben können, wie wichtig ihm z.B. die Größe oder die Dichte einer Gruppe ist. Aufgrund dieser Informationen sollten dann diejenigen Knotengruppen ausgegeben werden, die für die jeweilige Anwendung die interessantesten sind.

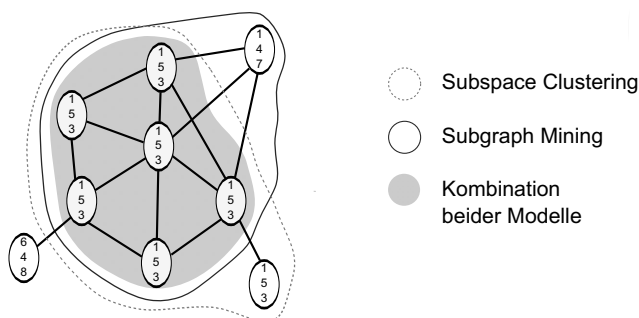


Abbildung 1: Wunschergebnis

Bei der Beschreibung des Modells wird von einem Graphen $G = (V, E)$ als Eingabe ausgegangen, wobei jeder Knoten $v \in V$ ein d -dimensionaler Vektor $(v[1], \dots, v[d])$ mit numerischen Werten $v[i] \in \mathbb{R}$ ist. $D = \{d_1, \dots, d_d\}$ bezeichnet die Menge der Dimensionen. Die Knoten eines Clusters sollen in einem Teilraum von D einen Subspace Cluster

bilden und gleichzeitig einen dichten Subgraphen von G darstellen.

Die Subspace Cluster werden hier analog zu *DOC* [PJAM02] definiert. Dabei dürfen sich die Knoten innerhalb eines Clusters in den relevanten Dimensionen höchstens um einen vorgegebenen Wert unterscheiden. Die dichten Subgraphen werden durch die γ -Quasi-Clique (vgl. z.B. [LW08]) definiert. Die neue Clusterdefinition kombiniert die Definitionen von Subspace Cluster und Quasi-Clique wie folgt:

Definition 1 (Cluster):

Ein Cluster $C = (O, S)$ besteht aus einer Knotenmenge $O \subseteq V$ mit $|O| \geq n_{min}$ und einem Teilraum $S \subseteq D$ mit $|S| \geq s_{min}$, die folgende Bedingungen erfüllen:

- O ist im Graphen zusammenhängend
- O ist eine γ_{min} -Quasi-Clique
- $C = (O, S)$ ist ein Subspace Cluster im Teilraum S

Dabei bezeichnet der Parameter n_{min} die minimale Knotenzahl eines Clusters und γ_{min} die minimale Dichte einer Quasi-Clique O . Der Parameter s_{min} gibt vor, wie groß der Teilraum eines Clusters mindestens sein muss.

Durch die erlaubte Überlappung der Cluster ist es möglich, dass sehr viele gültige Cluster gefunden werden, die sich teilweise nur geringfügig unterscheiden. Daher muss eine Auswahl von möglichst interessanten Clustern für die Ausgabe getroffen werden, so dass die ausgegebene Clustermenge eine übersichtliche Größe hat und keine redundanten Cluster enthält. Um die Interessantheit von Clustern zu bewerten, wurde eine Qualitätsfunktion $Q(C)$ für Cluster entwickelt. Diese Funktion stellt einen Kompromiss zwischen Knotenzahl, Dichte und Dimensionalität eines Clusters dar. Die Gewichtung der einzelnen Komponenten kann dabei durch den Benutzer vorgegeben werden.

Das Redundanzmodell legt fest, welche Cluster als redundant gelten und daher nicht ausgegeben werden. Ein Cluster C heißt redundant bezüglich eines anderen Clusters C' , falls C' eine höhere Interessantheit besitzt und einen bestimmten Anteil der Knoten und Dimensionen von C abdeckt. In Definition 2 wird diese Redundanz formal definiert.

Definition 2 (Redundanz):

Gegeben seien Parameter $s \in [0, 1]$ und $r \in [0, 1]$. Ein Cluster $C = (O, S)$ ist redundant bezüglich eines Clusters $C' = (O', S')$, falls $Q(C) < Q(C') \wedge \frac{|O \cap O'|}{|O|} \geq r \wedge \frac{|S \cap S'|}{|S|} \geq s$.

Durch die Parameter r und s kann die Anzahl und der Grad der Überlappung der ausgegebenen Cluster beeinflusst werden. Basierend auf dieser Definition kann nun eine Auswahl von Clustern getroffen werden, die keine untereinander redundanten Cluster enthält.

Weitere Ergebnisse. Im Rahmen der Diplomarbeit wurde gezeigt, dass das Clustering nach dem neuen Modell ein NP-schweres Problem ist. Es wurden ein exakter Algorithmus und ein Approximationsalgorithmus für das Clustering entwickelt. Die Effektivität und Effizienz dieser Algorithmen wurden in verschiedenen Experimenten untersucht. Die Experimente belegen, dass das entwickelte Clusteringmodell eine gute Kombination von Subspace Clustering und Subgraph Mining ermöglicht.

Literatur

- [DWP⁺07] N. Du, B. Wu, X. Pei, B. Wang und L. Xu. Community detection in large-scale social networks. In *WebKDD/SNA-KDD '07: Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis*, Seiten 16–25, 2007.
- [HZZL02] D. Hanisch, A. Zien, R. Zimmer und T. Lengauer. Co-clustering of biological networks and gene expression data. In *Proceedings of the Tenth International Conference on Intelligent Systems for Molecular Biology*, Jgg. 18, Seiten 145–154, 2002.
- [ITM06] R. Ichise, H. Takeda und T. Muraki. Research community mining with topic identification. In *10th International Conference on Information Visualisation (IV 2006)*, Seiten 276–281, 2006.
- [KMS03] J. Kubica, A. Moore und J. Schneider. K-groups: Tractable group detection on large link data sets. In *Proceedings of the 3rd IEEE International Conference on Data Mining (ICDM 2003)*, Seiten 573–576, 2003.
- [LW08] G. Liu und L. Wong. Effective Pruning Techniques for Mining Quasi-Cliques. In *Machine Learning and Knowledge Discovery in Databases, European Conference, ECML/PKDD 2008*, Seiten 33–49, 2008.
- [MCRE09] F. Moser, R. Colak, A. Rafiey und M. Ester. Mining Cohesive Patterns from Graphs with Feature Vectors. In *Proceedings of the SIAM International Conference on Data Mining, SDM 2009*, Seiten 593–604, 2009.
- [NAJ04] J. Neville, M. Adler und D. Jensen. Spectral Clustering with Links and Attributes. Bericht, Dept of Computer Science, University of Massachusetts Amherst, 2004.
- [PJAM02] C.M. Procopiuc, M. Jones, P.K. Agarwal und T.M. Murali. A Monte Carlo algorithm for fast projective clustering. In *SIGMOD '02: Proceedings of the 2002 ACM SIGMOD international conference on Management of data*, Seiten 418–427, 2002.
- [RK90] P.J. Rousseeuw und L. Kaufman. *Finding Groups in Data: An Introduction to Cluster Analysis*. Wiley, New York, 1990.
- [STM07] M. Shiga, I. Takigawa und H. Mamitsuka. A spectral clustering approach to optimally combining numerical vectors with a modular network. In *KDD '07: Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, Seiten 647–656, 2007.
- [ZCY09] Y. Zhou, H. Cheng und J. X. Yu. Graph Clustering Based on Structural/Attribute Similarities. In *Proceedings of the 35th International Conference on Very Large Data Bases (VLDB 2009)*, *PVLDB Journal*, Jgg. 2, Seiten 718–729, 2009.

Konzeption und Implementierung einer Regression Datenbank zu Unterstützung der IP-Entwicklung

Michael Otto

Wilhelm Büchner Hochschule
Fachgebiet Datenbanken und Informationssysteme
michael.otto@globalfoundries.com

Art der Arbeit: Diplomarbeit

Betreuer der Arbeit: Dr. Wolfgang Kliesch

Abstract: Integrierte Halbleiterschaltungen, die flexibel in größeren Designprojekten nachnutzbar sein sollen, werden in IP-Blöcke (Intellectual Property) separiert. Solche Schaltungen müssen hohen Ansprüchen bei der Verifikation genügen. Beim HDL-Schaltungsdesign findet solch eine Verifikation durch Regression statt. Dabei werden sehr große Datenmengen generiert. Diese Zusammenfassung beschreibt, wie die dabei anfallenden Daten gesammelt und dargestellt werden können.

1 Aufgabe des Regression DB-Systems

Die hier vorgestellte Datenbank dient der Unterstützung der IC Verifikation auf RTL-Ebene¹, einem frühen Schritt im EDA-Entwicklungsprozess, der an mehreren, global verteilten Standorten, durchgeführt wird. Die IP-Daten werden durch ein VMS² verwaltet. Ziel der Verifikation ist das möglichst vollständige Prüfen der Funktion von DUT's³ durch Simulationen. Zu diesem Zweck wird für jedes DUT eine Testbench[KCC97] erzeugt. Orientiert an Testplänen werden dafür Simulationen entwickelt. Sie bestehen aus Voreinstellungen, Ablaufsteuerungen und Stimuli. Bei jedem Regression-Lauf werden mehrere Simulationen gestartet. Die Vollständigkeit der Verifikation wird über verschiedene Formen von Coverage[SYN08] bestimmt.

Das Regression DB-System unterstützt:

- Das erstellen und verwalten von Testplänen für die Verifikation,
- die Daten- und Statusdarstellung von Regression und Simulation und
- das verwalten und Speichern von Coverage Daten.

¹ Register Transfer Level

² Versionsmanagement system

³ Device Under Test

2. Die Regression-Datenbank

Bei der Umsetzung des Systems wird ausschließlich Freeware verwendet. Um die für das Regression System definierten Aufgaben zu handhaben, wurde die in Abb. 1 dargestellte Struktur erzeugt. Sie soll jeweils eine IP mit ihren Anwendungen in verschiedenen Projekten verwalten.

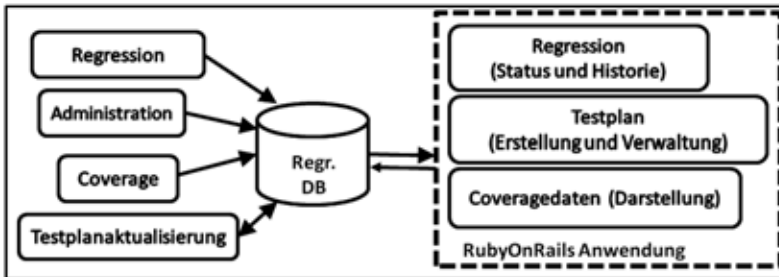


Abbildung 1: Elemente des Regression Systems

Als Datenbanksystem wird PostgreSQL[Pet03], [PGD06] eingesetzt. Regression-Daten, Testpläne und Coverage-Daten sind zusammen in einer Datenbank abgebildet[Gün00]. Die Konfigurationsdaten für die Regression-Steuerung und Datenextraktion wurden, im Interesse einer Datenbankunabhängigen Reproduzierbarkeit, nicht in die Datenbank aufgenommen. Sie werden in Form von ASCII-Dateien im VMS verwaltet.

Im Frontend der Datenbank kommen Perl-Skripte [Ran98] zum Einsatz. Ein **Regression-Skript** automatisiert das Ausfassen aus dem VMS und das Starten eines Build-Prozesses. Nachdem die Testbench in eine für den Simulator ausführbare Form Kompiliert wurde, werden die Simulationen gestartet⁴ und überwacht. Letztlich steuert das Regression-Skript das Speichern von Einstellungs- Status- und Ergebnisdaten von DUT und Testbench. **Administrations-Skripte** realisieren die Möglichkeit den Status von Simulationen manuell zu steuern und neue IP-Anwendungen in der Datenbank zu registrieren. Das **Coverage-Skript** verwaltet die durch das Simulationstool erzeugten Reporte und ist in der Lage, die Daten mehrerer Regression-Läufe zu Vereinigen. Auszüge aus den Coverage-Reporten werden in der Datenbank zur Visualisierung und Auswertung gespeichert. Ein weiteres Frontend Skript, die **Testplanaktualisierung**, wertet die in der Datenbank gespeicherten Simulationsdaten aus und schreibt daraus ermittelte Status in die zugehörigen Testpläne.

Als Frontend der Datenbank kommt eine RubyOnRails-Anwendung [DAV06] [DAV07] zum Einsatz. Dabei handelt es sich um ein Framework, mit dessen Hilfe die Inhalte der DB im Intranet des Unternehmens via WEB-Browser zugänglich gemacht werden. Es arbeitet nach der MVC-Architektur[DAV07]. Mittels WEB-Server⁵ wird dabei die Schnittstelle zum Nutzer hergestellt.

⁴ Die Simulationen werden mittels LSF (Load Sharing Facility) in einen Rechnercluster verteilt.

⁵ In dieser Anwendung kommt der WEB-Server Mongrel [DAV07] zum Einsatz

Alle drei Teilaufgaben sind über Controller mit zugehörigen WEB-Views in der Anwendung umgesetzt. Neben der Möglichkeit Abfragen zu erzeugen und deren Darstellung zu steuern wurde, wurde das Erstellen und verwalten von Testplänen implementiert.

Jede mit RubyOnRails erzeugte Abfrage kann über die URL im resultierenden WEB-Browser-View reproduziert werden. Auf diese Weise wird das Nachnutzen von Abfragen unterstützt. Durch zyklische Auswertung hunderter solcher Abfragen werden die Testplanaktualisierungen realisiert.

3 Performance

Ein wesentlicher Anspruch an eine Regression-Datenbank für den täglichen Einsatz in der Verifikation ist eine zumutbare Antwortzeit, auch im Fall komplexer DB-Operationen. Listenlängen in der DB können auf über 100-Millllionen Einträge anwachsen und Auswertungen sollen auf der gesamten Datenmenge durchgeführt werden. Um diesen Ansprüchen zu genügen wurden verschiedene Maßnahmen ergriffen.

Beobachtend: Die Aktivitäten des DB-Servers werden Anhand von Server-Load-Überwachung beobachtet. PostgreSQL verfügt über DB-Views[PGD06], die ebenfalls eine Bewertung der DB-Aktivitäten zuläßt. Letzlich verfügt das Framework über einfache Zeiterfassungsfunktion, deren Ergebnisse nach jeder Abfrage ausgewertet werden können.

Organisatorisch: Die meisten Regression-Läufe, und Backup-Operationen werden zu einem Zeitpunkt gestartet, an dem die wenigsten Suchanfragen des Weltweit organisierten Designerteams zu erwarten sind. Das DB-System wurde auf einem Dedizierten Rechnersystem installiert und Zeitaufwändige, sich wiederholende Abfragen werden im Voraus durchgeführt. Die bei suchabfragen auszuwertende Datenmenge wird limitiert, indem nur per Schlüsselwort Extrahierte Daten aus Simulations- und Coverage-Reporten der DB-gespeichert werden.

DB-Optimierung: Die Datenbank wurde nach allen Primär- und Fremdschlüsseln indiziert. Die Abfragealgorithmen wurden auf Basis von DB-Analysedaten optimiert. Abfragen, die längere Zeit in Anspruch nahmen, wurden zusätzlich durch Scheduler-Steuerung beschleunigt [Uwe08] [PGD06]. Zyklische Pflegearbeiten wie Vakuum und das Bereinigen nicht länger benötigter Daten werden regelmäßig durchgeführt.

Durch die Summe der Aktivitäten wurde eine erheblich höhere Antwortgeschwindigkeit des DB-Systems erreicht. Eine detailliertere Darstellung der Arbeiten ist in einer früheren Veröffentlichung dargestellt [Uwe08].

Literaturverzeichnis

1. **Chang, K. C.** *DIGITAL DESIGN and MODELLING with VHDL and SYNTHESIS*. Los Alamitos, Ca : Wiley-IEEE Computer Society Press, 1997. ISBN: 978-0-8186-7716-8.
2. **Dave Thomas, Chad Fowler, Andy Hunt.** *Programming Ruby*. Dallas, Texas : The Pragmatic Bookshelf, 2006. ISBN 0-9745140-5-5.
3. **Dave Thomas, David Heinemeier Hanson.** *Agile Web Development with Rails*. Dallas, Texas : The Pragmatic Bookshelf, 2007. ISBN-10: 0-9776166-3-0.
4. **Eisentraut, Peter.** *PostgreSQL Das Offizielle Handbuch*. Bonn : mitp-Verlag GmbH, 2003. ISBN 3-8266-1337-6.
5. **Günter Matthiessen, Michael Unterstein.** *Relationale Datenbanken und SQL*. München : Addison-Wesley Verlag, 2007. ISBN 978-3-8273-2656-0.
6. **James Martin, James J. Odell.** *Objektorientierte Modellierung mit UML: Das Fundament*. München : Prentise Hall, IMPRINT der MarktTechnik Buch- und Software-Verlag GmbH, 1999. ISBN 3-8272-9580-7.
7. **Peter Zöller-Greer, Otto Mildenberger.** *Softwareengineering für Ingenieure und Informatiker: Planung, Entwurf und Implementierung*. Braunschweig, Wiesbaden : Vieweg Teubner Verlag, 2002. ISBN 3-528-03939-6 .
8. **Randal L. Schwartz, Tom Christiansen.** *Einführung in Perl*. Köln : O'REILLY Verlag, 1998. ISBN 3-89721-434-2.
9. **Raymond, Scott.** *Ajax on Rails*. Köln : O'REILLY Verlag, 2007. ISBN 978-3-89721-716-4.
10. **Thaller, Georg Erwin.** *Software-Test: Verifikation und Validierung*. Hannover : Heise, 2000. ISBN 3-88229-183-4 .
11. **Uwe Hatnik, Kai Eichhorn, Michael Otto.** A Freeware based Regression Framework. [Buchverf.] Rainer G. Spallek, Günter Elst René Schüffny. *Workshop Entwurf integrierter Schaltungen*. Dresden : Fraunhofer-Institut für Integrierte Schaltungen, Institutsteil Entwurfsautomatisierung, 2008.
12. **SYNOPSIS.** *Unified Coverage Reporting User Guide*. [PDF-Dokument] Mountain View, CA : SYNOPSIS VCS, SYNOPSIS, Mai 2007. Version Y2006.06-SP1.

Konstruktion und Optimierung von Diskriminierungsnetzwerken

Christoph Terwelp

christoph.terwelp@rwth-aachen.de

Abstract:Die Bedeutung der Datenverarbeitung für den Menschen, für Unternehmen und die Gesellschaft steigt stetig, da immer größere Datenmengen konsumiert werden. Hierfür werden diese entsprechend den Konsumanforderungen aufbereitet, was häufig mit Regelbasierten Systemen (RBS) vorgenommen wird. Dieses Papier beschreibt eine Optimierung für die Erhöhung der Verarbeitungsgeschwindigkeit der Daten in RBS nach dem Prinzip der Dynamischen Programmierung.

1 Regelbasierte Systeme

Ein RBS wendet eine Menge vordefinierter Regeln auf eine große Datenmenge an und besteht im wesentlichen aus den folgenden vier Komponenten:

Arbeitsspeicher enthält die Daten, welche vom RBS verarbeitet werden und in Form von typisierten Tupeln (Fakten) vorliegen.

Regelbasis enthält die Regeln, welche auf die Fakten des Arbeitsspeichers angewendet werden. Regeln bestehen aus dem Bedingungsteil und einem Aktionsteil.

Regelinterpretierer überprüft für welche Fakttuple die Regelbedingungen erfüllt sind.

Ausführungseinheit führt die Aktionen der erfüllten Regeln aus.

2 Diskriminierungsnetzwerke

Diskriminierungsnetzwerke (DN) überprüfen Regelbedingungen auf ihre Erfüllung durch große Datenmengen. Dazu werden die Regelbedingungen in atomare Bedingungen (Tests) aufgespalten, die jeweils auf Teilen von Fakttuple operieren. Die Tests werden von Filtern vorgenommen, die jeweils einen Teil eines Fakttuple auf einen Wahrheitswert abbilden. Eine Regelbedingung wird daher als eine Menge von Filtern dargestellt. Es wird zwischen Alpha- und Beta-Filtern unterschieden, wobei Alpha-Filter nur einzelne Fakten testen und Beta-Filter mehrere Fakten eines Tupels. Ein DN besteht aus Knoten mit Ein- und Ausgängen, die entsprechend der enthaltenen Filter in Alpha- und Beta-Knoten eingeteilt werden (Abbildung 1).

Der Wurzelknoten im DN erzeugt für jeden hinzugefügten oder entfernten Fakt des Arbeitsspeichers eine Nachricht an seinem Ausgang, an dem das Alpha-Netzwerk (Abbildung 1) angeschlossen ist. Dieses besteht aus Alpha-Knoten. Alpha-Knoten haben einen Eingang, einen Ausgang, einen internen Speicher und einen Alpha-Filter. In einem Alpha-Knoten wird der Test des Filters auf die Fakten der eingehenden Nachrichten angewendet. Wenn der Test wahr ist, wird der Speicher des Knotens aktualisiert und die Nachricht an den Ausgang weitergeleitet.

Unter dem Alpha-Netzwerk liegt das Beta-Netzwerk (Abbildung 1), das aus Beta-Knoten besteht. Diese haben mehrere Eingänge und erzeugen aus den daran anliegenden Faktttupeln und den eingehenden Nachrichten neue Faktttupel, auf welche der Test des Beta-Filters angewendet und des Weiteren wie bei Alpha-Knoten verfahren wird.

Eine Regelbedingung wird durch ein DN realisiert, indem alle erforderlichen Alpha- und Beta-Knoten angelegt werden, diese zu Teilbedingungen verbunden und mit einem Endknoten zusammengefasst werden. Am Endknoten kommen alle Faktttupel an, wel-

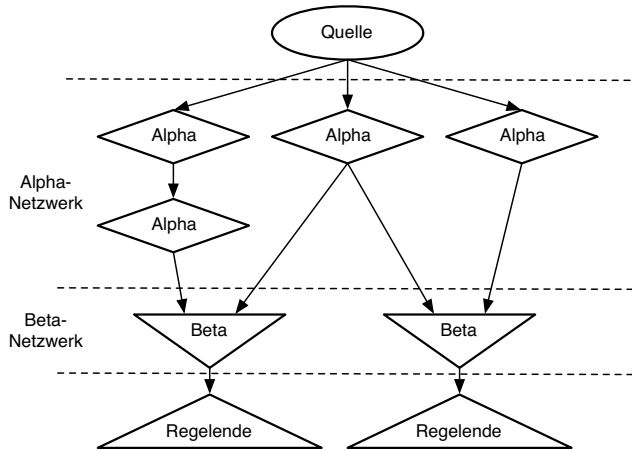


Abbildung 1: Beispiel für ein Diskriminierungsnetzwerk

3 Existierende Optimierungsansätze

Optimierungsansätze für DN nehmen häufig eine Einschränkung der Struktur des DN vor, um eine effizientere Verarbeitung der Daten zu erreichen. Beispiele hierfür sind RETE [For82], TREAT [Mir87], A-TREAT [Han92] und GATOR [Eri93]. Diese Ansätze haben die Optimierung eines DN für jeweils eine Regel als Ziel. Die Verwendung eines gemeinsamen DN für mehrere Regeln bietet weitere Optimierungsmöglichkeiten, indem Teile des DN durch mehrere Regeln gemeinsam genutzt werden (Node Sharing). In einem existierenden Ansatz [Han02] werden für die Realisierung neuer Regeln, bereits existierende Teile des DN wiederverwendet. Ein existierendes DN kann jedoch in einer Form vorliegen, die Node Sharing verhindert, obwohl ein funktional äquivalentes DN dies ermöglichen würde.

4 Optimierung durch Node Sharing

Die Erfahrung zeigt, dass sich die Regeln eines RBS im Vergleich zu den Fakten wesentlich seltener ändern. Eine Änderung der Regelbasis erfordert eine vollständige Rekonstruktion des DN, da eine Anpassung meistens aufwendiger ist. Die Optimierung des Alpha- und Beta-Netzwerks in Hinsicht auf Node Sharing wird dabei getrennt vorgenommen, da die Schnittstelle zwischen den beiden Netzwerken gleich bleibt.

Das Alpha-Netzwerks wird pfadweise konstruiert. Dazu wird der am häufigsten vorkommende Filter aus der Menge noch nicht realisierter Alpha-Filtermengen ausgewählt und dafür ein Alpha-Knoten unterhalb des Wurzelknotens angelegt. Anschließend wird der Filter ausgewählt, welcher am häufigsten in Kombination mit dem schon angelegten Filter vorkommt, und als nächster Alpha-Knoten an den Pfad angehängt. Die weiteren Filter werden analog realisiert, wobei jeweils der am häufigsten vorkommende Filter (in Kombination mit den Filtern des schon angelegten Pfades) gewählt und realisiert wird. Wenn keine weiteren Filter für einen Pfad gefunden werden, wird ab dem vorletzten Knoten im Pfad fortgefahren bis alle Alpha-Filterkombinationen durch Alpha-Knoten realisiert

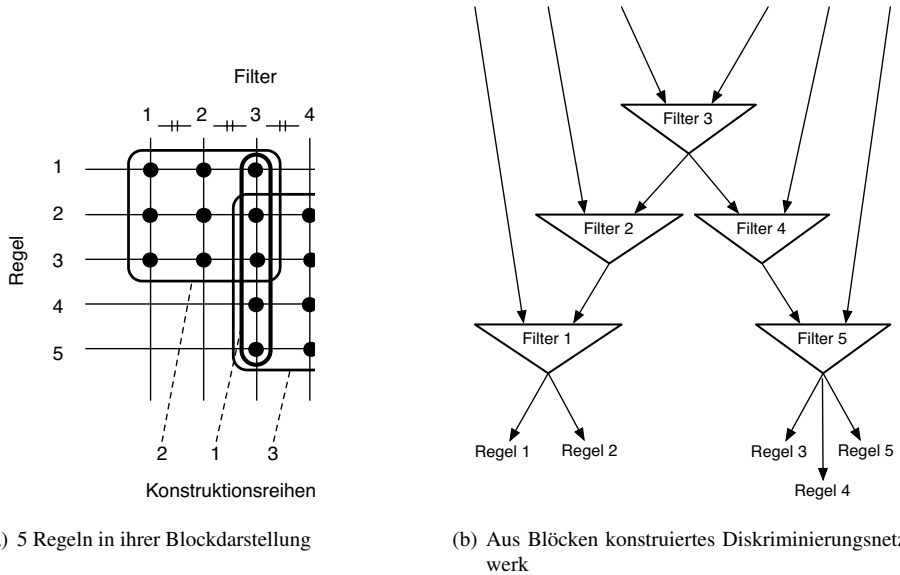


Abbildung 2: Beispielnetzwerke für die Lösung von Konflikten

Zur Optimierung des Beta-Netzwerks muss erwähnt werden, dass nicht alle Überschneidungen von Regeln gemeinsam genutzt werden können. Dies liegt daran, dass die gemeinsame Nutzung eines Netzwerks durch Regeln, die weitere Verwendung von Teilen dieses Netzwerks einschränken kann. Da Filter zu früh in diesem Netzwerk auf Faktupel angewendet werden wäre eine Rekonstruktion des Netzwerks erforderlich, die jedoch das ursprüngliche Node Sharing unmöglich machen würde.

Des Weiteren kann durch ungünstiges Node Sharing eine Vergrößerung der Faktupel in bestimmten Knoten eintreten, was zu einer Vergrößerung der Menge der zu überprüfenden Faktupel führt. Beide Probleme werden auf Konflikte zwischen Beta-Filtern reduziert. Zwei Beta-Filter stehen in Konflikt, wenn sie mindestens einen Fakt des Tupels gemeinsam verwenden. Netzwerkbereiche, die gemeinsam genutzt werden, werden durch Blöcke repräsentiert, die jeweils aus einer Menge Regeln und einer Menge Beta-Filtern bestehen. Die Beta-Filter eines Blocks müssen durch Konflikte einen zusammenhängenden Graphen bilden.

Blöcke dürfen sich überschneiden (Block 2 und 3 in Abbildung 2(a)), wodurch implizite Blöcke entstehen (Block 1). Blöcke dürfen sich jedoch nicht berühren, da ansonsten zwei Blöcke sowohl gemeinsame Regeln als auch Filter, die miteinander in Konflikt stehen enthalten würden. Dies würde zu den geschilderten Problemen führen.

Solche Blöcke werden als in Konflikt stehend bezeichnet. Eine Menge von paarweise konfliktfreien Blöcken kann konstruiert werden, indem für jede Regel und jeder ihrer Beta-Filter jeweils ein neuer Block erzeugt wird, der zunächst nur die Regel und den jeweiligen Filter enthält. Dieser wird dann um diejenigen Filter ergänzt, die notwendig sind, um Konflikte mit bereits bestehenden Blöcken zu eliminieren. Der Block ist nun konfliktfrei. Danach wird der Block um so viele Regeln und abschließend um so viele Filter wie möglich ergänzt, ohne dadurch neue Konflikte zu erzeugen. So erzeugt Blöcke, die vollständig von bereits existierenden Blöcken umschlossen sind, werden verworfen. Die übrigen Blöcken werden nach aufsteigender Filterzahl sortiert und nacheinander in Teilnetzwerke umgewandelt (Abbildung 2(b)).

5 Ergebnisse

Zur Optimierung von DN wurden Algorithmen entwickelt, die eine gemeinsame Nutzung von Teilnetzen des DN durch mehrere Regeln ermöglichen. Hierfür wurde die Blocknotation entwickelt, die sowohl eine effiziente als auch anschauliche Prüfung von Teilnetzen auf Konfliktfreiheit ermöglicht. Die Verwendung von Blöcken aus Filtern und Regeln beschleunigt die Konstruktion von DN erheblich.

Die entwickelten Algorithmen wurden für das RBS Jamocha [Jam07] implementiert. So wurde mit verschiedenen Messkonfigurationen eine Beschleunigung der Datenverarbeitung des RBS von ca. 15 Prozent erreicht.

Eine Profilinganalyse ergab, dass ca. 40 Prozent der Zeit für die Verarbeitung der Fakten zur Umwandlung der Darstellung von der Eingabesprache CLIPS [CLI05] in das interne Format des RBS benötigt wurde. Es ergibt sich also ein Geschwindigkeitsgewinn von bis zu ca. 25 Prozent für die Verarbeitungsgeschwindigkeit von Fakten. Die Zeit zur Verarbeitung der Regeln stieg deutlich an, was jedoch aufgrund der geringen Änderungsrate der Regelbasis vernachlässigt werden kann.

Literatur

- [CLI05] *CLIPS Reference Manual: Advanced Programming Guide*, 6.23. Auflage, Januar 2005.
- [Eri93] Eric Hanson and Mohammed S. Hasan. Gator: An Optimized Discrimination Network for Active Database Rule Condition Matching. Bericht, 1993.
- [For82] Forgy, Charles L. Rete: A fast algorithm for the many pattern/many object pattern match problem. *Artificial Intelligence*, 19(1):17–37, 1982.
- [Han92] Hanson, Eric N. Rule condition testing and action execution in Ariel. In *SIGMOD '92: Proceedings of the 1992 ACM SIGMOD international conference on Management of data*, Seiten 49–58, New York, NY, USA, 1992. ACM.
- [Han02] Hanson, E. N. and Bodagala, S. and Chadaga, U. Trigger Condition Testing and View Maintenance Using Optimized Discrimination Networks. *IEEE Trans. on Knowl. and Data Eng.*, 14(2):261–280, 2002.
- [Jam07] Jamocha Community. Jamocha Homepage, 2007.
- [Mir87] Miranker, Daniel P. TREAT: A Better Match Algorithm for AI Production Systems; Long Version. Bericht, Austin, TX, USA, 1987.

Ansätze zur Integration unterschiedlicher Data Warehouse-Architekturen

Vi Hoa Tu

FernUniversität in Hagen
vihoa.tu@fernuni-hagen.de

Zur Vorlage bei der Gesellschaft für Informatik e.V. (GI) – Informatiktag 2010

Art der Arbeit:	Masterarbeit (Abgabetermin: 12.11.2009)
Betreuer der Arbeit:	Prof. Dr. Gunter Schlageter, Dr. Wolfgang Wilkes Lehrgebiet Informationssysteme und Datenbanken

Abstract: Zur Unterstützung von Managemententscheidungen werden heute OLAP-Systeme (OnLine Transaction Processing) eingesetzt, die auf Data Warehouse-Systemen basieren. Die Anwendung unterschiedlicher Ansätze zur Entwicklung eines Data Warehouse führt zu unterschiedlichen Data Warehouse-Architekturen. Historisch bedingt entstehen in den Unternehmen unterschiedliche Data Warehouse-Systeme. Zunehmend ist die Tendenz festzustellen, dass diese Sattelitensysteme zusammengeführt werden müssen. Dies erfordert eine Integration der Architekturen. Das Ziel dieser Arbeit ist es, Ansätze zur Integration unterschiedlicher Data Warehouse-Architekturen herauszufinden, mit denen Data Warehouse-Systeme harmonisiert und konsolidiert werden können. Erkenntnis daraus ist die Notwendigkeit eines stufigen Aufbaus der Datenarchitekturen in der Datenhaltungsebene, um alle zukünftige Auswertungsanforderungen erfüllen zu können.

1 Grundlagen der Data Warehouse-Architektur

Die Architektur eines Data Warehouse-Systems beschreibt die Strukturebenen, die darin enthaltenen Komponenten und deren Beziehungen untereinander. Abbildung 1 gibt eine typische Data Warehouse-Architektur, die sich aus drei Ebenen zusammensetzt: In der Datenbeschaffungsebene werden extrahierte Quelldaten durch ETL-Prozesse (Extract Transform Load) in einem temporären Arbeitsbereich, sog. Data Staging Area bereinigt und anschließend in ein für Auswertungszwecke erforderliche Formt transformiert. Die aufbereiteten Daten werden in der Datenhaltungsebene, die das eigentliche Data Warehouse darstellt, je nach Data Warehouse-Ansatz in unterschiedlichen Arten von Datenbasen gespeichert. In der Datenaufbereitungsebene geht es um die zweckmäßige Aufbereitung von relevanten Informationen für OLAP-Anwendungen durch einen OLAP-Server. Bei der Umsetzung der OLAP-Schicht mit relationalen Datenbanken wird

es ROLAP (Relational OLAP), bei der Realisierung mit multidimensionalen Datenbanken MOLAP (Multidimensional OLAP) genannt. Die Organisationsform der Daten in der Datenhaltungsebene bestimmt die Architektur eines Data Warehouse. Die Formen reichen von einer auf relationalem Schema basierten zentralen Datenbasis über mehrere ebenso auf relationalem Schema realisierte ODS (Operational Data Store) bis zu den auf multidimensionalem Schema ausgerichteten Datenbasen, sog. Data Marts.

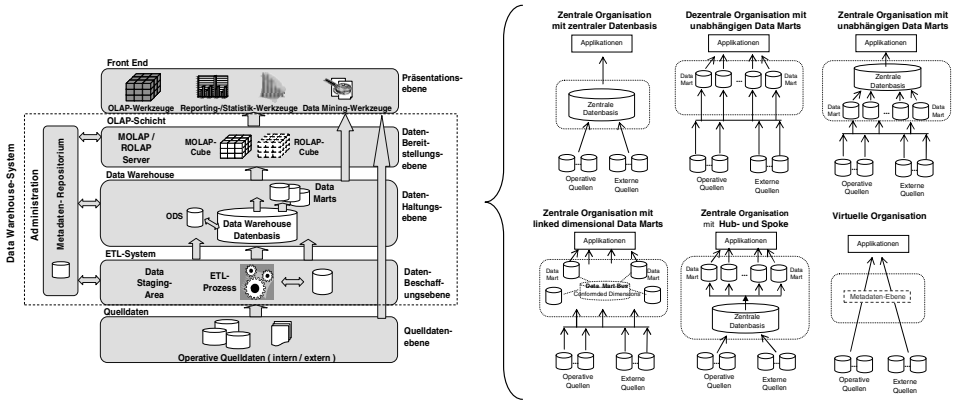


Abbildung 1 Architektur eines Data Warehouse-Systems - Organisationsformen in der Datenhaltungsebene

2 Ansätze zur Integration der Data Warehouse-Architekturen

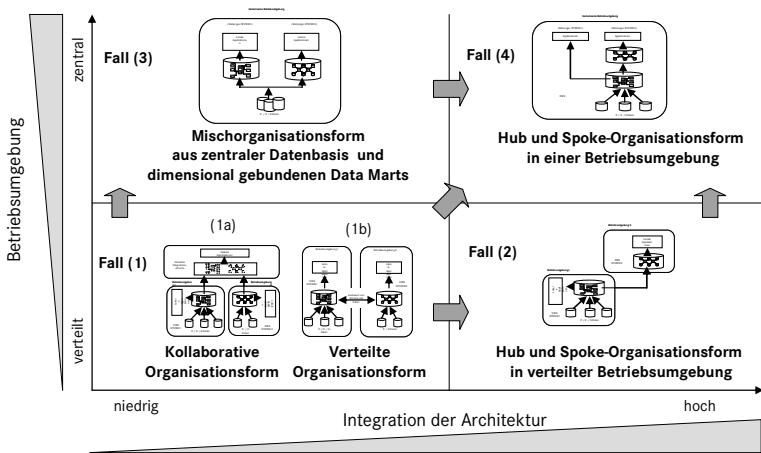


Abbildung 2 Data Warehouse-Integrationsstrategie / Integrationsformen

Ausgehend von zwei Data Warehouse-Systemen, die sich in zwei getrennten Betriebsumgebungen befinden, werden zwei unterschiedliche Data Warehouse-Architekturen untersucht. Die Untersuchung bedarf einer Betrachtung der jeweiligen BI-Architekturen. Die Feststellung der Architekturunterschiede dient als notwendige

Voraussetzung zur Ableitung der Ansätze zur Integration unterschiedlicher Data Warehouse-Architekturen. Die hinreichende Bedingung erfordert jedoch eine weitere Betrachtung des Merkmals Betriebsumgebung – im Sinne von verteilter oder zentralisierter Betriebsumgebung. Die zwei Unterscheidungsmerkmale Integration der Architektur und Verteilung/Zentralisierung der Betriebsumgebung führen zu vier Unterscheidungsfällen von Integrationsformen (Abbildung 2). Bei Nicht-Zentralisierung der Betriebsumgebung und keiner Integration der Architektur führt es zum Fall (1) mit seiner „kollaborativen bzw. verteilten Organisationsform“. Fall (2) mit seiner „Hub und Spoke-Organisationsform in verteilter Betriebsumgebung“ tritt ein, wenn keine Zentralisierung der Betriebsumgebung aber eine Integration der Architektur angestrebt wird. Fall (3) mit seiner „Mischorganisationsform“ ergibt sich, wenn eine Zentralisierung der Betriebsumgebung aber keine Integration der Architektur angestrebt wird. Die höchste Integrationsform wird im Fall (4) mit seiner „Hub und Spoke-Organisationsform in zentralisierter Betriebsumgebung“ erreicht, wenn sowohl eine Zentralisierung der Betriebsumgebung als auch eine Integration der Architektur durchgeführt wird. Die Hub und Spoke-Organisationsform, die hier als Referenzarchitektur zur Data Warehouse-Integration benutzt wird, hat ihren Ursprung im Data Warehouse-Ansatz nach W.H.Inmon. Die Unterscheidung der vier Integrationsformen führt zu einer Integrationsstrategie für Data Warehouse-Architekturen mit drei möglichen Integrationsansätzen (Tabelle 1).

Integrationsansatz	1. Integrationsstufe (optional)	2. Integrationsstufe	3. Integrationsstufe
(A) Integrieren dann Zentralisieren	Fall (1): Kollaborative oder verteilte Organisationsform	Fall (2): Hub und Spoke- Organisationsform in verteilter Betriebsumgebung	Fall (4): Hub und Spoke- Organisationsform in zentralisierter Betriebsumgebung
(B) Zentralisieren dann Integrieren	Fall (1): Kollaborative oder verteilte Organisationsform	Fall (3): Mischorganisationsform in zentralisierter Betriebsumgebung	Fall (4): Hub und Spoke- Organisationsform in zentralisierter Betriebsumgebung
(C) Direktes Zentralisieren und Integrieren	Fall (1): Kollaborative oder verteilte Organisationsform	(keine)	Fall (4): Hub und Spoke- Organisationsform in zentralisierter Betriebsumgebung

Tabelle 1 Data Warehouse-Architekturen - Integrationsansätzen

3 Ausblick - Betrachtung der Datenbasen in der Datenhaltungsebene

Die Wichtigkeit der Organisationsformen führt zu einer notwendigen Kategorisierung der Datenarchitekturen in der Datenhaltungsebene unter Betrachtung von zwei Merkmalen: 1) Die Betrachtung der Artenunterschiede bzgl. der Datenbasen: Relationales und Multidimensionales Datenschema. 2) Die Berücksichtigung des Grades der Anwendungsorientierung. Anhand dieser zwei Merkmale können vier Kategorien von Datenschemata unterschieden werden (Abbildung 3). Der Gedanke zur Kategorisierung unterschiedlicher Datenschemata in der Datenhaltungsebene führt zu einer Implementierungsstrategie, die einen stufigen Aufbau der unterschiedlichen Datenbasen ermöglicht. Durch den stufigen Aufbau solcher Datenarchitekturen in der

Datenhaltungsebene sollen zukünftige, unvorhersehbare Auswertungsanforderungen unterschiedlicher Natur erfüllt werden.

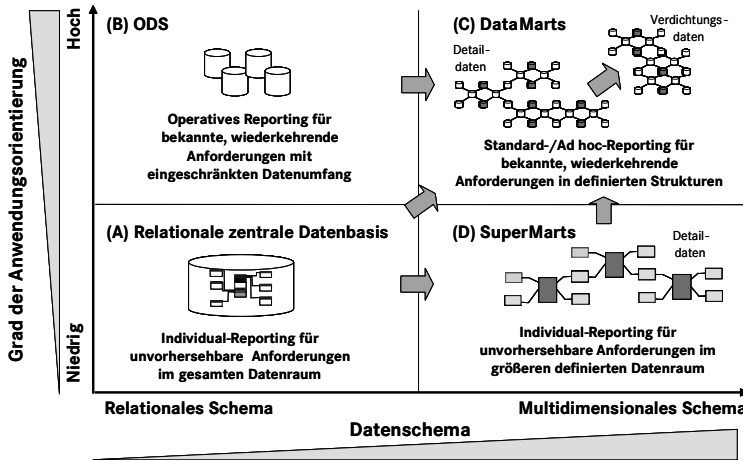


Abbildung 3 Datenarchitekturen in der Datenhaltungsebene - Kategorisierung der Datenschemata

Literaturverzeichnis

- [1] Bauer, A. / Günzel, H. (Hrsg.): Warehouse-Systeme - Architektur, Entwicklung, Anwendung. dpunkt.verlag, 2009, 3. Auflage
- [2] Goeken, M.: Entwicklung von Data-Warehouse-Systemen (Dissertation, Phillips-Universität Marburg, 2005). Deutscher Universitäts-Verlag, 2006, 1. Auflage
- [3] Inmon, W.H.: Building the Data Warehouse (Forth Edition). Wiley Publishing, Inc., 2005
- [4] Kimball, R. / Caserta, J.: The Data Warehouse ETL Toolkit - Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data. Wiley Publishing, Inc., 2004
- [5] Kimball, R. / Ross, M./ Thornthwaite, W./ Mundy, J./ Becker, B.: The Data Warehouse Lifecycle Toolkit (2nd Edition). Wiley Publishing, Inc., 2008
- [6] Lehner, W.: Datenbanktechnologie für Data-Warehouse-Systeme. dpunkt.verlag, 2003, 1. Auflage
- [7] Malinowski, E. / Zimányi, E.: Advance Data Warehouse Design – From Conventional to Spatial and Temporal Applications. Springer, 2008
- [8] Pintschovius, L.: Kooperativ verteiltes Data Warehouse - ein Architekturansatz für eine ganzheitliche Datenhaltung in der Hochschulverwaltung. Universität Mannheim, Dissertation, 2006
- [9] Sinz, E.J. / Böhnlein, M./ Plaha, M./ Ulbrich-vom Ende A.: Architekturkonzept eines verteilten Data Warehouse-Systems für das Hochschulwesen. Otto-Friedrich-Universität Bamberg, 2001.

Berechnung von Korrelation in Echtzeit auf der Cell BE Plattform

Autor: Tobias Beier
(tobias.beier@uni-jena.de)

Betreuer: Prof. Dr.-Ing. Dietmar Fey
(Dietmar.Fey@informatik.uni-erlangen.de)

Prof. Dr. Ing. Werner Erhard
(werner.erhard@uni-jena.de)

Abstract: Rechenintensive Aufgaben und die Möglichkeiten moderner Simulationsverfahren leuchten nach Rechenleistung. Trotz der Aufrechterhaltung des Mooreschen Gesetzes stagniert die erzielbare Single-Thread Rechenleistung auf aktuellen Systemen in den letzten Jahren. Dies liegt in den nur noch marginal steigenden Taktraten begründet. Massive Steigerungen der (theoretischen) Rechenleistung ist nur durch Anpassung der verwendeten Algorithmen auf die hoch-parallelen Architekturen und ihrer damit einher gehenden Parallelisierung möglich. Ein Ansatz hin zu einer heterogenen hoch-parallelen Architektur ist IBMs Cell BE Plattform. Die hier vorgestellte Arbeit beschäftigt sich mit den grundlegenden Eigenschaften der Architektur und versucht sie in vorhandene Schemata einzuordnen und einen speziellen Algorithmus anhand der gewonnen Erkenntnisse möglichst effizient auf die Cell BE Architektur zu portieren.

1 Einführung in das Themengebiet

Bei der Fertigung von z.B. optischen Oberflächen werden flächige Werkzeuge benutzt. Diese haben einen ausgedehnten Wirkbereich welcher als Werkzeug-Abtrags-Funktion (WAF) bezeichnet wird. Dieser Abtrag wirkt über einen bestimmten Zeitraum auf die Oberfläche, welcher über eine Verweilzeitfunktion (VWZ) repräsentiert wird. Das Werkzeug wird an verschiedenen Oberflächenpunkten aufgesetzt und es ergeben sich Überlappungen des Abtrags zu einem Gesamtabtrag. Mathematisch lässt sich der Vorgang durch eine Korrelation beschreiben. Diese ist rechtechnisch sehr aufwändig und sollte möglichst in Realzeit abgewickelt werden, da der Vorgang in einen Fertigungsprozess integriert ist. In der Arbeit wurde dafür die Cell BE Plattform auf ihre Eignung für diese Aufgabe untersucht.

2 Die Cell-Architektur

Die Cell BE Plattform ist eine heterogene Multi-Core-Prozessorarchitektur, die von IBM in Zusammenarbeit mit Toshiba und Sony entwickelt wurde. Sie ist heterogen, da sie ein Power5-Kern über ein spezielles Bus-System mit acht sogenannten SPEs verbindet. Diese Prozessorkerne sind über einen Speicher-Controller (MIC: „Memory Interface Controller“,) und einen Interface-Controller (BIC: „Bus Interface Controller“) mit dem Hauptspeicher bzw. anderen Komponenten (z.B. weiteren Cell-Prozessoren) verbunden.

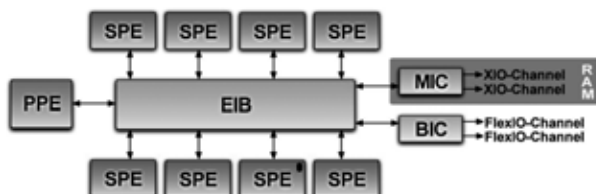


Abbildung 1: Die Grob-Struktur einer Cell BE

SPE steht dabei für „Synergistic Processor Element“. Ein SPE wiederum ist ein kompletter Prozessor für sich, der über eine Recheneinheit (SPU: „Synergistic Processing Unit“), einen eigenen lokalen Speicher (LS: „Local Store“) von 256KByte Größe und einen Speicher-Controller (MFC: „Memory Flow Controller“) verfügt.

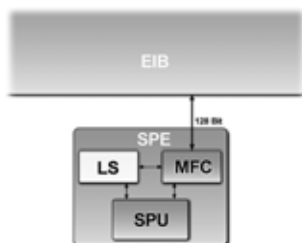


Abbildung 2: Aufbau eines SPE

Die SPU selbst ist das Herzstück des SPE. Es handelt sich hierbei um eine In-Order SIMD-Architektur, welche einen RISC-Befehlssatz verwendet und über zwei Befehlspipelines verfügt. Alle Single-Precision-Floating-Point-Befehle sind fully-pipelined. Dies ermöglicht eine parallele Abarbeitung von SIMD-Rechenoperationen und Speicherzugriffen sowie Permutationsoperationen in der SPU. Im Verlauf des Kapitels wird die Struktur der einzelnen Elemente der Cell BE Architektur weiter ausgeführt, sowie einzelne Befehle des umfangreichen Befehlssatzes erläutert [IBMa] [IBMb].

3 Portierung eines Beispielalgorithmus

Die Umsetzung des beschriebenen Abtragsalgorithmus erfordert eine variable Aufteilungsstrategie, auf Grund der beschränkten Größe des LS der SPEs. Denn diese können nicht direkt auf den Hauptspeicher zugreifen, so dass alle nötigen Daten im LS gehalten werden müssen. Sobald die Größe der Datensätze die des LS übersteigt ist eine Aufteilung der Daten erforderlich. Da die VWZ-Funktion auf Grund der realen Größenverhältnisse generell wesentlich mehr Datenpunkte besitzt ist es am sinnvollsten, zuerst diese zu partitionieren und verschiedene Gebiete auf verschiedenen SPEs rechnen zu lassen. Sobald allerdings auch die WAF nicht mehr im LS gehalten werden kann wird es notwendig, diese ebenfalls aufzuteilen. Dadurch erhält man für jeden Teil der WAF eine eigene neue Abtragsfunktion, deren Ergebnisse sich im Anschluss wiederum zum Gesamtabtragsfeld summieren. Der Aufwand dieser Summation ist glücklicherweise im Vergleich zum Rechenaufwand der Korellation vernachlässigbar gering.

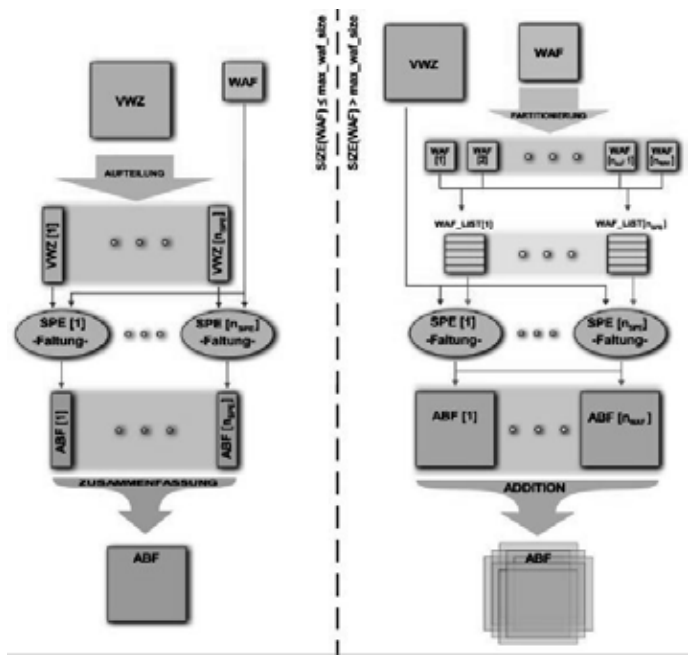


Abbildung 3: Globaler Ablauf der Abtragsberechnung

Im Vergleich zu optimiertem Code auf einem PC (Multithreading, SSE) welcher eine Rechenleistung von ca 10,4 GFlops/s erreichte, stellt die auf der Cell BE erreichte Leistung von 140 GFlops einen beachtlichen Leistungssprung dar.

4 Auswertung

Die Cell BE Plattform ist eine hoch-parallele Architektur, welche energieeffizient hohe Rechenleistung zur Verfügung stellt. Mit optimiertem Code ist es möglich, einen großen Teil der theoretischen Rechenleistung zu erreichen. So wurden mit der in der Arbeit beschriebenen Vorgehensweise 17,5 GFlop/s pro SPE erreicht, bei 25,6 GFlops theoretischer Rechenleistung. Dieser Wert ließe sich durch komplettes Auflösen der Schleifenstrukturen noch weiter erhöhen, würde aber in sehr unübersichtlichem Code resultieren. Offensichtlich ist der Programmieraufwand mittels der zur Verfügung gestellten C-Intrinsics sehr hoch. Algorithmen, welche viele Sprünge/Verzweigungen produzieren sind generell zu vermeiden, da sie auf Grund fehlender dynamischer Sprungvorhersage kaum effizient umzusetzen sind. Diese Nachteile sind auf die simpel gestrickte Befehlsverarbeitung der Cell BE Architektur und deren langen Pipelines zurück zu führen. Bei genauer Betrachtung wird offensichtlich, dass bei der Entwicklung viele Innovationen der letzten Jahre über Bord geworfen wurden. Genannte Einschränkungen machen die Software-Entwicklung für die Cell BE Plattform leider schwierig und mitunter sehr langwierig.

Literatur

[IBMa] IBM. C/C++ Language Extensions for Cell Broadband Engine Architecture Version 2.5.

[IBMb] IBM. Cell Broadband Engine Architecture.

Decodability Model for Dynamically Adaptive Prioritization Algorithm in H.264/SVC Video Streaming

Tsheten Dorji

The Srindhron Thai-German Graduate School of Engineering
The King Mongkut's University of Technology North Bangkok, Thailand
in cooperation with
Department of Computer Science / RWTH Aachen University, Germany

tshetendorji@hotmail.com

Art der Arbeit: Master Thesis Work

Betreuer/in der Arbeit: Prof. Dr. Dr. h.c. O. Spaniol

Abstract: In this paper, a novel approach for H.264/SVC packet prioritization for real-time streaming is presented. This approach is based on the decodability concept to assess the relative importance of delay sensitive packets. This scheme exploits the information of the bitstream structure to measure the relative importance of the frame w.r.t the impact of error propagation due to data dependencies. A decodability model for H.264/SVC is proposed. The algorithm is tested with different video streams and a modified JVSM-9.18 encoder. The simulation results show that the proposed algorithm generates priority numbers for every frame in a GOP as unique as possible, and can be used in conjunction with a proper packet scheduling algorithm.

1 Introduction

Real-time video streaming over heterogeneous networking environments demands effective solutions in providing optimal quality of services (QoS). In order to meet these requirements, H.264/SVC [4, 5, 6] was developed which is an extension of the H.264/AVC standard [3]. However, the packet scheduling algorithm should be able to deliver the video packets to multiple clients with optimal QoS in respective of the transmission channel. The H.264/SVC stream has a layered structure for scalability supports, and thus a good packet prioritization scheme for scheduling is needed. Several schemes were proposed in [1, 2, 8] to address the problems which prioritized the packets based on the information at the frame level of stream structure and did not consider errors due to the transmission channel or packet dropping. However, [7] proposed a unique approach introducing the decodability concept but applied it to H.263 only, and ignored the use of B frames and multiple reference pictures. In this work the concept is extended to scalable video streams, and a decodability model for H.264/AVC is presented.

2 H.264/SVC Stream Analysis Overview

The JSVM encoder is used in its single coding mode for simplicity. Frames of the test input video sequence (MOBILE_176x144_15.yuv) are encoded with multiple temporal resolutions, and the required video data statistics for stream analysis and simulations are extracted. The scalable video stream has a layered dependency structure with multiple reference pictures which are managed in two lists where List-0 stores all forward reference pictures (FRP) and List-1 stores all backward reference pictures (BRP) for decoding depending frames. Any error in a frame will get propagated to all depending frames, and affects their decodability. However, this effect is limited to frames in the same GOP only. A longer GOP length has a higher probability of error propagation, and impacts the decodability of the frame. For a stream encoded with a given GOP length the dependency structure for all GOPs in the entire stream will basically be the same.

3 Decodability Model for H.264/SVC Stream

The decodability is the probability by which a frame can be decoded without error. The aim of this algorithm is to optimize the number of decodable frames. The decodability of the frame depends on the decodability of all its macro-blocks (MB) without error. The decodability $I_{k,l}$ of the MB depends on its reference blocks, an expected error rate and MB types: 1) intra, 2) inter or 3) skip blocks are referred to as 1) inter-P- or -B-block, 2) intra-I- or -P- or -B-block, or 3) skip-P- or -B-block based on the frame type it belongs. An intra block is decodable independently of other frames. A skip block gets decoded by simply copying the content of the previous frame. When decoding an inter block, it depends on one or two reference blocks in the reference pictures. Assume that frame k contains n total blocks given by the sum of $n_{I,k}$ intra I blocks, $n_{P,k}$ intra P or B blocks, $n_{B,k}$ inter P or B blocks, and $n_{S,k}$ skip blocks. The decodability I_k of a frame can be defined as the mean value of the decodability of all blocks. The simplified formula for the decodability of frame k is

$$I_k = \frac{1}{(n - (1 - \varepsilon_k)n_{P,k})} [n_{S,k} \cdot I_{k-1} + (1 - \varepsilon_k) n_{I,k} + n_{B,k} \{ \alpha_k \cdot I_{k-m(k)} + \beta_k \cdot I_{k+n(k)} \} (1 - \varepsilon_k)] \dots\dots\dots(2)$$

Where α_k and β_k are average value of decision parameters for using FRP or BRP respectively by all inter blocks of frame k such that $0 \leq \alpha_k \leq 1$, $0 \leq \beta_k \leq 1$ and $\alpha_k + \beta_k = 1$ for a B frame; $\alpha_k = 1$ and $\beta_k = 0$ for a P frame. $I_{k-m(k)}$ and $I_{k+n(k)}$ are the decodability values of the reference frames. ε_k is the expected error rate of frame k . If $I_k = 1$ the frame is completely decodable without error. If $0 < I_k < 1$ the frame is either partially decodable or only with some error probability. Otherwise, the frame is not at all decodable. The decodability of a video sequence is the mean value of the decodability of all frames.

Since a scheduling decision (especially the decision to transmit a packet or not) changes the expected error of a frame, it changes the decodability of the video stream. To optimize the presentation quality packets have to be transmitted in a way optimizing the decodability of the video stream, i.e. the scheduler should transmit packets which strongly affect the decodability in a positive way.

4 Simulation Results

All stream statistics are extracted by the JSVM-9.18 encoder configured in single layer coding mode. Different sets of these data are extracted by varying the encoder parameters, particularly the GOP size, Maximum frame rate and input-output frame rate with default values for the remaining encoder parameters. The first data set is GOP8s from the stream encoded with GOP size=8 and Maximum Frame Rate (MFR), Frame Rate In (FRI) and Frame Rate Out (FRO) set to 30Hz. The decodability model is evaluated for all data sets based on the expected error rate of frame k given the value of all other variables. The examples show the model with two cases: without and with an expected error.

Case-I: Assume $e=0$ for every frame, so all frames are completely decodable as depicted in figure-1 for one example stream. The x-axis represents the frames, and is subscripted with the frame ids of the first two GOPs in decoding order. The y-axis depicts the decodability value $l[k]$ of the frame and its structure by showing the percentage of the different block types it consists of.

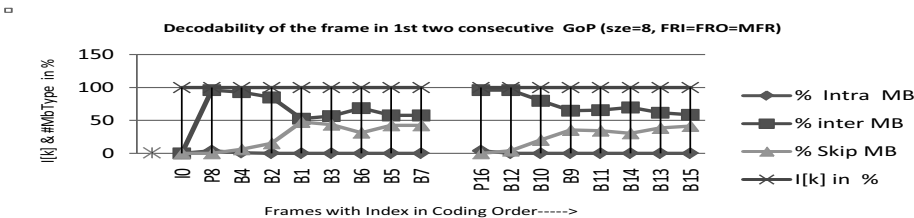


Figure-1: Decodability of the Frames without Error in Coding Order for GOP8s

Case-II: To further highlight the properties of the decodability concept we will highlight what happens when it is not possible to reliably transmit every frame. We start with the assumption that everything can be transmitted (see $e=0$ in left graph of figure-2). Now we assume that different frames may get lost with a probability of $e=25\%$. Due to this error it is not only impossible to decode the frame itself but also the decodability of depending frames is reduced. The more important a frame is the more depending frames are affected, and the effect usually is stronger. This error propagation is limited to one GOP. Suppose, that any frame from the highest temporal layer say B7 is dropped by scheduler then $e=1$, it only effects the frame itself. In the right graph of figure-2 you see how the error probability of a frame affects its decodability: The more likely a loss is, the lower the decodability gets. A change in the expected reception probability of a packet changes its decodability.

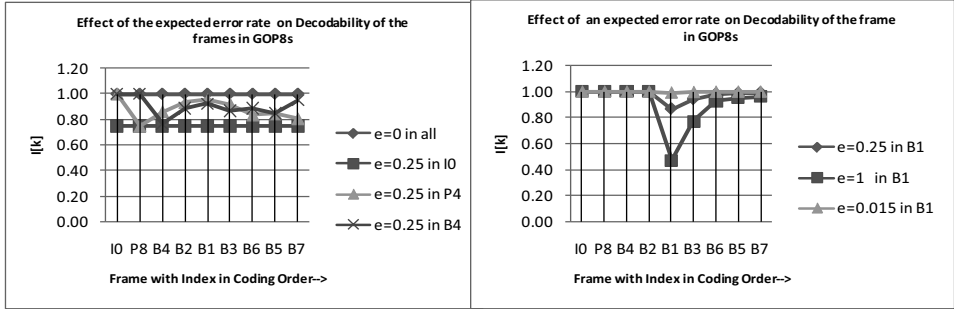


Figure-2: Effects of Error Propagation on the Decodability of Frames

5 Conclusion

In this paper, we have proposed a unique scheme for video packet prioritization to assess the relative importance of frames in a H.264/SVC scalable video stream. The scheme uses information about the structure of a stream to estimate the relative importance of frames based on the impact of error propagation on decodability of other frames. The simulation results show how this approach can be applied to H.264/SVC to prioritize the frames in GOP. In conjunction with a proper scheduling algorithm these priority values support packet dropping in case of limited bandwidth.

Bibliography

- [1] D. G allucci, A. F iandrotti, E . Masala an d J uan C arlos: C ontent-Adaptive R obust H.264/SVC Video Communications over 802.11e Networks, I nternational Conference on Advance Information Networking and Applications, 2009
- [2] D. G allucci, A. F iandrotti, E . Masala an d J uan C arlos: T raffic P rioritization of H.264/SVC V ideo over 802.11e A d Hoc W ireless Networks, i n Proc. of the IEEE ICCCN, 17th Intl. Conference on Computer Communication and Networks, 2008
- [3] J. Lee, H. K alva: The VC-1 and H.264 Video Compression Standards for Broadband Video Services, April 2008
- [4] H. S chwarz, M. . Wein: The Scalable V ideo C oding E xtension of the H.264/AVC Standard, i n IEE Signal Processing Magazine[135], March 2008
- [5] H. S chwarz, T. Wiegand, C ontribution of the I mage C ommunication Group: Scalable Video Coding (SVC) amendment of the H.264/AVC Standard, August 2008.
- [6] T. Wiegand, H. S chwarz, IEEE: Over View of Scable Video Coding Extension of the H.264/SVC Standard i n IEEE Transactions on Circuits Systems for Video Technology, VOL.17, No.9, September 2007.
- [7] J. Kritzn er, U. Horn, and M. Kampmann: Priority generation for video streaming using stream decodability: i n Proceedings of the Packet Video 2004, Irvine, CA, Dec. 2004.
- [8] Y. C ho, D. K K won, C .-C Kuo, R. Huang, C. L imma, J. D B lack, and J. J Pan: Video Streaming over 3G Networks with GOP- based Priority Scheduling, i n Proc. Intelligent Information hiding and Multimedia Signal Processing, Dec. 2006, pp. 201-204

Verifikation der Java-Echtzeitfähigkeit für den Einsatz in zeitkritischen Systemen

Dominik Franke

RWTH Aachen

Lehrstuhl Informatik 11

dominik.franke@rwth-aachen.de

Art der Arbeit: Diplomarbeit

Betreuer/in der Arbeit: Prof. Dr.-Ing. Stefan Kowalewski

Abstract: Die *Real-Time Specification for Java* ist eine Erweiterung für Standard-Java, mit der Echtzeitfähigkeit in Java-Programmen garantiert werden kann. In der vorliegenden Arbeit wird Java Real-Time System 2.1, eine Implementierung der Real-Time Specification for Java, auf drei ausgewählte Echtzeit-Kriterien getestet: zeitliches Verhalten bei Allokationen in allen vorhandenen Speicherbereichen, zeitliche Vorhersagbarkeit während der Kommunikation mit WaitFreeQueues und die Pünktlichkeit des Systems. In allen Tests werden die Möglichkeiten, die Real-Time System 2.1 bietet, um deterministische Ergebnisse zu erlangen, durchleuchtet und verwendet. Dabei werden sowohl Vor- als auch Nachteile von Real-Time System 2.1 und der Dokumentation besprochen. Die Ergebnisse zeigen, dass die echtzeitfähige Version von Java ein deutlicher Fortschritt gegenüber Standard-Java ist, decken jedoch auch Lücken auf, welche die Echtzeitfähigkeit von Programmen gefährden können.

Motivation

Echtzeitsysteme sind heutzutage in vielen Bereichen der Industrie und Technik präsent. Zum Beispiel werden in vielen sicherheitskritischen Systemen echtzeitfähige Lösungen benötigt, um die korrekte Arbeitsweise in einem vorgegebenen zeitlichen Rahmen zu gewährleisten. In Flugzeugen sorgen Echtzeitsysteme dafür, dass das Fahrwerk auf Wunsch innerhalb einer bestimmten Zeit ausgefahren ist. In Kernkraftwerken sind Echtzeitsysteme dafür verantwortlich, bei einem Notfall die Reaktoren in einen sicheren Zustand zu bringen [Bra03]. In beiden Beispielen hätte ein Fehler in dem entsprechenden Echtzeitsystem gravierende Auswirkungen. Da sich Echtzeitsysteme im Vergleich zu nicht-Echtzeitsystemen keine oder nur beschränkt Ausfälle im Betrieb erlauben dürfen, ist es wichtig, solche Systeme zu testen und ständig zu verbessern.

Oftmals sind Echtzeitsysteme auf Grund ihres speziellen Einsatzes und der hohen Anforderung an die Verfügbarkeit und Zuverlässigkeit individuelle und damit kostspielige Lösungen. Dabei sind die Aufgaben dieser Systeme häufig durch handelsübliche PCs lösbar. Doch diese können selten die Anforderungen nach Zuverlässigkeit und Echtzeitfähigkeit erfüllen. Ein handelsüblicher PC, der Echtzeitanforderungen genügt, ist in vielerlei Hinsicht attraktiv für den heutigen Echtzeit-Markt:

- Er ist eine kostengünstige Alternative zu den heutzutage häufig verwendeten speziellen Echtzeit-Lösungen.
- Im Gegensatz zu den individuellen Systemen kann er schneller an andere Aufgaben angepasst werden.
- Durch die weite Verbreitung des PCs ist er bei einem Ausfall oder bei Wartungsarbeiten schneller und günstiger zu ersetzen.
- Mit Java oder anderen Programmiersprachen betrieben, benötigt man keine so aufwendigen und teuren Schulungen des Personals wie für spezielle Lösungen.

Die *Real-Time Specification for Java* ist eine Spezifikation, auf deren Basis echtzeitfähige Implementierungen von Java realisiert werden können [GB00]. Aber wozu benötigt man neben den bereits vorhandenen Möglichkeiten echtzeitfähiges Java? Moderne Programmiersprachen wie C++ und Java sind gegenüber Programmiersprachen wie Assembler und C, die heutzutage noch in vielen Echtzeitsystemen der Industrie verwendet werden, besser geeignet, um komplexe Vorgänge abzubilden [ME07]. Des Weiteren stellen objektorientierte Programmiersprachen den aktuellen Stand der Technik dar, um große und komplexe Systeme zu erstellen. Da Java unter den objektorientierten Programmiersprachen relativ neu ist, beinhaltet sie viele Lösungen für Probleme, die die Objektorientierung mit sich bringt [Mar98]. Java ist zum Beispiel robust, verteilt und sicher, portabel, dynamisch und ermöglicht Multithreading [JG00]. Java hat jedoch auch Nachteile, die insbesondere keine Echtzeitfähigkeit bei Standard Java zulassen. Die Real-Time Specification for Java soll diese Mängel beseitigen, sodass diese moderne objektorientierte Programmiersprache in Echtzeitsystemen Anwendung findet.

Java Real-Time System ist ein Versuch von Sun, eine echtzeitfähige Version von Java auf den Markt zu bringen. Ziel dieser Arbeit war es, die Echtzeitfähigkeit dieser Implementierung zu untersuchen und dazu ein entsprechendes Testframework zu implementieren.

Testkriterien

Den Kern des fertigen Testframeworks bilden drei Black-Box-Tests [HMS07]. Als Black-Box gilt dabei Java Real-Time System, dessen Quellcode nicht frei verfügbar ist. Die Tests wurden nach folgenden Kriterien ausgesucht:

- allgemeine Anforderungen an Echtzeitsysteme (Rechtzeitigkeit, Verfügbarkeit, ...)

- Anforderungen der Real-Time Specification for Java an ihre Implementierungen (Verhält sich die Allokationszeit in einem Speicherbereich so, wie spezifiziert?, ...)
- Besonderheiten von Java Real-Time System gegenüber anderen Echtzeitsystemen
- persönliche Erfahrungen mit Echtzeitsystemen

Die drei resultierenden Tests decken die Bereiche Allokationsgeschwindigkeit, Kommunikationsgeschwindigkeit und Timer-Pünktlichkeit ab. Der Allokationstest misst die Allokationszeiten in den unterschiedlichen Speicherbereichen Scoped-, Immortal- und Heap-Memory, indem 14 unterschiedliche Allokationsgrößen, von 32 Byte bis 256 Megabyte, in jedem Speicherbereich jeweils 100.000 Mal alloziert werden. Beim Kommunikationstest wird die Kommunikationsgeschwindigkeit zwischen einem Standard-Java Thread und einem Echtzeit-Thread gemessen, welche über WaitFreeQueues miteinander kommunizieren. Im Timertest wird sowohl die durchschnittliche, als auch die worst-case Unpünktlichkeit der Timer berechnet. Dabei feuert der Timer periodisch in neun unterschiedlichen Intervallen von 100 Mikrosekunden bis 1 Sekunde.

Ergebnisse

Die Resultate der Arbeit zeigen, dass sich Java Real-Time System 2.1, zumindest im Bereich der weichen Echtzeit, als durchaus echtzeitfähig erweist. Das neuartige Speicherkonzept mit mehreren unterschiedlichen Speicherbereichen ist bei richtiger Verwendung durchaus hilfreich. So werden zum Beispiel während des Allokationstests in 500 Megabyte RAM-Speicher 50 Terabyte Daten ohne Zwischenfälle alloziert. Auch die Speicherbereiche, die eine zeitliche Vorhersage über ihre Allokationszeit machen, entsprechen ihrer Spezifikation und halten diese Zeiten ein. Bei diesem Test ist jedoch aufgefallen, dass ein nebenläufiger Standard-Java Thread die Stabilität der Allokationszeiten negativer beeinflusst, als ein nebenläufiger Echtzeit-Thread. Wenn der Scheduler ausschließlich die in der *Real-Time Specification for Java* spezifizierten Echtzeit-Threads verwendet, so sind die Ausführungszeiten stabiler und voraussagbarer. Eine andere überraschende Feststellung beim Allokationstest ist, dass zwei als unterschiedlich propagierte Speicherbereiche (*LT-Memory* und *VTMemory*) nahezu exakt das gleiche zeitliche Allokationsverhalten zeigen. Wir vermuten, dass die Implementierung der beiden Speicherbereiche in Java Real-Time System identisch ist.

Beim Kommunikationstest, bei dem nebenläufige Standard-Java und Echtzeit-Threads mit unterschiedlichen Prioritäten über WaitFreeQueues kommunizieren, konnten wir bestätigen, dass dieses Kommunikationskonzept gut und zuverlässig in Java Real-Time System umgesetzt ist. Zudem ist diese Art der Kommunikation, abhängig von der Anzahl und Last der nebenläufigen Threads und Prozesse, schnell. Mit einem nebenläufigen Thread beträgt die maximale Schreibrate auf dem Testrechner 181.000 Schreibvorgänge pro Sekunde. Die maximale Lese rate beträgt 250.000 Lesevorgänge pro Sekunde. Der Scheduler, der bei dieser Kommunikation eine wichtige Rolle spielt, konnte dabei sowohl die Echtzeitbedingungen einhalten, als auch die Prioritäten der unterschiedlichen Threads beachten.

Bei den Timertests lag die Timer-Latenz im Bereich 60-1000 μ s, abhängig vom Timer-Intervall. Das ist mehr, als bei kritischen Anwendungen, z.B. in der Motorsteuerung akzeptabel ist. Des Weiteren konnten wir Aussetzer bei einem regelmäßig feuernenden Timer messen. Für weiche Echtzeitanwendungen, die im Millisekunden-Bereich aktiv sind, reichen die Timer mit den gemessenen Eigenschaften oftmals dennoch aus.

Sowohl positive als auch negative Erfahrungen mit Java Real-Time System konnten wir mit einer sehr aktiven Community und dem Java Real-Time System Entwickler-Team über zahlreiche Foren teilen. Somit haben wir während dieser Zusammenarbeit einige Software-Bugs aufgedeckt und konnten Verbesserungs-Vorschläge für die kommenden Java Real-Time System Releases mitteilen. Im September 2009 ist ein neues Release von Java Real-Time System erschienen, in dem viele unserer Verbesserungsvorschläge bereits umgesetzt sind.

Ein besonderer Dank gilt John Schommer für die interessante Themenstellung, die stete Förderung dieser Arbeit und die anregenden wissenschaftlichen Diskussionen.

Literatur

- [Bra03] Rüdiger Brause. *Betriebssysteme: Grundlagen und Konzepte*. Springer-Verlag, 3. Auflage, 2003.
- [GB00] Ben Groszogl Greg Bollella, James Gosling. *The Real-Time Specification for Java*. Addison-Wesley, 2000.
- [HMS07] Richard Seidl Harry M. Sneed, Manfred Baumgartner. *Der Systemtest*. Hanser Verlag, 2007.
- [JG00] Guy Steele James Gosling, Bill Joy. *The Java Language Specification*. Addison-Wesley, 2000.
- [Mar98] Robert C. Martin. Java and C++: a critical comparison. In *Java Gems: jewels from Java Report*, Seiten 51–68, New York, NY, USA, 1998. Cambridge University Press.
- [ME07] David Boubekeur Mc Enery, James Hickey. Empirical evaluation of two main-stream RTSJ implementations. In *JTRES '07: Proceedings of the 5th international workshop on Java technologies for real-time and embedded systems*, Seiten 47–54, New York, NY, USA, 2007. ACM.

Entwurf eines generischen, applikationsspezifischen, transportgesteuerten Prozessor-Modells in VHDL und Validierung auf einem FPGA

Stefan Hauser

hausers@cs.tu-berlin.de

Hintergrund: Die stetig fortschreitende Entwicklung von Prozessoren hat in den letzten Jahren eine gravierende Änderung erfahren. Bisher galt es stets die Frequenz der einzelnen CPUs zu erhöhen. Mittlerweile geht der Trend jedoch zunehmend in Richtung Mehrkernprozessoren. Neben dieser generellen Tendenz werden auch andere neue Wege im Bereich der Prozessorarchitekturen eingeschlagen. Transportgesteuerte Datenflussarchitekturen bieten in diesem Bereich eine interessante Alternative. Des Weiteren gewinnen auch applikationsspezifische Prozessoren an Bedeutung. Diese Arbeit beschreibt den Entwurf, die Implementierung sowie die Validierung eines applikationsspezifischen und transportgesteuerten Prozessor-Modells.

1 Einleitung

Der grundsätzliche Aufbau einer transportgesteuerten Architektur besteht aus unterschiedlichen Funktionseinheiten, welche über ein Verbindungsnetzwerk Daten untereinander austauschen können (siehe [Cor98]). Der Datenaustausch wird durch Transportoperationen gesteuert, welche den Datenfluss, und damit den Ablauf der Applikation, innerhalb des Prozessors regeln. Corporaal beschreibt dabei ausschließlich sehr einfache Funktionseinheiten, die über eine einzige Operation verfügen. Im Entwurf von Menge (siehe [Men03]) werden komplexere Funktionseinheiten verwendet, denen zusätzliche Steuerinformationen für die Auswahl der richtigen Operation zur Verfügung gestellt werden müssen. Auch der vorliegende Entwurf sieht vielseitige Funktionseinheiten vor. Darüber hinaus kann die Art und Anzahl der Einheiten in Abhängigkeit der eigenen Applikation festgelegt werden. Das Design lässt sich ohne größeren Aufwand um eigene Funktionseinheiten erweitern, sofern diese die spezifizierte Schnittstelle einhalten. Des Weiteren können nicht nur die Funktionseinheiten an die eigene Applikation angepasst, sondern auch das Verbindungsnetzwerk kann für die eigene Anwendung adaptiert werden. Die Auswahl einer geeigneten Konfiguration erfolgt idealerweise mit Hilfe von Analyseverfahren, wie sie beispielsweise in [Gre08] vorgestellt werden.

2 Realisierung

Entscheidend für den erfolgreichen Einsatz einer derartigen Architektur ist das Verbindungsnetzwerk. Eine Datenübertragung mit geringster Latenzzeit stellt dabei das wesentlichste Entwurfskriterium dar. Daher sind schalterbasierte Netzwerke, wie sie für Mehrkernprozessoren verwendet werden, ungeeignet. Im vorliegenden Design wurde aus diesem Grund auf eine klassische busbasierte Verbindungsstruktur gesetzt. Neben der Anzahl der Busse kann auch festgelegt werden, welche Funktionseinheiten mit welchen Bussen verbunden werden. Das bietet mehr Freiraum für die Adaption an die eigene Applikation, da eine Verknüpfung aller Einheiten mit allen verfügbaren Bussen zwar maximale Flexibilität bietet, jedoch steigt damit auch signifikant der Hardwareaufwand und die Verzögerungszeit. Eine Spezifikation des Verbindungsnetzwerkes kann im festgelegten XML-Format angegeben werden und wird von den implementierten Werkzeugen weiterverarbeitet.

Bei der Anbindung der Funktionseinheiten zeigt sich eine wesentliche Herausforderung der Architektur. Bei fast allen Instruktionen handelt es sich um dyadische Operationen. Demzufolge konsumieren die meisten Funktionseinheiten je Operation zwei Werte, liefern aber jeweils nur ein Resultat. Ein effizienter Betrieb ist nur möglich, wenn nahezu alle Funktionseinheiten ausgelastet sind. Um das zu ermöglichen, wurden verschiedene Lösungsansätze entwickelt. Als erstes wurde eine erweiterte Speicherschnittstelle entworfen, welche dem Netzwerk mehrere zusammenliegende Datenworte anbietet und somit den parallelen Zugriff auf Speicherbereiche ermöglicht. Ferner können Werte über einen Bus mehreren Konsumenten zur Verfügung gestellt werden. Diese Form des Busbetriebes wird als Multicast bezeichnet.

Wie in [MHG09] beschrieben wurde, zeigen einfache Analysen von Programmen, dass Werte direkt von weiteren Operationen konsumiert werden. Dabei lässt sich oft eine spezielle Abfolge (z.B. Multiplikation und Addition) finden, die sehr häufig auftritt. Zur Unterstützung derartiger Befehlsfolgen können Funktionseinheiten direkt miteinander verbunden werden. Dadurch wird unter Umständen eine weitere Ausdünnung der Netzvermaschung möglich. Weiterhin lässt sich anhand von Applikationen erkennen, dass Werte häufig mehrfach von einer Einheit benötigt werden. Diese können, angestoßen durch spezielle Operationen, in den Eingangsregistern der Funktionseinheiten vorgehalten werden. Bei den Programmuntersuchungen fällt ebenfalls auf, dass Resultate oft von der sie erzeugenden Funktionseinheit weiterverarbeitet werden. Dieses Prinzip entspricht im Grunde einem 1-Addressrechner, bei dem das Ergebnis über ein Akkumulatorregister rückgekoppelt wird. Eine derartige Arbeitsweise der Funktionseinheiten wird von der Architektur ebenfalls unterstützt und verringert die Kommunikation über das Netz. Alle vorgestellten Mechanismen zur Reduktion der Netzauslastung lassen sich über spezielle Operationen aktivieren und werden durch eine spezielle Schnittstelle, welche die eigentlich Funktionseinheit mit dem Netzwerk verbindet, realisiert.

Da nicht alle erzeugten Werte im Folgetakt weiterverarbeitet werden können, ist eine Pufferung in prozessorinternen Registern erforderlich. Ein gemeinsamer Registerspeicher, der den Einheiten über das Verbindungsnetzwerk zur Verfügung steht, erschien nicht optimal. Im vorliegenden Design wurde auf einen verteilten Registerspeicher gesetzt, welcher im Prinzip aus den den Funktionseinheiten nachgeschalteten Ergebnisregisterspeichern be-

steht. Dadurch kann ein gegebenenfalls unnötiges Sichern von Werten in einem entfernten Register vermieden werden. Konsequenzen aus dieser Form des Registersatzes ergeben sich für das Programmiermodell, welches erheblich an Komplexität zunimmt. Für die Evaluation des Designs wurde ein Werkzeug konzipiert, welches einen Assembler für konkrete Prozessorinstanzen generieren kann.

Die erzeugten Binäroperationen steuern neben dem Verbindungsnetzwerk auch die einzelnen Funktionseinheiten. Daher kann zwischen der eigentlichen Transportoperation und der Kontrolloperation unterschieden werden. Interessant ist, wie die Kontrollinformation zur konkreten Funktionseinheit gelangt. Dazu wurden zwei unterschiedliche Wege untersucht: Zum einen die Bündelung von Kontroll- und Transportoperationen, welche sich für das Programmiermodell als übersichtlicher erwies, und zum anderen die unkonventionell wirkende Trennung beider Operationen. Klassische TTA Architekturen setzen auf die erste Variante, da die Funktionseinheiten sehr einfach gehalten sind und die Unterscheidung zwischen verschiedenen Operationen über virtuelle Adressen erfolgt.

3 Ergebnisse

Eine derartige Trennung erwies sich als äußerst vorteilhaft, da die Steuerinformationen gezielt für jede Einheit abgelegt werden können. Sind diese Instruktionen an die Transportoperation gekoppelt, so muss das Netz nicht nur die Daten zur richtigen Zieleinheit transportieren, sondern zusätzlich auch die Steuerinformationen richtig weiterleiten. Einsparungen zeigen sich sehr deutlich in der Synthese der unterschiedlicher Konfigurationen. Eine Abschätzung der notwendigen Hardwareressourcen kann bereits vor der Synthese anhand von empirisch ermittelten Formeln erfolgen.

Synthetisiert wurden unterschiedlichste Konfigurationen für FPGAs der Firma Xilinx. Dabei zeigte sich nicht nur eine deutliche Verringerung der benötigten Hardware bei der Verwendung von ausgedünnten Verbindungsnetzwerken, sondern auch eine spürbare Verringerung des kritischen Pfades. Daher ist ein spezialisiertes Netz mit minimaler Verbindungsstruktur stets zu bevorzugen. Allerdings erfordert dieser Entwurf deutlich höhere Entwicklungszeiten. Die Arbeit zeigt den manuellen Aufbau des Verbindungsnetzwerkes anhand von Datenflussgraphen, welche für die Zielapplikation erzeugt wurden.

Um die Funktionalität des Prozessors unter Beweis zu stellen, wurden einfache Applikationen ausgewählt, zu denen eine passende Konfiguration erstellt wurde. Programmierung und Registerallokation erfolgten per Hand, dabei erwies sich die Trennung von Transport- und Kontrolloperationen weniger hinderlich als anfangs angenommen. Es konnte sogar eine Verringerung des Speicherbedarfs um 27 % festgestellt werden. Ursache dafür ist im Wesentlichen eine Verkürzung der Transportoperationen durch den Wegfall der Kontrollinformation. Die zusätzlich hinzugekommenen Steuerinformationen je Funktionseinheit wiegen das aber bei weitem nicht auf.

Die Entwicklung einer eigenen Assemblersprache mit dazugehörigen Werkzeugen verkürzte den Validierungszyklus. So konnte beispielsweise die Berechnung der Fibonaccizahlen erfolgreich für eine speziell angepasste Prozessorinstanz parallelisiert werden.

Literatur

- [Cor98] Henk Corporaal. *Microprocessor architectures from VLIW to TTA*. Wiley, Chichester [u.a.], 1998.
- [Gre08] C. Gremzow. Quantitative global dataflow analysis on virtual instruction set simulators for hardware/software co-design. In *Computer Design, 2008. ICCD 2008. IEEE International Conference on*, Seiten 377–383, 2008.
- [Men03] Matthias Menge. *Zen-I, Prozessor mit kontrollflussgesteuertem Datenfluss / M. Menge. Technische Universität Berlin, Fakultät IV - Elektrotechnik und Informatik. [Hrsg.: Die Professoren der Fakultät IV - Elektrotechnik und Informatik der Technischen Universität Berlin]*. Leiter der Fachbibliothek Informatik, Sekretariat FR 5-4, Berlin, 2003.
- [MHG09] Nico Moser, Stefan Hauser und Carsten Gremzow. Reduzierung der Kommunikation in TTA-Verbindungsnetzen mittels Laufzeitanalyse. In *Methoden und Beschreibungssprachen zur Modellierung und Verifikation von Schaltungen und Systemen*, Seiten 107–116, 2009.

Entwurf und Realisierung eines Adapters für UniLoG zur Lastgenerierung an IP-basierten Schnittstellen

Martin Kulas

Universität Hamburg, Department Informatik
Arbeitsgruppe Telekommunikation und Rechnernetze (TKRN)
2kulas@informatik.uni-hamburg.de

Abstract: Rechnernetze sind allgegenwärtig. Man erwartet, dass die Netze so leistungsstark sind, dass sie die Anforderungen der Anwendungsdienste und ihrer Benutzer erfüllen. Ein Spezialwerkzeug zum Ermitteln der Leistungsfähigkeit der Netze sind Lastgeneratoren. Dieser Beitrag stellt einen echtzeitfähigen Lastgenerator für IP-Netze dar und zeigt seinen Einsatz im Rahmen einer umfangreichen Fallstudie für ein Video-on-Demand-Szenario in einem WLAN nach Standard IEEE 802.11g.

1 Einführung

Rechnernetze auf Basis des Internet-Protokolls IP sind weit verbreitet. Häufig bestehen diese Netze aus einem Konglomerat diverser Vernetzungstechniken. Beispielsweise unterscheiden sich Gigabit-Ethernet und WLAN nach IEEE 802.11 durch ihre Reichweite, Durchsatz und Bitfehlerwahrscheinlichkeit stark voneinander. Zur Komplexität heutiger Netze trägt zudem die Vielzahl verwendeter Anwendungsdienste bei. Die Dienste stellen dabei unterschiedlich hohe Anforderungen an Netze, um die Dienstgüte – insbesondere aus Benutzersicht – zu garantieren. Während z. B. IP-Telefonie und Online-Actionspiele niedrige Paketlaufzeiten fordern, ist beim Dateitransfer (z. B. über HTTP) ein möglichst hoher Durchsatz wichtig. Die große Heterogenität und Komplexität der zu erbringenden Dienste erschwert es, die Leistungsfähigkeit der Netze zu prognostizieren. In solchen Situationen helfen Lastgeneratoren, denn sie ahmen das Benutzerverhalten im Netz nach und produzieren dabei gewünschte Lastniveaus. Ein Netzbetreiber stellt mit Lastgeneratoren fest, ob ein Netz erwartete Lasten bewältigt, und ein Entwickler eines Dienstes überprüft, ob sein Netz unter bestimmten Hintergrundlasten benutzbar ist.

Die Arbeitsgruppe TKRN hat das Lastgeneratorkonzept UniLoG („Unified Load Generator“) entwickelt, welches zwischen der Lastspezifikation und der Lastgenerierung differenziert [CoW06]. Dabei ist die (Arbeits-)Last $L = L(E, S, IF, T)$ als eine Abfolge von Aufträgen definiert, welche eine Umgebung E einem Bediensystem S über eine Schnittstelle IF während eines Zeitintervalls T anbietet. Zur formalen Lastspezifikation verwendet man Benutzerverhaltensautomaten. Mittels dieser endlichen, erweiterten Automaten ist man in der Lage, realitätsnahe Lasten auf flexible und anschauliche Art an ausgewählten Schnittstellen zu modellieren. Dazu produzieren die R -Automatenzustände die Last-generierenden abstrakten Aufträge und die D -Zustände ermitteln die Zwischenan-

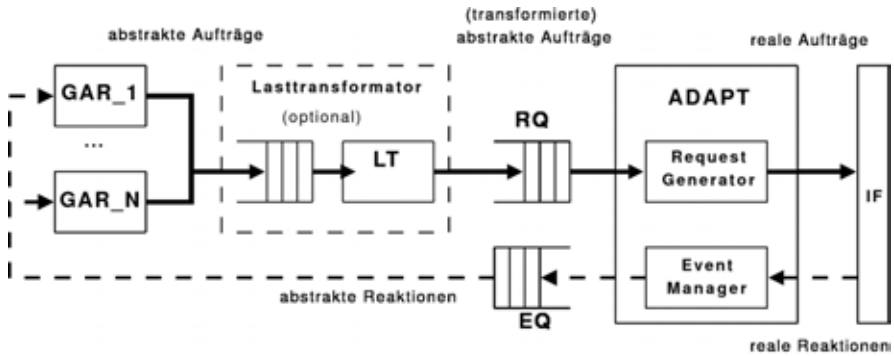


Abbildung 1: Aufbau des UniLoG-Lastgenerators RTLG

kunftszeiten dieser Aufträge. Für die Lastgenerierung ist eine echtzeitfähige Ausführungskomponente verantwortlich, welche Schnittstellen-spezifische Adapter benötigt [Kol08].

2 Der UniLoG-Adapter für IP-basierte Schnittstellen

Die Arbeitsgruppe TKRN besitzt einen UniLoG-konformen Lastgenerator namens RTLG (Real-Time Load Generator). Abbildung 1 zeigt seinen Aufbau. Er besteht u. a. aus einem Generator (GAR) abstrakter Aufträge, Warteschlangen für abstrakte Aufträge und Reaktionen (RQ und EQ) und einem Schnittstellen-spezifischen Adapter (ADAPT). Optional wandelt ein Lasttransformator (LT) eine Primärlast in eine Sekundärlast um [CoW06].

Die Aufgaben des IP-Adapters bestehen darin, (1) abstrakte Aufträge in Schnittstellenkonforme Aufträge umzuwandeln, (2) reale Aufträge an der IP-Schnittstelle zu vorgegebenen Zeitpunkten zu übergeben, (3) reale Reaktionen der IP-Schicht zu erfassen (ICMP-Nachrichten und Statusmeldungen des Betriebssystems) und (4) diese Reaktionen in abstrakte Reaktionen zu konvertieren. Für diese Aufgaben sieht der Entwurf des IP-Adapters die Komponenten Request-Generator und Event-Manager vor.

Die IP-Schicht bietet an ihrer Schnittstelle den Dienst an, ein IP-Paket in das Netz zu injizieren. An diesem Dienst orientiert sich der abstrakte Auftrag `InjectPacket`, welcher die Pflichtattribute `Frachtlänge` und `Zieladresse` enthält. Weitere einstellbare Auftragsattribute sind die `Quelladresse` des zu injizierenden Pakets, die `Art der Verkehrsklasse` oder die `maximale Anzahl der zu passierenden Router`.

Eine nicht-funktionale Anforderung an den IP-Adapter ist die Echtzeitfähigkeit. Dies hat zur Folge, dass (1) der Adapter die abstrakten Aufträge in reale Aufträge in Echtzeit konvertieren und (2) die realen Aufträge zu den richtigen Zeitpunkten an der Schnittstelle zur IP-Schicht übergeben muss. Die Erfüllung dieser Echtzeitanforderungen spielt für eine präzise Auftragsübergabe eine wichtige Rolle. Es ist wenig sinnvoll, einen Lastgenerator zu benutzen, welcher unfähig ist, eine spezifizierte Last exakt zu erzeugen. Denn ein Experimentator würde mit einem dazu unfähigen Lastgenerator falsche Schlüsse aus seinen



Abbildung 2: Das 250. ausgegebene Bild unter VBR-Lasten mit 1480 Byte langen IP-Frachten

Lastexperimenten ziehen [DBP07]. Damit der IP-Adapter und der RTLG die Echtzeitanforderungen erfüllen, sind beide in der Programmiersprache C++ realisiert.

3 Fallstudie

Die Fallstudie untersucht die Leistungsfähigkeit eines WLANs (IEEE 802.11g) für einen Video-on-Demand-Dienst (VoD-Dienst). Dazu empfängt ein Benutzer den Audio- und Videostrom eines qualitativ hochwertigen Films über das WLAN mittels RTSP/RTP/UDP (RFC 2326, RFC 3550). Neben dem Filmempfänger stellt der UniLoG-Lastgenerator mit IP-Adapter andere WLAN-Benutzer dar, indem er Hintergrundlasten im WLAN erzeugt. Es stellen sich u. a. die Fragen, ab welcher Hintergrundlast sich die Filmübertragung signifikant verschlechtert und ob die Art der Hintergrundlast eine Rolle spielt.

Das Fallstudienetz besteht aus einem voll-duplex Fast-Ethernet-Segment mit VoD-Server und Lastsenke und einem daran angeschlossenen Access-Point, in dessen Funkzelle der VoD-Client mit dem Lastgenerator um den Zugriff auf das Übertragungsmedium konkurriert. Die geringe Entfernung der aufgestellten WLAN-Stationen zueinander vermeidet ein „Hidden-Station“-Problem, weshalb der RTS/CTS-Mechanismus während der Experimente deaktiviert ist. Der Animationskurzfilm „Big Buck Bunny“ (H.264-Video in HDTV-Auflösung von 1280x720 Pixeln; MPEG-4-Audio) dient als Streamingmaterial [BBB]. Die verwendeten Hintergrundlasten sollen eine Engpassanalyse des WLANs erlauben. Es werden Hintergrundlasten mit konstantem (CBR) und variablem (VBR) Durchsatz auf der IP-Ebene benutzt, so dass ein mittlerer IP-Durchsatz von 1, 2, 4, 8, 16, 20, 24 oder 32 MBit/s entsteht. Dabei haben die injizierten IP-Pakete eine konstante Frachtlänge von 50 bzw. 1480 Byte sowohl für CBR- als auch für VBR-Hintergrundlasten. Die Zwischenankunftszeiten der IP-Pakete des VBR-Verkehrs sind exponentiell verteilt. Als Qualitätsmaße werden die Verzögerungsschwankung, der mittlere IP-Durchsatz auf Seiten des VoD-Clients, die Zahl der Sequenzfehler, der Paketduplikate und der verlorenen Pakete betrachtet.

Ein Erhöhen der Hintergrundlast verschlechterte wie erwartet die Dienstgüte. Im Hochlastbereich erhielt die VoD-Abspielsoftware zu wenig Daten, um Bilder und Ton korrekt wiederzugeben (siehe Abbildung 2). Des Weiteren wird festgestellt, dass Hintergrundverkehr mit kleinen Frachtlängen (50 Byte) bereits bei niedrigen Hintergrundlastniveaus (4 MBit/s) das WLAN in den Hochlastbereich treiben (siehe Abbildung 3). Dies ist durch die diversen Techniken der Bitübertragungs- und Datensicherungsschicht zur zuverlässigen Datenübertragung im WLAN begründet [Rec08].

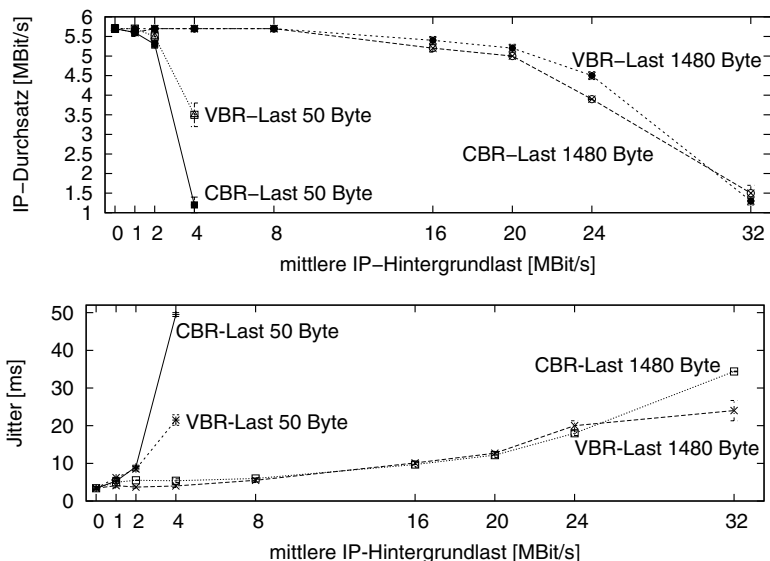


Abbildung 3: IP-Durchsatz beim VoD-Client und Jitter als exemplarische Qualitätsmaße

4 Resümee

Der IP-Lastgenerator bietet die Möglichkeit, flexibel Lasten zu modellieren, und erfüllt die Echtzeitanforderungen an die Lastgenerierung. In einer Fallstudie ist sein praktischer Nutzen demonstriert worden. Neben detaillierten Leistungs- und Dienstgüteuntersuchungen für Rechnernetze unter gezielt generiertem Hintergrundverkehr ist man mit IP-Lastgeneratoren beispielsweise auch in der Lage, verteilte Dienstverweigerungsangriffe (DDoS) kontrolliert durchzuführen und ihre Auswirkungen zu studieren.

Literatur

- [BBB] „Big Buck Bunny“ – ein frei erhältlicher computer-generierter Kurzfilm (Länge: 10 Minuten), <http://www.bigbuckbunny.org>, Online-Ressource, zuletzt besucht am 11.12.2009.
- [DBP07] Dainotti, A.; Botta, A.; Pescapè, A.: Do you know what you are generating?, Proceedings of the 2007 ACM CoNEXT conference, ACM, S. 1–2, New York, 2007.
- [CoW06] Cong, J.; Wolfinger, B. E.: A Unified Load Generation Approach Based On Formal Load Specification and Load Transformation, Value Tools 2006, First Intern. IEEE Conf. on Performance Evaluation Methodologies and Tools, Pisa, Oktober 2006.
- [Kol08] Kolesnikov, A.: Konzeption und Entwicklung eines echtzeitfähigen Lastgenerators für Multimedia-Verkehrsströme in IP-basierten Rechnernetzen, Echtzeit 2008.
- [Rec08] Rech, J.: Wireless LANs: 802.11-WLAN-Technologie und praktische Umsetzung im Detail, 3. Auflage, Heise Zeitschriften Verlag GmbH & Co. KG, Hannover, 2008.

Abstract der Diplomarbeit: Performant-Optimierung eines Echtzeit-Multimedia-Netzwerkes

Roman Messmer
FernUniversität Hagen
roman.meszmer@orf.at
Wien, 10.1.2010

1 Einleitung

Ziel der Arbeit war die Routingoptimierung eines existierenden glasfaserbasierten Echtzeit-Multimedia-Netzwerks, das in der Domäne Fernseh- und Hörfunk-Studioteknik seit kurzer Zeit eingesetzt wird. Es wurde ein hochperformanter Multicast-basierender, verteilter Routingalgorithmus entwickelt und evaluiert, mit dessen Hilfe eine optimale Leitungsauslastung und ein möglichst latenzzeitarmes Schalten von Netzwerkknoten erreicht werden kann. Dieser Algorithmus muss drei Vorbedingungen erfüllen:

1. Die möglichst optimale Auslastung der Verbindungsleitungen ist zu gewährleisten.
2. Die Latenzzeit für Schaltungsvorgänge muss begrenzt werden. Vorgabe: Latenz bis max. 150ms für den gesamten Schaltvorgang(psychologische Schranke).
3. Eine vorgegebene Gesamtverzögerung der geschalteten Leitungssignale muss eingehalten werden (Äquivalent zur Länge der größten Entfernung zwischen Quellknoten und Multicast-Zielen).

Im ersten Fall führt der Einsatz von Multicast-Technologie zu einer Optimierung der Ausnutzung der Kanäle. Der Minimierung der Latenzzeit der Schaltungsvorgänge kann man durch optimierten Einsatz der Befehlsverarbeitung und der Anwendung einer Kombination zwischen Offline- und Online-Algorithmus näher kommen. Dem dritten Aspekt begegnet man mit Algorithmen, die auf Basis eines sog. beschränkten Pfadroutings (Constraint Path Algorithm)[SICH05] und der Überwachung der Summe der Laufzeit der jeweiligen Verzögerungspfade arbeiten. Das Problem der Ermittlung eines Spannbaums mit minimalem Gewicht, eines sogenannten minimalen Steinerbaums, ist für allgemeine Graphen und beliebige Kantenkosten NP-vollständig[VERB08]. Neben Spezialfällen, die in polynomieller Zeit lösbar sind, existieren Approximationsalgorithmen als praktische Lösungsansätze. Ein klassisches Beispiel dafür ist die *MST-Steinerbaum-Heuristik*.

2 Multimedia-Multicast-Algorithmus (MulMic)

Als Vorbedingung für die Erfassung und Repräsentation des Netzwerk-Graphen wird der Umstand ausgenutzt, dass im Ausgangssystem bereits eine Kommunikationsschicht existiert. Diese arbeitet ähnlich wie das Multicast-Open-Shortest-Path-First (MOSPF) Protokoll [WILL07]. Hier wird in einem Netzwerk zwischen je zwei direkt verbundenen Netzwerkknoten eine Nachbarschaftsbeziehung aufgebaut. Die resultierende Topologieinformation wird dem zentralen Koordinatorknoten kommuniziert. Ergebnis der Abgleichvorgänge ist eine regelmäßig aktualisierte Strukturinformation, die direkt die Adjazenzmatrix liefert. Auf Basis dieser Information werden von den lokalen Knoten mit Hilfe des Floyd-Warshall-Algorithmus regelmäßig die kürzesten Pfade zwischen allen Knoten errechnet und in einer lokalen Routingtabelle mit ihren Kosten gespeichert. Wird nun ein Multicastschaltaufruf mit dem Quellknoten s und einer Menge $\{d_1, \dots, d_m\}$ mit $d_i \in M$ (M sei die Menge der Multicast-Knoten) gestartet, so werden aus der Routingtabelle die kürzesten Pfade zwischen dem lokalen Quellknoten und den Destinationsknoten ausgelesen.

1. Die Zielknotendistanzen werden nach ihrer Länge sortiert, um den kürzesten Pfad zwischen dem Quellknoten und dem topologisch am entferntesten gelegenen Zielknoten zu ermitteln.
2. Dieser Pfad wird als erster Teilgraph in den Multicastbaum aufgenommen.
Die weiteren Teilgraphen werden nun folgendermaßen ermittelt:
3. Sei i ein laufender Index von $1, \dots, |M|$ als Nummerierung für die Zielknoten.
Es wird geprüft, ob die Distanz des Knoten v_i zum Knoten v_{i-1} geringer ist als die Entfernung des Knoten v_{i-1} vom Quellknoten s .
4. Falls ersteres zutrifft, wird der kürzeste Pfad zwischen Knoten v_i und Knoten v_{i-1} in den Multicastpfad aufgenommen.
Ansonsten wird der Pfad zwischen Quellknoten s und Knoten v_{i-1} gewählt. Dazu ist es notwendig, die Distanz zwischen den einzelnen Destinationsknoten zu kennen. Diese werden der Routingtabelle entnommen. Ein auf einer Kante verlaufender Teilpfad ist nur einmal in den Multicastpfad aufzunehmen. Wenn ein Zielknoten bereits im Multicastpfad enthalten ist, wird er aus der Liste der noch zu bearbeitenden Zielknoten entfernt.
Zusätzlich wird geprüft, über wie viele Knoten das Signal bereits geschaltet wird, um unzulässige Signalverzögerungen verhindern zu können. Überschreitet diese Anzahl eine vorgegebene Grenze, so wird der nächste Knoten vom Quellknoten aus über den kürzesten Weg verbunden.
5. Der ermittelte Pfad muss auf Schleifen getestet werden. Dazu reicht die Ermittlung eines beliebigen spannenden Baumes mit den Zielknoten als Blättern über diesen Graphen.

Es resultiert ein nicht minimaler Spannbaum, dessen Summe der Teilpfadlängen zu den Blättern in allen Beispielfällen kleiner oder im Worst-Case maximal gleich groß ist, wie

sie bei einem Baum mit kürzesten Pfaden vom Quellknoten zu den einzelnen Destinationsknoten über direkte Pfade wäre. Der Vorteil gegenüber Algorithmen, die sämtliche kürzeste Wege mit Pfadlängenbeschränkung ermitteln, ein optimales Ergebnis berechnen und in dieser Kombination als einer multi-kriteriellen Optimierung einem NP-vollständigen Problem unterliegen, ist eine äußerst kurze Ausführungszeit und ein speziell in realistischer Umgebung eindrucksvoll optimiertes Ergebnis, wie sich in den Praxistests zeigt. Die Korrektheit des Algorithmus und die Laufzeit für den schaltungsrelevanten Teil werden in der Arbeit genau untersucht. Die mittlere Gesamtlaufzeit des Algorithmus, speziell bei einer größeren Zielknotenanzahl, liegt wegen der Ausnutzung des Ausschlusses während der Bearbeitung bereits gefundener Knoten deutlich unter dem ermittelten Wert von $O(|M| \cdot \log |M| + |V| + |E|)$. Wie Tests zeigen, läuft die Implementierung des Algorithmus umso effizienter, je mehr Zielknoten erreicht werden sollen.

3 Bandbreitenmanagement

Weiters wurde ein Vorschlag zur Erhöhung der Geschwindigkeit des Ermitteln verfügbarer Bandbreiten präsentiert. Aus der Praxis (Kreuzschienentechnik) ist bekannt, dass sendungsspezifisch nur eine begrenzte Anzahl von Signaltypen geschaltet wird. Diese Tatsache kann man insofern ausnützen, als für potenzielle Schaltaufträge nur eine sehr beschränkte Auswahl an zu schaltenden Bandbreiten in Frage kommen. In der Praxis sind Signalgruppen mit etwa fünf Bandbreitenklassen zu schalten. Der erstellte Graph führt in der Datenbank bereits die momentane Kapazitätsbelegung für jede Leitung und Leitungsrichtung extra mit. Diese Information kann für die Implementierung ausgenutzt werden.

4 Beurteilung und Ausblick

Die Musterimplementierung lässt erkennen, dass der Algorithmus schleifenfreie Multicastbäume ermittelt und sämtliche Zielknoten erreicht. In vielen Fällen liefert der Algorithmus ein nahezu optimales Routingergebnis. Nicht zusammenhängende Graphen oder Schleifenbildungen sind in keinem Fall aufgetreten und sind auch aufgrund der Ergebnisse der theoretischen Betrachtungen nicht möglich. Der Algorithmus läuft sogar im Worst Case deutlich schneller, als alle bisher für die gestellte Aufgabe entwickelten Heuristiken. Berücksichtigt man die kurze Laufzeit des Algorithmus für die Ermittlung von Multicastbäumen und den Fakt, dass eine Erhöhung von Multicastzielknoten auf Grund der Streichung bereits erfasster Zielknoten und Pfade eine verbesserte Effizienz des Algorithmus mit sich bringt, so kann davon ausgegangen werden, dass eine Implementierung des im Rahmen dieser Arbeit entwickelten Algorithmus in der Eingangs erwähnten Netzwerkumgebung oder verwandten Topologien wesentlich zur Beschleunigung der Berechnung von Multicast-Bäumen bzw. Punkt-zu-Multipunkt-Schaltungen beitragen kann.

Literatur

- [FRAN04] Franzl, G.: Path Selection Methods With Multiple Constraints in Service-Guaranteed WDM Networks
IEEE/ACM Transactions on Networking, IEEE Press, Piscataway, NJ, USA, Vol. 12, No. 1, February 2004
- [GERL96] Gerla M. et al.: Multicast protocols for high-speed, wormhole-routing local area networks
SIGCOMM '96: Conference proceedings on Applications, technologies, architectures, and protocols for computer communications
ACM, New York, N.Y., USA, 1996, pp 184-193
- [GUMM03] Gummadi, P. K.: An efficient primary-segmented backup scheme for dependable real-time communication in multihop networks
IEEE/ACM Transactions on Networking, IEEE Press, Piscataway, NJ, USA, Vol. 11, NO. 1, February 2003
- [JIA98] Jia, X. et al.: A Distributed Algorithm of Delay-Bounded Multicast Routing for Multimedia Applications in Wide Area Networks
IEEE/ACM Transactions on Networking, IEEE Press, Piscataway, NJ, USA, Vol. 6, No. 6, December 1998, pp 828-837
- [JUFR02] Jukan, A.; Franzl G.: Constraint-based path selection methods for on-demand provisioning in WDM networks
Proc. IEEE INFOCOM, June 2002, pp. 827-836
- [KOMP93] Kompella, V. P. et al.: Multicast Routing for Multimedia Communication
IEEE/ACM Transactions on Networking, IEEE Press, Piscataway, NJ, USA, Vol. 1, No. 3, June 1993, pp. 286-292
- [LI07] Li, Z.; Garcia-Luna-Aceves, J. J.: Finding multi-constrained feasible paths by using depth-first-search
Kluwer Academic Publishers, Hingham, MA, USA, Vol. 13, No. 3, June 2007, pp 323-334
- [OLIV05] Oliveira, C. A. S. et al.: Optimization Problems in Multicast Tree Construction
AT&T Labs Research Technical Report TD-6DTLXR, June 29, 2005
- [REEV00] Reeves D. S.; Salama H. F.: A distributed algorithm for delay-constrained unicast routing
IEEE/ACM Transactions on Networking, IEEE Press, Piscataway, NJ, USA, Vol. 8, No. 2, April 2000, pp 239-250
- [SHEN05] C. C. Shen; C. Jaikao: Ad Hoc Multicast Routing Algorithm with Swarm Intelligence
Mobile Networks and Applications, Kluwer Academic Publishers, Hingham, MA, USA, Vol. 10, No. 1-2, February 2005, pp. 47-59
- [SICH05] Sichalou, S.; L. Georgiadis: Algorithms for Precomputing Constrained Widest Paths and Multicast Trees
IEEE/ACM Transactions on Networking, IEEE Press, Piscataway, NJ, USA, Vol. 13, No. 5, October 2005, pp 1174-1187
- [VERB08] Verbeek, R.: Kurs Effiziente Graphenalgorithmen, FernUniversitat Hagen, WS 2008
- [WILL07] Williamson, B.: Developing IP Multicast Networks Volume I
Cisco Press, 2007

Service Brokering for Efficient Service Selection in Wireless Mesh Networks

Paveesuda Chaijiroj
paveenew@gmail.com

Supervisor: Prof. Dr. Dr. h.c. Otto Spaniol

Abstract: Usually, there is a large amount of services running inside a network. For example, file transfer, telnet, printing and scanning services. In order to help a user get a good quality and appropriate service, a reliability of those services needs to be indicated. Some criteria are needed and service brokering mechanism is a good answer of this problem. Our purpose is developed to manage and control all services before present them to a user's side. Moreover, it can evaluate some service information, such as, hopcount, Link Quality (LQ), Expected Transmission Count (ETX), Expected Transmission Time (ETT) which show service's quality and status at that moment. These considerable information are measured by service discovery unit inside our mesh node. Not only these factors can effect user's decision, but also some criteria which cannot be measure inside network, such as, physical distance should be considered.

1 Introduction

A Wireless Mesh Network (WMN) [AWW05] is a kind of wireless network which uses a mesh topology to perform a network. It composes of collections of mesh nodes. There are two main types of nodes depending on their functionality; mesh routers and mesh clients. Mesh routers have a static specification and form a mesh backbone and network gateways. A routing functionality is operated in this part. Mesh client nodes normally have independent movement. They can leave or enter a network at any time. Moreover, a non-routing mesh client needs neither good processing nor huge memory because it does not take care of routing mechanism and main functionality. A wireless mesh network has been improved from a mobile ad hoc network (MANET) [BCG05] which was developed for using in US military. Because of many advantages of MANETs; simplicity, ease of use and so on, a wireless mesh network brings these advantages and is added an infrastructure to be more robust. Then its backbone can be used for more complicated task and calculation. The architecture of a wireless mesh network is necessary to be self-organizing, self configuring and self-healing.

UMIC-Mesh Testbed [UMI09] is a hybrid wireless mesh network which composes of mesh backbone and mesh clients. 51 mesh routers are connected with a management server by network cable, in order to manage and control the whole network. In the testbed, a service discovery mechanism has already been provided. It collects all registered services

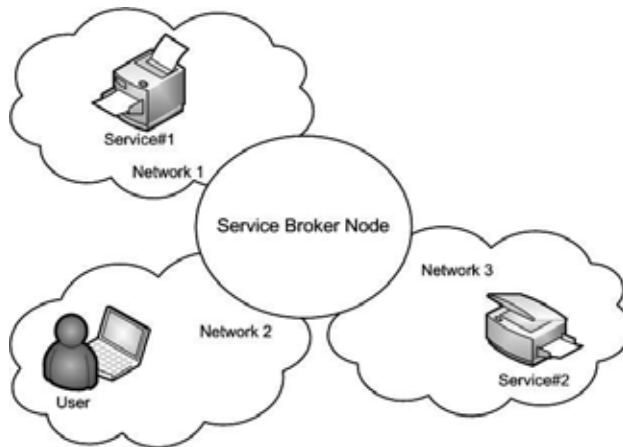


Figure 1: The service brokering model

in our wireless mesh network. However, there is no mechanism to reveal qualities of those links and the service ranking is still missing. A client has never known which service is good and should be selected. This paper proposes an approach to manage the services, evaluates the quality information and do the ranking for those services which have been provided in the network.

A mobile ad hoc network which implements a flat structure cannot guarantee performance when a network becomes larger. An idea of clustering can help to achieve scalability in an environment which has a great amount of nodes and high mobility. A clustering infrastructure divides mobile nodes into many groups of nodes. The close nodes perform as the same cluster. Clustering can be divided into many types depending on its functionality [YC05]. First, a clusterhead is a node which connects, manages all nodes in its cluster and handles data flow. Second, a clustergateway which is not clusterhead links to another node in different clusters. Thus its duty is to access a neighbor cluster. Last, a clustermember is a member in cluster which is not both a clusterhead and a clustergateway.

2 Service Brokering

In a wireless mesh network which has a dynamic environment, a client wants to access a service which has mobility characteristic. A client and service can possibly move from a network to another, or a link is unstable. Those make a service unreliable. Figure 1 describes a relation among service brokering, user (service client) and service providers [KV04]. The service brokering manages this transaction between a service and a client.

A Service Broker Node (SBN) [DSL04] is a middleware which efficiently manages all services and the queries. Not only service query translation is concerned, but also services registration translation is another approach of service broker node. First, a service query

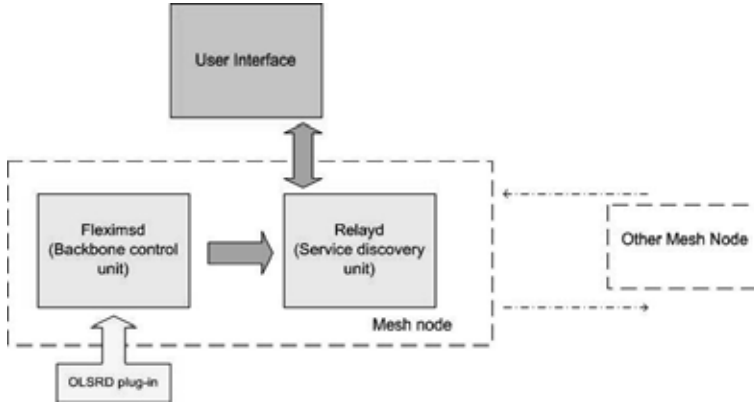


Figure 2: Service brokering backend architecture

is initiated by a user. Then a service broker node receives and sends it into the translation mechanism. The registered services from a network are collected and translated by a service broker node. When a query matches the services, those services are offered to the user to be selected afterward. That means the service selection has started.

3 Fleximspd and Relayd

The major structures of our network backbone consist of two parts; The first part is a backbone control unit or Fleximspd. The second part is service discovery unit or Relayd which is a daemon which provides a back-end process to manage the service information in a wireless network. The Fleximspd daemon is launched to set up a clustering infrastructure of a network. Furthermore, it connects with OLSRD plug-in where it discovers services and collects some link information from the network. It works as a backbone of the network and runs on every mesh router. It holds a neighbor table of each node. This is an important information which Relayd accesses to it. OLSRD is a routing daemon which plugs into Fleximspd. It implements the OLSR routing protocol for the network. Besides, it calculates a routing table and sends some packets to measure the routing metric between the nodes. In this approach, Relayd acts as a service brokering part.

4 Methodology

An OLSRD plug-in plugs into Fleximspd to feed datas from the network as shown in Figure 2. After OLSRD daemon has been launched, Fleximspd is started to set up the cluster structure. Then routing paths are created and a service discovery mechanism takes place. Fleximspd and Relayd run on our backbone mesh router, in order to collaborate. The Re-

layd daemon which works as a service brokering grabs some useful information from a neighbor table in Fleximsd. For example, names of services, service providers, ETX, ETT and so on. These useful information are used in the ranking calculation, in order to process the services and rank the top-quality of service for a user to select.

4.1 User's point-of-view

A user communicates with the user interface and requests for his required services by giving a keyword. The query is send to Relayd within our backend. A service brokering matches a user 's request services with the registered services and extracts the link information from Fleximsd. Finally, the results are sent back to a user. Therefore, he can see a list of his service with its criteria. At this point, he can see which service is suitable to be used. For example, the printing service number 1 has no task load at this moment and is located not far from his position.

5 Conclusion

A Relayd module can retrieve the information from fleximsd and display the services with their ranking results on a user interface. In the last step, we perform measurements and analyze them to clarify that how the result is after implement a service brokering compare with a wireless mesh network without service brokering. A user can consider and makes his own decision to select or not select the services. This method makes a selected service match user's requirement the most.

References

- [AWW05] Ian F. Akyildiz, Xudong Wang, and Weilin Wang. A Survey Wireless mesh networks. *Computer Networks*, pages S23–S30, September 2005.
- [BCG05] Raffaele Bruno, Marco Conti, and Enrico Gregori. Mesh Networks: Commodity Multi-hop Ad Hoc Networks. *Ad Hoc Networks*, pages 123–131, March 2005.
- [DSL04] Seema Degwekar, Stanley Y. W. Su, and Herman Lam. Constraint Specification and Processing in Web Services Publication and Discovery. *Web Services, IEEE International Conference on*, 0:210, 2004.
- [KV04] Teemu Koponen and Teemupekka Virtanen. A Service Discovery: A Service Broker Approach. *Hawaii International Conference on System Sciences*, 9:90284b, 2004.
- [UMI09] UMIC-Mesh.net - A hybrid wireless mesh network testbed, 2009. [Online] available: <https://www.umic-mesh.net/>; last accessed 7.11.2009.
- [YC05] Jane Y. Yu and Peter H.J. Chong. A Survey of Clustering Schemes for Mobile Ad Hoc Networks. *IEEE Communications Surveys and Tutorials*, 7, 2005.

Performance Analysis of Transmission Power Control in Wireless Mesh Networks

Watinee Sapkaew

RWTH Aachen University
Lehrstuhl für Informatik 4
chomphu_22@hotmail.com

Type of work: Master Thesis

Supervisor: Prof. Dr. Dr. h.c. Otto Spaniol

Abstract: Transmission power control (TPC) is a classical parameter to reduce interferences and the overall performance in wireless mesh networks. An intelligent transmission power control (TPC) has been widely investigated in simulations such as NS-2 but it is still unclear if simulation results can show the same performance gain in a real testbed. In this paper, we present our results based on a real mesh network testbed, namely U-MIC-Mesh.net, by varying the transmission power parameter and analyzing the resulting connectivity. The results are analyzed regarding 1-hop, 2-hop, and overall network connectivity of the network performance in wireless mesh networks by using Optimized Link State Routing Protocol (OLSR).

1 Introduction

Wireless Mesh Networks (WMNs) [1] consist of mesh routers and mesh clients who are built automatically from the idea of ad hoc networking. They are connected as mesh under the characteristics of self-organizing and self-configuring. WMNs provide redundancy by high connectivity but this may cause many problems such as interferences, contention, collisions when many neighbors try to transmit their data packets. WMNs consist of a static backbone of fixed mesh routers and a number of mobile routing or non-routing clients. Transmission power control (TPC) [3, 4] is a popular technique to enhance the network connectivity by varying a node's transmission power level. To reduce the interference during packet transmission, the node coverage interference is decreased which improves the sender/receiver-based interference by using a transmission power level control. This local adjustment has effects on the performance of the whole system. A too high power level may result in a too dense topology which causes interferences. A too low power level may result in a sparse and poorly connected topology, see Fig. 1.

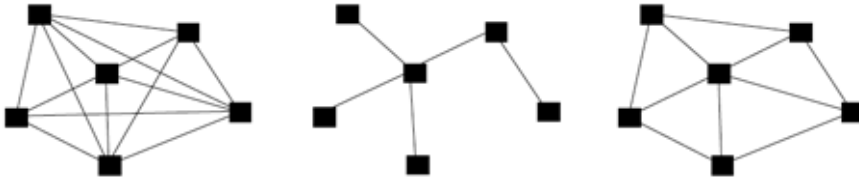


Figure 1: Network connectivity: Too high (left), too low (center), good (right)

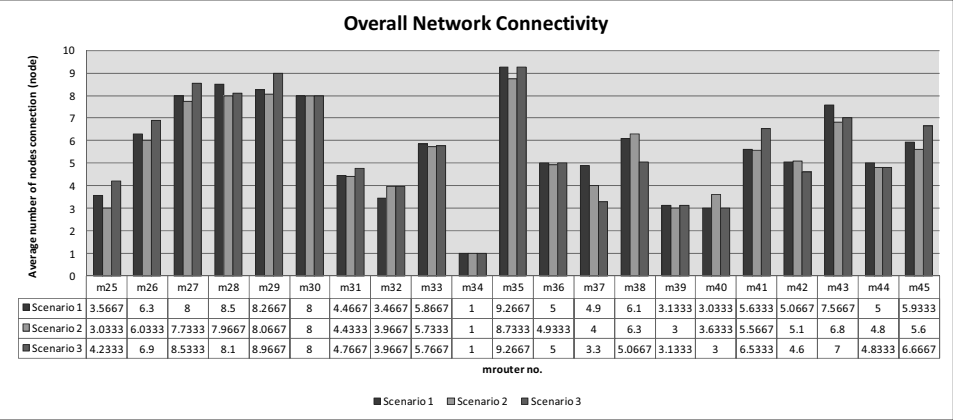


Figure 2: Measurement of overall network connectivity

To investigate the performance gain of TPC, the measurements in this paper are conducted in a real testbed. UMIC-Mesh.net Testbed [2] is a hybrid wireless mesh testbed which is under control of RWTH Aachen University, Germany

2 Measurements

An investigation has been carried out to find suitable values for the transmission power level by selecting the best value from 1-hop connectivity of the experiment, specifying the selected power level to each mesh node, testing the connectivity (routing algorithm), varying the power level, and comparing with the other result. The best results are compared to select the best value of transmission power level for each mesh node such as the bar of mrouter 25 is the best result at tx power 19 dBm then it is firstly indicated at 19 dBm as an example.

Several investigations have been adjusted to be a suitable value of transmission power level in each mesh node and enhanced the network connectivity. The chosen are based on 3 main scenarios which are the best results on network connectivity. Network performance is measured via OLSR among 21 mesh nodes (mrouter 25 – mrouter 45).

The figures show the different transmission power level which is assigned to each node starting from mrouter 25 to mrouter 45. It is resulted in 1-hop, 2-hop, and overall network connectivity. The worse cases are in mrouter 25, and 34 which have a little bit affected in research results. The highest power level is not required in mrouter 25, and 34 to increase network connectivity because of interference problem. Although the same power level is specified the same mesh node but the effect of overall network connectivity are not the same that depended on node surrounding. Figure 2 shows a comparison of the scenarios of overall network connectivity. It is demonstrated that a suitable value can achieve appropriate network connectivity in a real testbed. Therefore, the locations, node congestion, transmission power level of mesh node are considered in the results.

3 Conclusion

The main objective of this paper was to analyze the network performance of transmission power control in wireless mesh networks (WMNs). It is concluded that transmission power control (TPC) is on influence to limit the interference. Generally, the OLSR routing protocol can reduce the overhead of flooding in a wireless multihop network. The scope of this work is tuning of transmission power control (TPC) for improving network connectivity in a real wireless mesh testbed. It can be considered that these scenarios may enlarge the network connectivity but the interference still is the problem for a node.

References

- [1] Akyildiz, I.F. Xudong Wang : *A survey on wireless mesh networks*, Communications Magazine, IEEE, Vol.43, No. 9, pages 23-30, 2005.
- [2] Zimmermann, A. Schaffrath, D. Wenig, M. Hannemann, A. Gunes, M. Makram,: *Performance Evaluation of a Hybrid Testbed for Wireless Mesh Networks*, IEEE International Conference on Mobile Adhoc and Sensor Systems, 2007. MASS 2007., pages 1-10, October 2007
- [3] Kawadia, V. Kumar, P.R.: *Principles and protocols for power control in wireless ad hoc networks*, Selected Areas in Communications, IEEE Journal on , Vol. 23, Pages 76-88, January 2005
- [4] R. Ramanathan and R. Rosales-Hain., *Topology control of multihop wireless networks using transmit power adjustment*, vol.2, pp.404-413, INFOCOM 2000, March 2000.

Exploiting Object Orientation to Parallelize and Optimize C++ Applications

Christopher Schleiden, Christian Terboven

Rheinisch-Westfälische Technische Hochschule Aachen
{christopher.schleiden, christian.terboven}@rwth-aachen.de

1 Introduction and Motivation

In the last decades the High-Performance Computing community has mostly used procedural programming languages like C or FORTRAN. However, in recent years with projects and applications with increasing complexity many developers have switched to C++ in order to benefit from the advanced (object-oriented) language features needed by modern software development practices.

One of the problems in making this switch is that the currently established parallelization paradigms (e.g., OpenMP for Shared-Memory parallelization) and application programming interfaces (API) of parallel libraries have been designed and implemented to fit procedural concepts. Although the use of these interfaces in procedural languages may be easy and elegant, they regularly do not fit into the object-oriented nature of C++.

Another challenging aspect of the transition of high performance applications to C++ is the integration of hardware specific optimizations. In order to overcome the memory bandwidth limitations of bus-based SMP systems, shared memory computers frequently are cc-NUMA (cache-coherent non-uniform memory architectures) machines. In these types of machines each processor is equipped with local memory and thus the access speed depends on a processor's location relative to the position of the accessed memory location. In order to realize the full potential of these machines it is important to make data structures and calculations aware of these hardware characteristics.

2 Approach

In [TSaM09] a class library for linear algebra calculations, called *laperf*, has been designed and implemented in order to compare different parallelization paradigms in an object-oriented fashion. The library implements data types for (sparse) matrices, vectors, as well as operations like the inner product; a matrix vector product; or a vector triad using overloaded C++ operators.

The laperf library uses the concept of *expression templates* [Vel95] to translate an expression into a corresponding template expression which resembles an abstract syntax tree (see Figure 1). The compiler is then able to *unroll* this expression into a single `for`-loop which can be evaluated in parallel. The unique feature of laperf is that it provides a parallel version and at the same time hides all parallelization specific code from a user. This

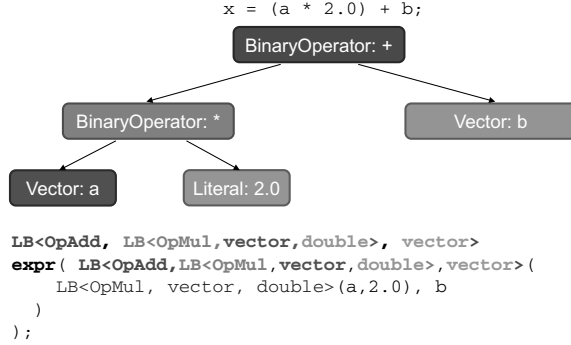


Figure 1: Expression Template mechanism: the C++ compiler converts the mathematical expression into a template expression resembling the corresponding AST.

way a programmer is able to write linear algebra expressions in a very straightforward and simple manner (see Figure 2) while still getting the performance of hand-crafted parallelized C code. As a use case to demonstrate and evaluate this library a generalized minimal residual (GMRES) [Saa03] linear algebra solver, taken from the application *DROPS* [GPRR02], was chosen. GMRES belongs to the class of Krylov subspace methods and

```

vector<double, OpenMPInternalParallelization> y,b,x;
matrix_crs<double> A; // matrix in CRS format
y = b + A * x; // matrix times vector

```

Figure 2: Usage of the laperf library.

as in other Krylov subspace methods the main computational operation is a sparse matrix times vector operation (SpMxV). In [WOV⁺07] it has been shown that the performance of the SpMxV kernel depends on the available memory bandwidth. Thus the challenge in parallelizing SpMxV on a cc-NUMA machine is the distribution of the sparse matrix rows across the available processors in a balanced way.

In this thesis different distribution strategies are developed and described: The *Distributed* policy distributes the rows of the matrix according to a specified OpenMP loop schedule (see [Boa08]). This distribution scheme is simple and easy to implement but does not take the structure of the sparse matrix (so-called *sparsity pattern*) into account thus resulting in a sometimes not optimal distribution of work across processors.

The *Chunked* policy uses a distribution scheme derived from the structure of the matrix:

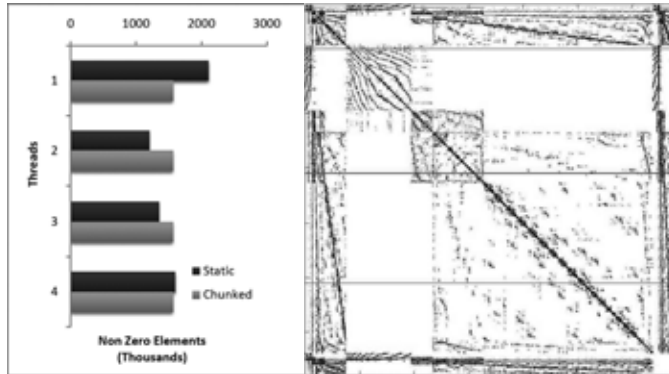


Figure 3: DROPS matrix with a dimension of 67385 and 6253281 nonzero elements distributed across 4 processors; horizontal lines mark the distribution according to the Chunked policy.

the rows are weighted and then distributed ensuring that every processor is assigned approximately the same amount of work (which corresponds to nonzero entries).

A comparison of a matrix distributed row-wise according to the Distributed policy with a *static* OpenMP schedule and according to the Chunked policy is shown in figure 2. Both of these distribution strategies are implemented as Policy Classes and can be passed to an *laperf* datatype during declaration.

3 Evaluation and Conclusion

In order to evaluate the performance of the policy based data types benchmarks were run on a number of machines at the Center for Computation and Communication of the RWTH Aachen University. The benchmark variants discussed below were measured on an IBM eServer LS24 equipped with four AMD Opteron 8356 quad-core processors with an individual clock rate of 2.3 GHz and a total of 32GB memory. Connected through the proprietary AMD interconnect called *HyperTransport* the system features cc-NUMA characteristics and employs four NUMA nodes.

Figure 3 shows the Speedup (serial divided by parallel runtime) of the GMRES kernel with a restart value of 20 and 100 completed iterations for the three available policies: Standard (no explicit data distribution), Chunked, and Distributed. With two threads all variants perform equally well with a speedup of about two. When using four threads, however, the Chunked memory policy still shows a near linear speedup of four while the other variants are slower. This trend continues up to 16 threads, where Chunked results in a speedup of nearly 12 while Standard just delivers approximately 3.5. As can be seen, with more active threads and processors, data distribution in general becomes more important. It can also be seen that the consideration of the sparsity pattern results in a better work distribution and higher performance.

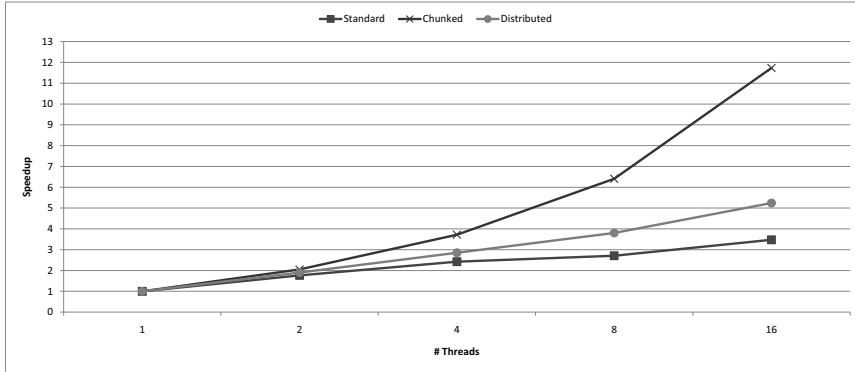


Figure 4: GMRES on AMD Barcelona, 100 iterations

In this thesis it has been shown that it is possible to hide code dealing with parallelization paradigms completely and effectively behind an object-oriented class interface, making it transparent to users. Well known and established C++ concepts have been adapted for use in a parallel environment. They have also been used to integrate the developed data distribution strategies into object-oriented data types in order to exploit cc-NUMA architectures. The benchmarking results show that despite some additional overhead due to the object-oriented encapsulation an efficient parallelization has been created.

References

- [Boa08] Architecture Review Board. OpenMP Application Program Interface, version 3.0, May 2008.
- [GPRR02] S. Gross, J. Peters, V. Reichelt, and A. Reusken. The DROPS Package for Numerical Simulations of Incompressible Flows Using Parallel Adaptive Multigrid Techniques. *IGPM-Report 211*, 2002.
- [Saa03] Y. Saad. *Iterative Methods for Sparse Linear Systems, Second Edition*. Society for Industrial and Applied Mathematics, April 2003.
- [TSaM09] C. Terboven, C. Schleiden, and D. an Mey. Comparing Programmability and Scalability of Multicore Parallelization Paradigms with C++. In *Second Workshop on Programmability Issues for Multi-Core Computers (MULTIPROG-2)*, January 2009.
- [Vel95] T. Veldhuizen. Expression Templates. *C++ Report*, 7(5):26–31, June 1995. Reprinted in C++ Gems, ed. Stanley Lippman.
- [WOV⁺07] S. Williams, L. Oliker, R. Vuduc, J. Shalf, K. Yelick, and J. Demmel. Optimization of Sparse Matrix-Vector Multiplication on Emerging Multicore Platforms. In *SC '07: Proceedings of the 2007 ACM/IEEE conference on Supercomputing*, pages 1–12, New York, NY, USA, 2007. ACM.

Konzeption und Umsetzung einer RIA zur untersuchungsbegleitenden Erfassung von RNFLT-Scans und Untersuchung von Klassifikatoren für die diagnostische Unterstützung bei neurodegenerativen Erkrankungen am Beispiel der Multiplen Sklerose.

Sebastian Bischoff, Mediber GmbH, Berlin

Fachhochschule Brandenburg
Informatik in den Lebenswissenschaften
bischoffs@freenet.de

Art der Arbeit: Diplomarbeit

Betreuer/in der Arbeit: Dipl. Inform. Ingo Boersch, M. Sc. Sebastian Mansow-Model

Zielsetzung: Erfassung von Retinal Nerve Fiber Layer Thickness (RNFLT) Scans für eine medizinische Studie durch eine Rich Internet Applikation (RIA), sowie die Erweiterung der Auswertung dieser Messungen von der etablierten, reinen Mittelwertberechnung, zu einer Auswertung durch Kurvenanalyse und Verfahren des maschinellen Lernens, zum Ziel der Generierung eines Klassifikators zur automatischen Klassifikation von Messungen.

Hintergrund: Multiple Sklerose (MS) ist eine neurodegenerative Erkrankung, die das Zentrale Nervensystem (ZNS) befällt. Die Diagnose dieser Erkrankung ist schwierig und teuer. Der Augennerv gehört zum ZNS und die Vermessung der Nervenfaserschicht des Auges (RNFLT) mittels optischer Kohärenztomographie, ist einfach und preiswert. Man vermutet, dass die Abnahme der Nervenfaserschichtdicke im Auge einen direkten Rückschluss auf eine neurodegenerative Erkrankung zulässt. So konnte an Hand des arithmetischen Mittelwertes der RNFLT-Messung die Erkrankung an MS mit einer Sensitivität von 19,4 % bei einer Spezifität von 95 % in einer Studie nachgewiesen werden (vgl. [BBD+09] und [FFF+08]). Die RNFLT-Messung wird bereits erfolgreich zur Diagnose von verschiedenen Augenerkrankungen genutzt, weshalb sie an vielen Kliniken verfügbar ist.

Methoden: Die Erfassung der Messungen erfolgt mit einer Silverlight Website oder einem C# .NET Programm, welche eine manuelle und eine halbautomatische Eingabe ermöglichen. Zur Auswertung der 523 Messungen mit ihren 199 selbst generierten Attributen wurde RapidMiner verwendet. Bei der Auswertung wurden der Naive Bayes Algorithmus und eine Support Vector Machine (SVM) mit linearem und polynomialem Kernel untersucht. Ebenfalls wurden verschiedene Auswahlen der Attribute getestet.

Ergebnisse: Es entstand ein halbautomatisches Eingabeprogramm, welches manuell exportierte Datensätze automatisch erfassen kann. Außerdem entstand eine RIA zum Anzeigen und Verwalten dieser Messungen und der Auswertungen. Es wurden neue Attribute durch Verfahren der Kurvendiskussion, Bildverarbeitung, Stochastik und durch Expertenwissen generiert. Die Auswertung mit einer Auswahl der Attribute, welche nur Mittelwerte beinhaltet ergab bei der SVM mit linearem Kernel eine Sensitivität von 32 % bei einer Spezifität von 98 %. Bei einer anderen Auswahl, die Attribute enthält, die an Hand der Bewertung der Relevanz durch den Chi-Quadrat-Test ausgewählt wurden, lag das Ergebnis bei 43 % Sensitivität bei einer Spezifität von 98 %.

Diskussion: Das halbautomatische Eingabeprogramm sowie die Anzeige- und Verwaltungswebsite sind in einer Probestellung bereits im Einsatz, um die Messungen in einer Studie zu erfassen. Das Auswertungsverfahren, welches bisher nur die Mittelwerte der Messungen verglichen hat, konnte bereits alleine durch die Verwendung neuer Attribute um 34,4 % verbessert werden. Durch die Verwendung neuer Attribute, der verschiedenen Lernalgorithmen und fortschreitender Verbesserung der Messgeräte, kann das Auswerten der RNFLT-Messungen in Zukunft die Diagnose der MS ergänzen.

Quellen:

- [BBD+09] Bock, Markus ; Brandt, Alexander U. ; Dörr, Jan ; Pfueller, Caspar F. ; Ohlraun, Stephanie ; Zipp, Frauke ; Paul, Friedemann: *Time Domain and Spectral Domain Optical Coherence Tomography in Multiple Sclerosis : a Comparative Cross Sectional Study*. 2009. - Unveröffentlicht
- [FFF+08] Frohman, Elliot M. ; Fujimoto, James G. ; Frohman, Teresa C. ; Calabresi, Peter A. ; Cutter, Gary ; Balcer, Laura J.: Optical coherence tomography: a window into the mechanisms of multiple sclerosis. In: *nature clinical practice NEUROLOGY* 4 (2008), S. 12

A Skeletal Approach for Parallel Medical Imaging

Mathias Brink

mathias.brink@wwu.de

1 Introduction

Parallel programming is becoming more and more important in application development. Especially time-consuming and time-critical applications employ parallel program paradigms. Developers of parallel programs have to handle several difficulties like synchronization and communication of the processing units. The choice of an appropriate and efficient paradigm is another problem.

Medical imaging is a computationally intensive field of application, because huge amounts of data have to be processed. Positron Emission Tomography (PET) is a non-invasive needed imaging technology to visualize metabolic processes in a – human or animal – body. A radioactive substance (tracer) is applied to the body. The tracer accumulates within certain tissues and decays. During decay positrons are emitted by the tracer and finally a pair of γ -rays is emitted for every positron. The body is placed within a scanner which records these pairs of γ -rays, so-called “events”. Usually between 10^7 and $5 \cdot 10^8$ events are recorded during a single investigation. A reconstruction algorithm is used to compute a 3D image of the tracer’s distribution from the recorded events.

2 Motivation

The research group “Parallel and Distributed Systems” at the WWU Münster developed parallel implementations of algorithms for medical imaging. These implementations are designed for many architectures (Multicore, Cluster [HSGL08], GPU [SVGM08]) using several parallel programming paradigms (*OpenMP*, *Threading Building Blocks*, *CUDA*, ...). It is usually necessary to take into account the memory hierarchy for an efficient memory management, such that different paradigms are necessary for different architectures to achieve a high performance. Some architectures require even more than one paradigm (e.g., for cluster computers, MPI and OpenMP are used together [Qui04]). Each algorithm’s implementation had to be developed from scratch by choosing suitable program paradigms, i.e. there is no code re-use.

```

for each (iteration  $k$ ) {
  for each (subset  $l$ ) {
    for each(event in subset  $l$ ) {
      computeTracersDistribution()
      computeCorrectionImage() }
    updateImage() } }

```

Listing 1: List-Mode OSEM algorithm

2.1 List-Mode OSEM: a PET algorithm

List-Mode Ordered Subset Expectation Maximization (List-Mode OSEM) is an iterative PET reconstruction algorithm [REFO98] that calculates a 3D image. First the tracer’s distribution is computed by Siddon’s algorithm [Sid85]. Then a correction image is calculated that is voxel-wise multiplied to the image. The entire algorithm is shown in Listing 1. Customary computers take several hours to execute the List-Mode OSEM algorithm. The algorithm can be performed in parallel to accelerate the reconstruction .

2.2 Skeleton Libraries

The goal of a skeleton-based approach is to abstract from parallel program paradigms and the underlying architecture. Skeletons are higher-order functions that implement a concrete parallel communication or calculation pattern and are provided to the user as ready-to-use program constructs. Users do not have to specify parallelism or handle side effects like race conditions. Task- and data-parallel Skeletons [Col89] are collected in so-called “Skeleton Libraries”. Examples for Skeleton Libraries are *DatTel* [Bis05], *Müsli* [CPK09] and *eSkel* [Col04].

These are often restricted to one or two parallel program paradigms so they can not be used on any arbitrary architecture. All these libraries provide generic types but no data structures for e.g. 3D images. The Medical Imaging programmer is forced to map the needed on available structures. A toolkit that provides a higher abstraction level is essential.

3 Objectives

A domain-specific skeleton library for medical imaging is created to support a simple and efficient program development in the area of medical imaging. In particular, it should provide several skeletons to describe calculations on multidimensional data structures. *SkIRT* (Skeletal Image Reconstruction Toolkit) is to the best of our knowledge a first approach to implement such a library in C++.

SkIRT provides interfaces for skeletons and data structures, needed to implement the List-

Mode OSEM algorithm. A prototypic implementation was developed to study whether the interfaces are implementable and to estimate the implementation's costs. The aimed architecture is the *AMD-64* architecture that can be found in off-the-shelf CPU's. *POSIX Threads* were used to implement the parallel patterns. The library is based on a generic vector data structure which can be configured for an arbitrary multi-dimensional address scheme (e.g., a 3-dimensional address scheme for images). Several skeletons are provided in *SkIRT* to manipulate such structures, in particular, three well-known skeletons:

map A unary function is applied to all elements of an input,

zip Combination of two structures of the same dimension using a binary function,

reduce Reduces an input by one dimension using a binary, associative function.

Additionally, data structures for concrete dimensions are included in *SkIRT*:

Vector (1D), Matrix (2D), Image (3D)

4 Experimental results

The List-Mode OSEM algorithm was implemented using *SkIRT* to compare the performance with other implementations. This evaluation already shows acceptable results. Table 1 shows the comparison to an *OpenMP*-based implementation on a quad-core *i7* CPU using Hyperthreading. The calculation is slowed down by the factor 1,5 but both implementations scale similarly well.

# Threads	Runtime <i>SkIRT</i>	Speedup	Runtime <i>Open-MP</i>	Speedup
sequential	143,9	–	89,1	–
2	74,9	1,92	47,0	1,90
4	42,1	3,42	36,3	2,45
8	35,4	4,06	22,3	3,99
16	34,5	4,17	22,3	3,99

Table 1: Runtime and speedup on a quad-core *i7* CPU using Hyperthreading

5 Conclusion and future work

Using a modern real-world algorithm for medical imaging, a prototype of a domain-specific high-level library containing algorithmic skeletons and data types for medical imaging was created.

SkIRT is not a fully elaborated skeleton library. Further algorithms for image processing will be studied to identify additional skeletons.

References

- [Bis05] H. Bischof. *Systematic Development of Parallel Programs Using Skeletons*. Berichte aus der Informatik. Shaker Verlag, 2005. ISBN 3-8322-4112-4.
- [Col89] M. Cole. *Algorithmic Skeletons: Structured Management of Parallel Computation*. MIT Press, 1989.
- [Col04] M. Cole. Bringing skeletons out of the closet: a pragmatic manifesto for skeletal parallel programming. *Parallel Computing*, 30, 2004.
- [CPK09] P. Ciechanowicz, M. Poldner, and H. Kuchen. The Münster Skeleton Library Muesli — A Comprehensive Overview. In Jörg Becker et al., editors, *Working Paper No. 7*. Working Papers, European European Research Center for Information Systems, 2009.
- [HSGL08] T. Hoefler, M. Schellmann, S. Gorlatch, and A. Lumsdaine. Communication Optimization for Medical Image Reconstruction Algorithms. In *Lecture Notes in Computer Science*, volume 5205, pages 75–83, Berlin/Heidelberg, 2008. Springer.
- [Qui04] M. J. Quinn. *Parallel Programming in C with MPI and OpenMP*. McGraw Hill Higher Education, 2004.
- [REFO98] A. J. Reader, K. Erlandsson, M. A. Flower, and R. J. Ott. Fast accurate iterative reconstruction for low-statistics positron volume imaging. *Physics in Medicine and Biology*, 43(4):823–834, April 1998.
- [Sid85] R. L. Siddon. Fast calculation of the exact radiological path for a three-dimensional CT array. *Medical Physics*, 12(2):252–255, March 1985.
- [SVGM08] M. Schellmann, J. Vörding, S. Gorlatch, and D. Meiländer. Cost-effective medical image reconstruction: from clusters to graphics processing units. In *CF '08: Proceedings of the 2008 conference on Computing Frontiers*, pages 283–292, New York, NY, USA, 2008. ACM.

Detektion von Strukturen in Fluoreszenzmikroskopiebildern

Markus Glaß

Martin-Luther-Universität Halle-Wittenberg, Institut für Informatik
markus.glass@student.uni-halle.de

Abstract: Fluorescence microscopy offers an important tool for visualizing and analyzing biological cells and their components. In this work an approach for detecting cells in 2D microscopy images is presented. The developed segmentation algorithm uses level set methods. Based on the well known Chan-Vese segmentation model a multi phase level set approach was developed and implemented that additionally preserves the topology of the detected objects.

Keywords: cell segmentation, level set, multi phase, topology preserving

1 Einleitung

Die Fluoreszenzmikroskopie bietet ein wichtiges Hilfsmittel zur Darstellung und Analyse von biologischen Zellen und deren Bestandteilen. Durch eine vollautomatische Analyse des gewonnenen Bildmaterials mit Methoden der digitalen Bildverarbeitung können objektive und reproduzierbare Ergebnisse erzielt werden. Um statistische Aussagen zur Verteilung und Lokalisation von fluoreszenzmarkierten Strukturen innerhalb von Zellen treffen zu können ist es nötig, die Zellgrenzen innerhalb derer sich diese Strukturen befinden, zu bestimmen. In der vorgestellten Diplomarbeit wurde ein Verfahren, basierend auf Level-Set-Methoden und dem Zwei-Phasen-Segmentierungsmodell von Chan und Vese [CV01] entwickelt, welches eine topologierhaltende Segmentierung einer beliebigen Anzahl von Zellen ermöglicht. Zusätzlich wurde ein Verfahren zur Segmentierung von Bilddaten für die Anwendung auf dreidimensionale Bilder erweitert und implementiert. Es wurde angewendet um zytoplasmatische granuläre Strukturen in diesen Bildern zu detektieren. Als Alternative zur Verwendung von Level-Sets können auch sog. Snakes [KT88] verwendet werden. Jedoch muss dann sichergestellt werden, dass es nicht zu Überlappungen der einzelnen Phasen kommen kann. In [ZOM05] wird eine Methode vorgestellt, die diesen Ansatz verfolgt.

2 Topologieerhaltende Multiphasen-Segmentierung

Multiphasen-Level-Sets Da sich bei Verwendung von Level-Set-Methoden die Topologie von Objekten normalerweise verändern kann, ist es möglich eine beliebige Anzahl

von Objekten zu erkennen, ohne dass diese Anzahl vor der Segmentierung bekannt sein muss. Eine Einteilung des Eingabebildes lediglich in die zwei Phasen Vorder- und Hintergrund kann zu schlechten Segmentierungsergebnissen führen, wenn sich die zu erkennenden Objekte in ihren Bildeigenschaften signifikant unterscheiden, z.B. wenn fluoreszenzmarkierte Objekte in einem Bild deutlich unterschiedliche Intensitätswerte besitzen. Eine Möglichkeit dies zu umgehen ist es, das Bild in mehr als nur zwei Phasen zu unterteilen. So können spezifische Eigenschaften einzelner Objekte oder Objektgruppen berücksichtigt werden. Level-Set-Methoden, die eine solche Segmentierung durchführen können, werden üblicherweise als Multiphasen-Level-Set-Methoden bezeichnet.

Zum Repräsentieren von n verschiedenen Phasen existieren unterschiedliche Ansätze. In der vorgestellten Arbeit wird der Ansatz von [LLT03] auf die direkte Minimierung des Energiefunktionales (1) nach [SC02] übertragen. Hierbei wird genau eine Level-Set-Funktion $\phi(p) : \Omega \mapsto \{1, \dots, n\}$ verwendet, die jedem Bildpunkt p aus dem Bildraum Ω eine der n Phasen zuordnet. Überlappungen und unsegmentierte Bereiche werden von diesem Ansatz von vornherein ausgeschlossen und es wird lediglich Speicherplatz für eine Level-Set-Funktion benötigt.

Das zu minimierende Energiefunktional ist dabei von dem Zwei-Phasen-Modell von Chan und Vese [CV01] abgeleitet. Es hat folgende Form:

$$E(\phi(p), \vec{c}) = \sum_{i=1}^n \lambda_i \int_{\Omega} |I(p) - c_i|^2 H_i(\phi(p)) dp + \alpha \sum_{i=1}^{n-1} \int_{\Omega} g(I(p)) K_i(\phi(p)) dp + \mu \sum_{i=1}^{n-1} \int_{\Omega} |\nabla H_i(\phi(p))| dp \quad (1)$$

Die Multiphasen-Heaviside-Funktion $H_i(\phi)$ ist definiert als

$$H_i(\phi(p)) := \begin{cases} 1 & : \phi(p) = i \\ 0 & : \phi(p) \neq i \end{cases}$$

Der erste Term in (1) modelliert die Annahme, dass jede Phase einen annähernd konstanten Grauwert c_i hat und bestraft die quadratischen Abweichungen davon. Der zweite Term soll die Ausbreitung der Phasen in Richtung von Bildkanten lenken. Die Funktion $g(I)$ bewertet Intensitätsunterschiede, sie wirkt also als Kantendetektor.

$$g(I(p)) := \frac{1}{1 + |\nabla(I(p) * G_{\sigma})|}$$

Das Bild wird zuerst mit einem Gaußfilter (G_{σ}) geglättet und vom Ergebnis dieser Glättung wird der Gradientenbetrag bestimmt. Die Funktion ist so gestaltet, dass ihr Funktionswert klein wird, wenn das Bild am Punkt p einen großen Gradientenbetrag aufweist. Im angegebenen Modell wird sie nur an den Konturen $C_i, i = \{1, \dots, n-1\}$ der Objektphasen ausgewertet (Phase n ist Hintergrund). Dafür wurde die Funktion $K_i(\phi)$ entworfen

$$K_i(\phi(p)) := \begin{cases} 1 & : p \in C_i \\ 0 & : p \notin C_i \end{cases}$$

Der dritte Term bestraft die Längen der Konturen der entstehenden Phasen. Dadurch können relativ glatte Objektgrenzen erzielt werden. Die Parameter λ_i, α und μ bestimmen, mit

welchem Gewicht die einzelnen Terme in die Gesamtenergie einfließen.

Die Optimierung wird in Erweiterung des Verfahrens von Song und Chan [SC02] durchgeführt. Dieser Algorithmus verzichtet auf das, für Level-Set-Verfahren übliche, Lösen von Differentialgleichungssystemen. Stattdessen wird das Energiefunktional direkt minimiert, indem iterativ für jedes Pixel geprüft wird, ob sich die Gesamtenergie verringert, wenn dieses Pixel einer anderen Phase zugeordnet würde.

Topologieerhaltung Die oft vorteilhafte Eigenschaft von Level-Set-Methoden, topologische Veränderungen der Phasen durchzuführen, kann in manchen Situationen unerwünscht sein. Eine solche Situation tritt zum Beispiel auf, wenn die Anzahl der zu segmentierenden Objekte vorher bekannt ist. Im Folgenden soll ein Ansatz vorgestellt werden, mit dem durch Einführung zusätzlicher Bedingungen dennoch eine topologieerhaltende Multiphasen-Segmentierung mittels Level-Set-Verfahren auf 2D-Bildern durchgeführt werden kann. Ein weiterer Ansatz zu topologieerhaltender Segmentierung mit Level-Set-Methoden findet sich in [HXP03].

Ausgehend von einer initialen Segmentierung des betrachteten Bildes werden die einzelnen Phasen durch Minimierung eines Energiefunktionals iterativ entwickelt. Damit die Topologie der initialen Segmentierung erhalten bleibt, wird gefordert, dass jede Objektphase aus einer Zusammenhangskomponente besteht. Die Anzahl n der Zusammenhangskomponenten bzw. Objekte, darf sich in den folgenden Iterationen nicht ändern. Zusätzlich zu den n Objektphasen gibt es eine Hintergrundphase, die aber aus verschiedenen Zusammenhangskomponenten bestehen kann. Damit sich das während der Entwicklung der Phasen nicht ändert, darf ein Bildpunkt p einer Phase r nur dann zu einer anderen Phase s flippen, wenn

1. p benachbart zu einem Bildpunkt aus s ist
2. die Zusammenhangskomponente von r durch das Flippen nicht gespalten werden kann, falls r eine Objektphase ist
3. das Flippen nicht zu einem Einschluss der Phase r in Phase s führen kann.

Die erste Bedingung impliziert, dass nur Bildpunkte, die auf der Kontur einer Phase liegen, überhaupt für ein Flippen in Frage kommen. Deshalb reicht es aus, in jeder Iteration nur diese Punkte zu betrachten. Die Konturpunkte werden nacheinander bearbeitet. Wenn ein Pixel p zu einer Objektphase r gehört, dann wird zuerst überprüft, ob ein Flippen von p die Zusammenhangskomponente von r aufspalten könnte. Das kann passieren, wenn p die einzige Verbindung zwischen zwei 4-Nachbarn ist, die ebenfalls zu Phase r gehören. Um zu überprüfen, ob so eine Nachbarschaftskonfiguration vorliegt, wird die 8-Nachbarschaft von p betrachtet. Wenn darin mehr als ein Wechsel von einem 4-Nachbarn von p oder einem, mit einem 4-Nachbarn von p verbundenen Pixel, von Phase r in eine andere Phase existiert, darf p nicht geflippt werden. Sollte dies nicht der Fall sein, wird als nächstes für alle Phasen $s \neq r$, aus denen 4-Nachbarn von p stammen, geprüft, ob es in diesen Phasen zu einem Einschluss der Phase r kommen könnte, wenn p zu der entsprechenden Phase flippen würde. Dies geschieht, indem in der 8-Nachbarschaft von p für jede in Frage kommende Phase s geprüft wird, ob mehr als ein Wechsel von verbundenen Pixeln der Phase s in eine andere Phase existiert. Alle Phasen für die dies nicht zutrifft werden vermerkt.

Wenn p zum Hintergrund gehört, wird nur die zweite Bedingung geprüft, da die Hintergrundphase aus mehreren Zusammenhangskomponenten bestehen darf. Am Ende wird für alle vorgemerkten Phasen berechnet, ob die Gesamtenergie abnehmen würde, wenn p zu der entsprechenden Phase geflippt würde. Sollte dies für eine oder mehrere Phasen der Fall sein, wird p gemäß der maximalen Energieverringering geflippt.

3 Ergebnisse

Als Eingabe dienten Mehrkanalbilder, die mit einem Epifluoreszenzmikroskop erstellt wurden. Zuerst wurden im DAPI-Kanal die Zellkerne detektiert. Diese wurden dann zur Initialisierung der Level-Set-Funktion verwendet. Da kein Farbstoff zur Verfügung stand, der spezifisch Zellmembranen färben kann, wurde die eigentliche Zelldetektion auf einem Kanal durchgeführt, in dem ein Protein angefärbt war, das im gesamten Zellplasma vorkam. Abbildung 1 zeigt ein Beispiel für die Zellsegmentierung. Zellen mit mehreren Zellkernen resultieren aus einer nicht optimalen Initialisierung, bei der nah beieinander liegende Zellkerne zu einem Objekt verschmolzen wurden. Die Ergebnisse müssen noch in einigen Bereichen verbessert werden, bilden aber bereits so eine vielversprechende Basis für biologische Analysen.

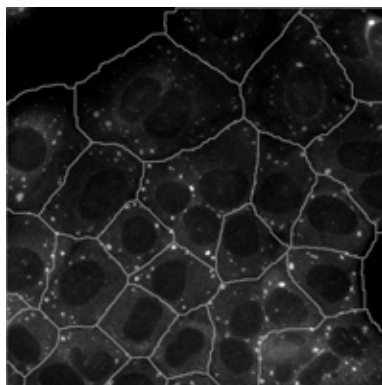


Abbildung 1: Beispiel für Zellsegmentierung; die Konturen der detektierten Zellen sind dem Eingabebild überlagert

Literatur

- [CV01] T. F. Chan und L. A. Vese. Active contours without edges. *Image Processing, IEEE Transactions on*, 10(2):266–277, 2001.
- [HXP03] Xiao Han, Chenyang Xu und Jerry L. Prince. A Topology Preserving Level Set Method for Geometric Deformable Models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25:755–768, 2003.
- [KT88] A. Kass, M. and Witkin und D. Terzopoulos. Snakes: Active contour models. *International Journal of Computer Vision*, 1(4):321–331, 1988.
- [LLT03] J. Lie, M. Lysaker und X. Tai. A Variant Of The Level Set Method And Applications To Image Segmentation. Bericht, Math. Comp, 2003.
- [SC02] B. Song und T. F. Chan. A Fast Algorithm for Level Set Based Optimization. *CAM-UCLA*, 68:02–68, 2002.
- [ZOM05] C. Zimmer und J. C. Olivo-Marin. Coupled parametric active contours. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 27(11):1838–1842, 2005.

Automatic Patient Pose Estimation

Robert Grimm

robert.grimm@informatik.stud.uni-erlangen.de

Abstract: Medical image acquisition using Magnetic Resonance Imaging scanners requires detailed knowledge of the patient's position. Only then, optimal measurement parameters can be configured. Today, the individual information is entered manually into the system by the examining radiologist. We propose methods to automatically detect both posture and detailed pose of a lying patient. A k-nearest neighbour classifier directly estimates the general posture. The algorithm to compute the detailed pose is model-based and relies on Powell's optimisation scheme. The sensor data are provided by a time-of-flight range camera acquiring a depth map of the scene. Thereby, the measurement setup is completely markerless.

1 Motivation

Today, imaging modalities such as Magnetic Resonance (MR) scanners are indispensable diagnostic tools in clinical routine. The technology is evolving very fast and the increasing image quality allows ever more applications. However, exactly because of this fast evolution, hospitals need to replace their systems every few years in order to provide up-to-date imaging facilities. Coupled with the complexity of such devices, acquisition and maintenance costs are very high. To compensate for these expenses, an efficient workflow is crucial such that as many patients as possible can be examined every day.

Although much progress has been made to reduce image acquisition times, each examination still requires the setting of individual measurement parameters. For example, the region of interest for image acquisition is defined manually. The patient's body height also needs to be entered, along with the position on the table.

This work is tedious and time-consuming. Our vision is a system that supports the radiologist by automatically classifying the posture and by computing the position of certain landmarks on the patient body prior to the MR scan. This information can then be used to estimate the location of the individual region of interest.

To truly improve the clinical workflow, the localisation information must be delivered robustly and without the need for additional preparations. Therefore, no marker-based pose estimation approaches can be applied. Further constraints are imposed by the Magnetic Resonance Imaging device itself, since no complex monitoring setups are possible due to the limited space.

This paper summarises the results of a diploma thesis carried out at the Chair for Computer Graphics, University of Erlangen, in collaboration with Siemens AG, Healthcare Sector, Erlangen, Germany, and Siemens Mindit Magnetic Resonance Ltd., Shenzhen, China. We

study the feasibility of a system to automatically determine a reclined patient's posture and pose. By **posture**, we refer to the fundamental directions a patient can be aligned on the examination table of a Magnetic Resonance scanner, namely *prone* (face-down), *supine* (on the back), or *left* or *right lateral* position. It also includes the **orientation**, which describes whether the patient is lying *head first* or *feet first* towards the magnet bore. By contrast, a patient's **pose** refers to the configuration of all joint angles and the lateral location or *translational offset* on the table. The pose also comprises the body height.

2 Modelling and Implementation

The thesis includes the review of suitable sensor technologies and the acquisition of a data set that allows first insights into the suitability. As a representative for range imaging sensor systems, a "2.5D" time-of-flight camera was selected for this study. Time-of-flight sensors are an emerging technology with promising potential especially in the field of computer vision [KBKL09], since they allow the acquisition of the depth map of a scene without the need for complex camera setups and reconstruction algorithms. By analysing the phase shift of an infrared light signal actively emitted by the camera system, the distance between objects and the camera is computed. The camera is mounted perpendicularly above the patient table, such that the field of view covers the full extent of the table.

The proposed algorithms process a single depth image as input. Therefore, the problem setting is slightly different than for motion tracking, for which an abundance of different approaches exists [MHK06]. These attempts mostly rely on data from multiple colour cameras and take advantage of observing the movement of limbs over time, but often do not provide a satisfactory solution for initialising the model in the first frame. By contrast, only one still image is available for analysis in our case.

In advance to the algorithmic work, data was acquired from 13 male and 3 female volunteers in typical clinical poses. In total, the data set comprises 152 different images.

Heuristic and instance-based algorithms allow to classify the subject's position efficiently. The task of posture classification can be divided into two stages. A first step determines heuristically, whether the patient's head or the feet are facing the magnet bore. The shortest distances between body and camera are encountered in the torso region. This information leads to a correct classification of the orientation in 97 % of the cases.

After that, the posture is further refined to prone, supine, left, or right lateral position by means of k-nearest neighbour classification [CH67]. Using the centroid-normalised, original data as feature vectors, we achieve an average accuracy of 87 %.

For pose estimation, a generative approach is considered that bases on an articulated model composed from simple geometric primitives. It is internally defined by a configuration vector of joint angles and limb lengths. The freedom of movement of each joint is restricted to the range that is expected in this clinical setting.

The transformations defining the pose of a human body are estimated indirectly by sampling the 3D model in the 2.5D space of depth maps. There, a pixel-based objective

function is used to quantify the similarity between model and observed data. By starting at a pre-defined initialisation and then traversing the pose space, the initial estimation is successively improved, based on the feedback of the error measure. For this optimisation task, we employ Powell’s Direction Set method [Pre07] which efficiently refines a single hypothesis. This scheme has been applied successfully in the context of motion tracking e.g. by Carranza, Theobalt, Magnor, and Seidel [CTMS03], and by Vidal [Vid09]. In order to escape local minima of the error function, domain knowledge is incorporated to choose good optimisation directions on the one hand and to control the expected error space on the other hand.

3 Results and Conclusions

Our experiments show that pose estimation works very well for the orientation of torso and legs. The presence of noise and the relatively low depth profile of the arms along with their high number of degrees of freedom make them considerably harder to detect. Nevertheless, in most cases a good estimation of the pose can be found. Exemplary input images and the resulting estimation of the pose, indicated by an overlay of the skeleton of the human body model, are shown in Figure 1. Note that in the second image from left a cushion is placed underneath the volunteer’s legs, and in the two rightmost images the volunteer is covered by a thin blanket.



Figure 1: Example input images and overlaid pose estimation results.

The accuracy is also analysed quantitatively by comparing the estimated coordinates of important external landmark points of the human body, such as the head, neck, shoulders, hips, and ankles with the respective manually localised position. The accuracy of the estimations typically lies within 3–10 cm, depending on the point and the posture.

Finally, after pose estimation, the body model indicates the patient's height. The average error is roughly 4–10 cm. Larger deviations are caused for example by the use of a leg cushion, a blanket, or a misleading foot pose.

In total, approximately 500 objective function evaluations are executed to find the pose in an image. A full run of our algorithm takes less than five seconds on a mid-range notebook computer.

The results prove that measurements from the time-of-flight sensor can be utilised to compute valuable information about a patient's pose on the examination table. Nevertheless, a combination with other sensor types would be advantageous, since the system then tolerates more sources of error, such as (partial) visual occlusion.

Of course, the usefulness of pose estimation from time-of-flight data is not restricted to MR imaging. Detailed knowledge of the patient position is required for the automatic positioning of C-arm Computed Tomography (CT) scanners [SRP⁺09], too. Our approach can also provide an initial estimation for human body motion tracking algorithms.

References

- [CH67] Thomas M. Cover and Peter E. Hart. Nearest neighbor pattern classification. *Information Theory, IEEE Transactions on*, 13(1):21–27, 1967.
- [CTMS03] Joel Carranza, Christian Theobalt, Marcus A. Magnor, and Hans-Peter Seidel. Free-viewpoint video of human actors. In *ACM Transactions on Graphics*, volume 22, pages 569–577, New York, NY, USA, July 2003. ACM Press.
- [KBKL09] Andreas Kolb, Erhardt Barth, Reinhard Koch, and Rasmus Larsen. Time-of-Flight Sensors in Computer Graphics. In M. Pauly and G. Greiner, editors, *Eurographics 2009 - State of the Art Reports*, pages 119–134. Eurographics Association, March 2009.
- [MHK06] Thomas B. Moeslund, Adrian Hilton, and Volker Krüger. A survey of advances in vision-based human motion capture and analysis. *Computer Vision and Image Understanding*, 104(2-3):90–126, November 2006.
- [Pre07] William H. Press. *Numerical Recipes: The Art of Scientific Computing*. Cambridge University Press, 3rd edition, September 2007.
- [SRP⁺09] Christian Schaller, Christopher Rohkohl, Jochen Penne, Michael Stürmer, and Joachim Hornegger. Inverse C-arm positioning for interventional procedures using real-time body part detection. In *MICCAI 2009, Part I, LNCS 5761*, Berlin - Heidelberg, September 2009.
- [Vid09] Marcel A. Vidal. Models and estimators for markerless human motion tracking. Master's thesis, Universitat Politècnica de Catalunya, June 2009.

Markush Structure Reconstruction

A Prototype for their Reconstruction from Image and Text into a Searchable, Context Sensitive Grammar based Extension of SMILES

Carina Haupt

Hochschule Bonn-Rhein-Sieg
Fachbereich Informatik
Grantham-Allee 20
53757 Sankt Augustin
chaupt2s@smail.inf.fh-brs.de

Abstract: In this master thesis a Markush structure reconstruction prototype is developed which is able to reconstruct bitmaps including Markush structures (meaning a depiction of the structure and a text part describing the generic parts) into a digital format and save them in the newly developed context-free grammar based file format *extSMILES*. This format is searchable due to its context-free grammar based design.

1 Introduction into the Concept of Markush Structures

Today publications are digitally available which enables researchers to search the text and often also the content of tables. On the contrary, images cannot be searched which is not a problem for most fields, but in chemistry most of the information are contained in images, especially structure diagrams.

Next to the “normal” chemical structures, which represent exactly one molecule, there also exist generic structures, so called Markush structures [Sim03]. An example therefore is shown in Figure 1. These contain variable parts and additional textual information which enable them to represent several molecules at once. This can vary between just a few and up to thousands or even millions. This ability lead to a spread of Markush structures in patents, because it enables patents to protect entire families of molecules at once. Next to the prevention of an enumeration of all structures it also has the advantage that, if a Markush structure is used in a patent, it is much harder to determine whether a specific structure is protected by it or not [Ust96]. To solve the question about the protection of a structure, it is necessary to search the patents. Appropriate databases for this task already do exist, but are filled manually. An automatic processing does not yet exist.

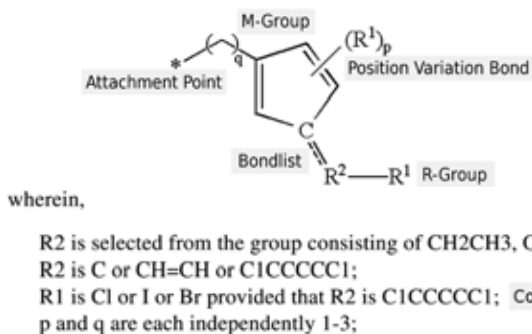


Figure 1: Example for a Markush structure containing all existing generic elements.

2 Task

In this master thesis [Hau09] a Markush structure reconstruction prototype is developed which is able to reconstruct bitmaps including Markush structures (meaning a depiction of the structure and a text part describing the generic parts) into a digital format and save them in the newly developed context-free grammar based file format *extSMILES*. This format is searchable due to its context-free grammar based design.

3 Concept and Implementation of Markush Structure Reconstruction

To be able to develop a Markush structure reconstruction prototype, an in depth analysis of the concept of Markush structures and their requirements for a reconstruction process was performed. Thereby it is stated, that the common connection table concept of the existing file formats is not able to store Markush structures. Especially challenging are conditions for most of the formats. The string-based format GENSAL [BLW81] is able to store Markush structures, but requires a modification of the R- and M-group labels, which makes a further context based processing impossible and so is not appropriate for this task.

Thus, a context-free grammar based file format is developed, which extends the SMILES format [Wei88]. This *extSMILES* called format assures the searchability of the results by its context-free grammar based concept, and is able to store all information contained in Markush structures. In addition it is generic, extendable and easily understandable.

The developed prototype for the Markush structure reconstruction uses *extSMILES* as output format and is based on the chemical structure recognition tool chemOCR [AZF⁺07] and the Unstructured Information Management Architecture UIMA [FL04] (see Figure 2). For chemOCR modules are developed which enable it to recognize and assemble Markush structures as well as to return the reconstruction result in *extSMILES*. For UIMA on the other

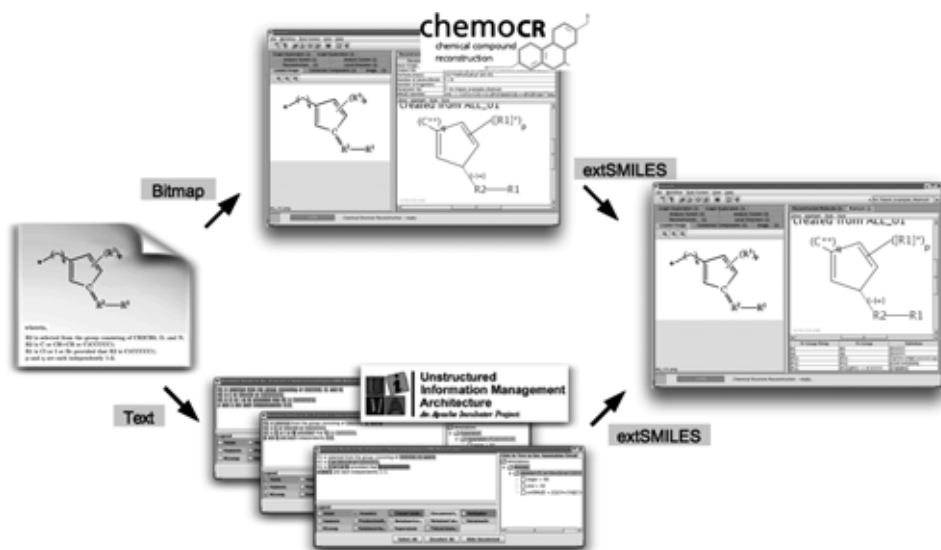


Figure 2: Workflow of Markush structure reconstruction prototype. The input is split up into the depiction of the core structure and the text, which then are independently processed by chemOCR and UIMA. These tools return *extSMILES* strings, which are then merged and presented in a newly developed viewer in chemOCR.

hand, a pipeline is developed, which is able to analyse and translate the input text files to *extSMILES*. The results of both tools then are combined and presented in chemOCR.

4 Evaluation and Results

An evaluation of the prototype is performed on a representative set of twelve structures of interest and low image quality which contain all typical Markush elements. Trivial structures containing only one R-group are not evaluated. Due to the challenging nature of the images, no Markush structure could be correctly reconstructed. But by regarding the assumption, that R-group definitions which are described by natural language are excluded from the task, and under the condition that the core structure reconstruction is improved, the rate of success can be increased to 58.4%.

5 Funding

This work was funded by the Fraunhofer Institute for Algorithms and Scientific Computing SCAI.

References

- [AZF⁺07] M. E. Algorri, M. Zimmermann, C. M. Friedrich, S. Akle, and M. Hofmann-Apitius. Reconstruction of Chemical Molecules from Images. In *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE*, pages 4609–4612, 2007.
- [BLW81] J. M. Barnard, M F. Lynch, and Stephen M. Welford. Computer storage and retrieval of generic chemical structures in patents. 2. GENSAL, a formal language for the description of generic chemical structures. *Journal of Chemical Information and Computer Sciences*, 21(3):151–161, August 1981.
- [FL04] D. Ferrucci and A. Lally. Building an example application with the unstructured information management architecture. *IBM Systems Journal*, 43(3):455–475, July 2004.
- [Hau09] C. S. Haupt. Markush Structure Reconstruction - A Prototype for their Reconstruction from Image and Text into a Searchable, Context Sensitive Grammar based Extension of SMILES. Master's thesis, Hochschule Bonn-Rhein-Sieg, December 2009.
- [Sim03] E. S. Simmons. Markush structure searching over the years. *World Patent Information*, 25(3):195–202, September 2003.
- [Ust96] E. Ustinova. Are Markush structures matters of chemistry and law or just figments of the imagination? *World Patent Information*, 18(1):23–31, March 1996.
- [Wei88] D. Weininger. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *Journal of Chemical Information and Computer Sciences*, 28(1):31–36, February 1988.

Prototypische Realisierung eines Clients zur Visualisierung von verteilten XML-Dokumenten für ein Informationssystem in der Landwirtschaft

Ronnie Kullick, Daniel Martini
KTBL
Bartningstr. 49
64289 Darmstadt
{r.kullick, d.martini}@ktbl.de

Hans-Peter Wiedling
Hochschule Darmstadt
Fachbereich Informatik
h.wiedling@fbi.h-da.de

Abstract: XML hat sich als wichtige Technologie für den elektronischen Datenaustausch im Web etabliert. Auch in der Landwirtschaft hat XML durch die Entwicklung von agroXML ihren Einzug erhalten. Gegenstand der Masterarbeit ist die Entwicklung eines Clients, der eine integrierte einheitliche Sicht auf verteilte und verknüpfte XML-Dokumente bietet. Wichtiges Ziel ist die Erweiterbarkeit und die Flexibilität der Anwendung, die durch den entwickelten Ansatz zur Laufzeit ein XML Data Binding herstellt. Aufbauend auf diesem Ansatz werden eine dynamische Formulargenerierung, sowie weitere Ansichten zur Visualisierung von verteilten XML-Dokumenten umgesetzt.

1 Einleitung

Wie in vielen Bereichen wächst auch der Informationsbedarf in der Landwirtschaft und deren vor- und nachgelagerten Bereichen kontinuierlich. Zunehmende Verflechtungen zwischen landwirtschaftlichen Betrieben und deren Partnern spielen eine entscheidende Rolle. Aus diesem Grund wird der digitalen Informationstechnologie seitens der Landwirtschaft steigende Aufmerksamkeit geschenkt.

Derzeit wird auf vielen landwirtschaftlichen Betrieben Software zum Betriebsmanagement (Farm Management Information System) eingesetzt. Bedingt durch die historische Entwicklung handelt es sich bei diesen Anwendungen überwiegend um monolithische geschlossene Systeme, die durch unterschiedliche Datenformate und individuelle Schnittstellen geprägt sind. Dies führt zu einem erschwerten elektronischen Datenaustausch und Hindernissen bei der Informationserfassung und -weiterverarbeitung aufgrund der geringen Kompatibilität und Inoperabilität bei gleichzeitigem Anstieg der Dokumentationsanforderungen.

Aus dieser Situation heraus entwickelte das KTBL gemeinsam mit der FH Bingen und den führenden deutschen Agrarsoftwareherstellern das XML-basierte Datenaustauschformat agroXML [agroXML]. Ziel ist es dabei, Datenaustauschprozesse in der Landwirtschaft zu vereinheitlichen und effizienter zu gestalten.

Im Rahmen des interdisziplinären Forschungsprojekts IT FoodTrace (IT supported Food Traceability, [ITFoodTrace]) wurden unter Verwendung von agroXML IT-gestützte Lösungen umgesetzt, um lückenlose Rückverfolgbarkeit und Qualitätssicherung entlang der Fleisch- und Fleischwaren Wertschöpfungskette – vom Futtermittel bis zum Endverbraucher – zu erzielen. Kernproblem dabei ist, dass es eine Reihe von Orten gibt, an denen für die Qualitätssicherung relevanten Daten zum Produkt oder zu dessen Verarbeitung erhoben und abgelegt werden.

2 Zielsetzung

Gegenstand der Master Thesis war die Entwicklung einer prototypischen Clientanwendung, die eine integrierte und einheitliche Sicht auf verteilt gehaltene Daten realisiert. Der Client wird in einer ressourcen-orientierte Architektur (ROA, siehe [Fi00]) eingesetzt und adressiert XML-Repräsentationen der Ressourcen per URI über HTTP. Ein Schwerpunkt des Clients lag in der Integration von XML-Dokumenten, die mittels XLink miteinander verknüpft sind (siehe Abbildung).

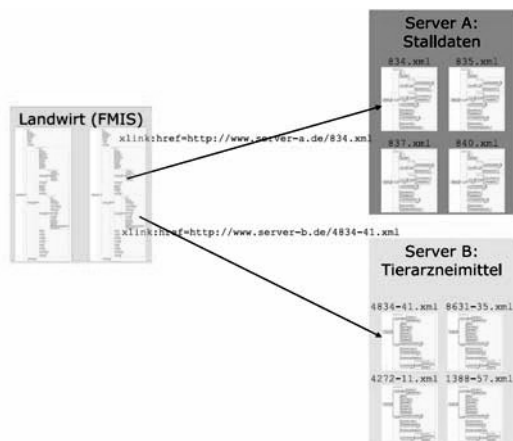


Abbildung: Verknüpfung landwirtschaftlicher Dokumente mit XLink

Es wurde sowohl eine formularbasierte Oberfläche als auch eine geeignete grafische Repräsentation zur Darstellung der verknüpften XML-Dokumenten vorgesehen. Im Anwendungsbeispiel können neue Betriebe, Partner oder Ställe in einer Prozessdokumentation eingebunden werden. Mögliche Änderungen an einem XML Schema ist ein weiterer Aspekt der Erweiterbarkeit, die vom Client berücksichtigt werden muss. Der Client muss eine Authentifizierungskomponente beinhalten, um auf geschützte Daten zugreifen zu können. Außerdem soll die Anwendung möglichst allgemein einsetzbar sein, und sich somit nicht ausschließlich auf den Einsatzbereich der Landwirtschaft oder auf agroXML beschränken.

3 Analyse von XML Binding Frameworks

Ein wichtiger Bestandteil der Arbeit war die Auswahl eines geeigneten XML Data Binding Frameworks in Java. XML Data Binding Werkzeuge bieten eine komfortable Möglichkeit, XML-Daten zu integrieren. Hierzu wurden die grundlegende Funktionsweise der Open-Source XML Data Binding Werkzeuge Castor XML, Apache XMLBeans und der Sun Spezifikation JAXB vorgestellt und anhand gestellter Kriterien und Eigenschaften gegenübergestellt.

Bei der Auswahl eines XML Data Binding Frameworks wurde das Hauptaugenmerk auf die geforderte Erweiterbarkeit gelegt. In diesem Zusammenhang wurden zwei Ansätze gegenübergestellt, wie ein XML Data Binding zur Laufzeit hergestellt werden kann und welche XML Data Binding Bibliothek dafür am besten geeignet ist.

Als Ergebnis dieser Analyse wurde die XMLBeans mit ihrem Schématypensystem ausgewählt. Hiermit ist ein neuer Ansatz implementiert, wo durch ein XML Binding für beliebige XML Schemata mit Hilfe von Metainformationen zur Laufzeit der Anwendung realisiert wird.

5 Umsetzung der Client Anwendung

Im praktischen Teil der Masterarbeit wurden die Anforderungen an die Clientanwendung erfasst. Es wurde die Ziele und Zielgruppen der Anwendung identifiziert.

Die Anwendung wurde auf Basis des Swing Application Frameworks [SAF] realisiert. Sie bietet dem Benutzer verschiedene Ansichten auf das zusammengesetzte integrierte XML-Dokument an. Mit Hilfe von JGoodies Forms [Le06] wurde ein formularbasierter Editor implementiert, in dem der Inhalt eines XML-Dokuments auf Basis seines XML Schemas bearbeitet werden kann. Des Weiteren wurde mit der Open-Source Visualisierungsbibliothek Prefuse [Prefuse] eine interaktive grafische Aufbereitung und Darstellung der verknüpften XML-Dokumente umgesetzt.

Anhand von zwei praktischen Szenarien wurde die Funktionsfähigkeit und Praxistauglichkeit der Anwendung demonstriert.

6 Verwertung

Der Client wurde im Rahmen verschiedener Anwendungsfälle zur Demonstration von Technologien im agro XML-Umfeld verwendet. Haupteinsatzbereich war im IT FoodTrace-Projekt. Die erarbeiteten Ergebnisse wurden auf der CeBIT am Stand des Bundesministerium des Inneren präsentiert. Eindrucksvoll konnte die Verknüpfung von Daten an verschiedensten Orten in der Lebensmittelkette visualisiert werden.

Auf der Tagung der Gesellschaft für Informatik in der Landwirtschaft (GIL) wurde die Anwendung für eine Live-Demonstration innerhalb eines Vortrags eingesetzt.

7 Ausblick

Mit der derzeitigen Ausbaustufe der Anwendung wurden die technischen Grundlagen eines Visualisierungstools für verteilte XML-Dokumente entwickelt. Für zukünftige Ausbaustufen sind einige technische Erweiterungen vorstellbar.

Eine grundsätzliche Einschränkung der entwickelten Clientanwendung ist die ausschließliche Unterstützung von zyklensfreien Verknüpfungen von XML-Dokumenten. Auch Graphstrukturen sollten in Zukunft angezeigt werden können.

Nach der Bearbeitung der XML-Dokumente werden diese derzeit lokal abgespeichert. Als Erweiterung kann die Realisierung der HTTP POST und PUT Methoden in Betracht gezogen werden, um die erzeugten XML-Dokumente wieder auf dem Server abzulegen.

In einer weiteren Ausbaustufe könnte die Zielgruppe des Endnutzers stärker berücksichtigt werden. Dafür ist es beispielsweise notwendig, im Formulareditor eine entsprechende muttersprachliche Beschriftung der Felder vorzusehen. Außerdem kann dem Endnutzer durch Ressourcenbeschreibungen weitere Unterstützung bei der Auswahl der zu ladenden Dokumente, z.B. durch Icons, geboten werden.

Literaturverzeichnis

- [agroXML] Kuratorium für Technik und Bauwesen in der Landwirtschaft e.V.: agroXML.
URL: <http://www.agroxml.de/>
- [Fi00] Fielding, R. T.: Architectural Styles and the Design of Network-based Software Architectures. Doctoral dissertation, University of California, Irvine, 2000.
URL: www.ics.uci.edu/~fielding/pubs/dissertation/top.htm
- [ITFoodTrace] Forschungsverbund IT FoodTrace: IT FoodTrace – Internet supported Food Traceability. Forschungsprojekt IT FoodTrace von der Universität Hohenheim, IBM Deutschland GmbH, comundus GmbH und KTBL.
URL: <http://www.itfoodtrace.de>
- [Le06] Karsten Lentzsch: JGoodies – Java User Interface Design – JGoodies Forms. 2006.
URL: <http://www.jgoodies.com/freeware/forms/index.html>
- [Prefuse] prefuse Information Visualization Toolkit: the prefuse visualization toolkit.
URL: <http://www.prefuse.org>
- [SAF] Sun Microsystems: Swing Application Framework.
URL: <https://appframework.dev.java.net/>

Quantitative measurement of regional wall motion in the right heart ventricle

Kerstin Lenk

Lausitz University of Applied Sciences
Course of studies: Computer Science
kerstin.lenk@hs-lausitz.de

Type of thesis: Diploma thesis
Supervisors: Prof. Dr. Barbara Priwitzer (Lausitz University of Applied Sciences),
Dr. Graeme P. Penney (King's College London)

Currently, there is no gold standard for measuring the wall motion of the right heart ventricle which plays an important role in congenital heart diseases. We developed a fully automatic method to allow an analysis of the wall motion in defined regions in the right ventricle.

1 Objective

We aimed to develop a new measurement tool for computing right ventricle motion from standard magnetic resonance (MR) image series (i.e. 3D+time) to enable a quantitative analysis of congenital heart diseases.

2 Background

The right ventricle plays an important role in congenital heart defects [SR08] such as Tetralogy of Fallot. Wall motion abnormalities are often involved in right ventricle failure [AHY05, BIC08, Ra09]. The main problem for assessment is its complex shape in comparison to the left ventricle. Thus it is difficult to develop a simple and reproducible method for accurate motion measurement. To the best of our knowledge there is no generally accepted method for the assessment.

3 Methods

We used short-axis and transverse 2D+time datasets, which were reformatted into 3D images (i.e. 3D+time), from five patients with heart disease and five volunteers without heart disease (i.e. $2 \cdot (5+5) = 20$ datasets with short-axis and transverse images) which came from the database of the 'Competence Network for Congenital Heart Defects'.

The end-diastolic phase was registered to all other phases of the cardiac cycle using the non-rigid registration algorithm by Rueckert et al. [Ru99, CMR04, CMR05] (A good overview about image registration is provided in [CHH04]. An example of non-rigid registration is shown in figure 1). Rueckert et al. developed a hierarchical non-rigid transformation model to register breast MR images before and after contrast agent

application. Firstly, the global motion was captured by an affine transformation. Secondly, an affine transformation was applied. Thirdly, a B-spline based free-form deformation model was used to transform the control points. The work showed that motion artifacts could be reduced [Ru99].

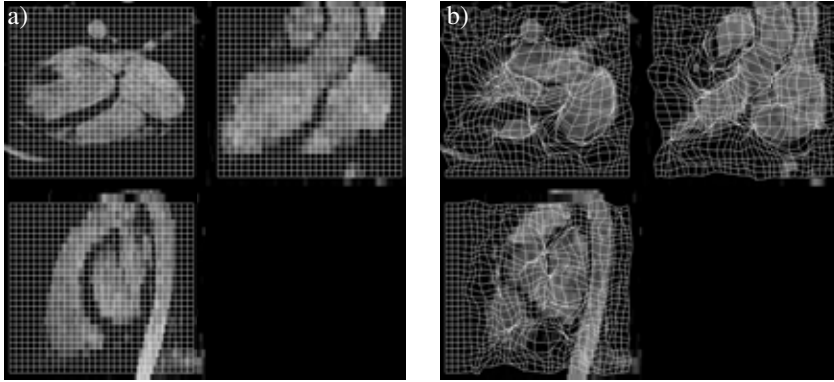


Figure 1 shows an example of non-rigid registration to two phases of the cardiac cycle. Figure 1(a) shows a transverse MR image of a volunteer and the respective undeformed motion grid at the end-diastole (phase 1) in three orthogonal views. The image at the first time point in the cardiac cycle was then mapped to all other phase in the cardiac cycle using a B-spline based free-form deformation model. Figure 1(b) displays the corresponding transverse MR image at the end-systole together with the respective deformed motion grid (yellow).

In order to find consistent regions around the right ventricle, we carried out a manual segmentation on a single dataset. This segmentation consisted of five anatomically relevant regions (orientated on Bodhey et al. [BBS08]) and was propagated to the other datasets using image registration. Results are presented in graphs (figure 2) which show the displacement of each region to a reference point (approximate centre of right ventricle).

4 Results and Discussion

The results demonstrate the potential of our approach to quantify motion characteristics from cardiac MR image series in a clinical setting (figure 2). Accuracy was assessed using visual inspection by experts.

The registrations from phase-to-phase and intrasubject registrations were shown to be robust and accurate. Although the propagation of defined segments using image registration was successful in all healthy volunteers, disease related alterations of the ventricular morphology reduced the success rate in patients.

The next step was to calculate the heart phase (e.g. the patient's dataset, figure 2, is represented by 20 phases) at which each anatomical segment was closest to the reference point of the right ventricle. The median value of these phase values was compared to the end-systolic phase determined using visual inspection by experts. They all agreed to within one cardiac phase. The displacements between the short-axis and the transverse datasets were calculated and ranged in between 0.35 and 4.67 mm which are respectable results.

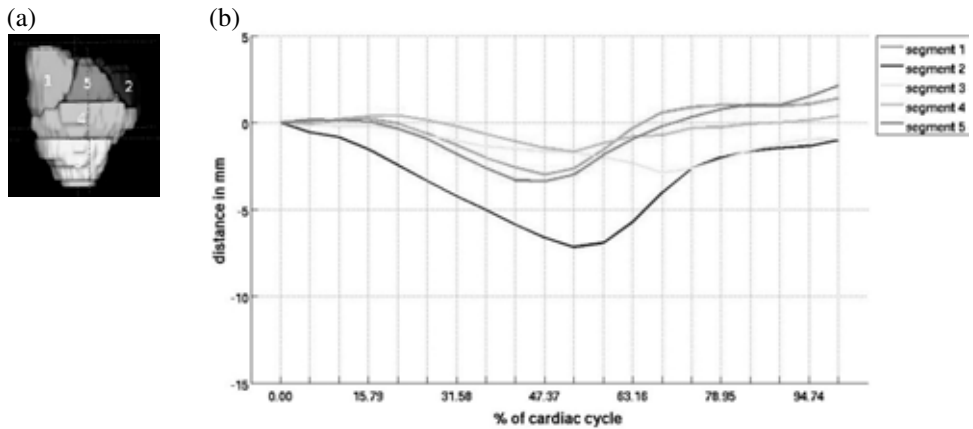


Figure 2: Results based on a short-axis dataset of a patient with heart disease; (a) segmentation of the right ventricle; (b) example graph of the distance measurement

This method will be used to analyze a large cardiac MR image database acquired from 710 patients with repaired Tetralogy of Fallot (the most common complex congenital heart disease - 10% of all congenital heart disease [HGB05]).

5 Conclusions

The developed motion analysis tool shows potential to accurately capture large scale motion of the defined right ventricle wall segments in reformatted 3D+time MR images.

Abstract publications about the subject:

- Lenk K., Beerbaum P., Penney G.P.: A new method for quantifying regional RV wall motion from standard 2D cine MR volumetric imaging in Tetralogy of Fallot. *Clin Res Cardiol* 2009;98:599. (talk at the 41st Annual Meeting of the German Society of Pediatric Cardiology in Weimar)
- Parish V., Bell A., Sarikouch S., Lenk K., Greil G., Razavi R., Beerbaum P.: Assessment of right ventricular contractile reserve in corrected Tetralogy of Fallot by Dobutamine Stress MRI. *Clin Res Cardiol* 2009;98:608. (poster at the 41st Annual Meeting of the German Society of Pediatric Cardiology in Weimar)

References

- [AHY05] Abd El Rahman M.Y., Hui W., Yigitbasi M., et al.: Detection of left ventricular asynchrony in patients with right bundle branch block after repair of Tetralogy of Fallot using tissue-Doppler imaging-derived strain. *Journal of the American College of Cardiology* 2005;45:915-921.
- [BBS08] Bodhey N.K., Beerbaum P., Sarikouch S., et al.: Functional Analysis of the Components of the Right Ventricle in the Setting of Tetralogy of Fallot. *Circulation-Cardiovascular Imaging* 2008;1:141-147.
- [BIC08] Bordachar P., Iriart X., Chabaneix J., et al.: Presence of ventricular dyssynchrony and haemodynamic impact of right ventricular pacing in adults with repaired Tetralogy of Fallot and right bundle branch block. *Europace* 2008;10:967-71.
- [CMR04] Chandrashekara R., Mohiaddin R.H., Rueckert D.: Analysis of 3-D myocardial motion in tagged MR images using non-rigid image registration. *Ieee Transactions on Medical Imaging* 2004;23:1245-1250.
- [CMR05] Chandrashekara R., Mohiaddin R.H., Rueckert D.: Comparison of cardiac motion fields from tagged and untagged MR images using nonrigid registration. In: Frangi ARPISAHM, ed. 3rd International Workshop on Functional Imaging and Modeling of the Heart. Barcelona, SPAIN: Springer-Verlag Berlin, 2005:425-433.
- [CHH04] Crum W.R., Hartkens T., Hill D.L.G.: Non-rigid image registration: theory and practice. *British Journal of Radiology* 2004;77:S140-S153.
- [HGB05] Hombach V., Grebe O., Botnar R.M., editors: *Kardiovaskuläre Magnetresonanztomographie. Grundlagen – Technik – klinische Anwendung*. Stuttgart, New York: Schattauer, 2005.
- [Ra09] Raedle-Hurst T.M., Mueller M., Rentzsch A., et al.: Assessment of left ventricular dyssynchrony and function using real-time 3-dimensional echocardiography in patients with congenital right heart disease. *American Heart Journal* 2009;157:791-798.
- [Ru99] Rueckert D., Sonoda L.I., Hayes C., et al.: Non-rigid registration using free-form deformations: Application to breast MR images. *Ieee Transactions on Medical Imaging* 1999;18:712-721.
- [SR08] Sheehan F., Redington A.: The right ventricle: anatomy, physiology and clinical imaging. *Heart* 2008;94:1510-1515.

Model-Based Respiratory Motion Compensation for Image-Guided Cardiac Interventions

Matthias Schneider*

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

Abstract: In this thesis we describe and validate a statistical respiratory motion model for motion compensation during image-guided cardiac interventions. In a preparatory training phase, a preoperative 3-D segmentation of the coronary arteries is automatically registered with a cardiac-gated biplane cineangiogram at different breathing phases to build a subject-specific motion model. This motion model is used as a prior within the intraoperative registration process for motion compensation by restricting the search space to the most dominant modes of the motion profile. In this way, the interventionalist is provided visual guiding assistance during cardiac catheterization procedures under live fluoroscopy. The model-constrained registration increases the robustness and accuracy of the dynamic re-registrations, especially for weak data constraints such as low signal-to-noise ratio, the lack of contrast information in fluoroscopic images, or an intraoperative monoplanar setting. This allows for reducing radiation exposure without compromising on registration accuracy. Synthetic data as well as phantom and clinical datasets have been used to validate the model-based registration in terms of accuracy, robustness and speed. We were able to significantly accelerate the intraoperative registration with an average 3-D error of less than 2 mm even for monoplanar settings and tracked guidewire data in the absence of contrast agent, which makes respiratory motion correction feasible during clinical procedures.

Moreover, we propose a new methodology for global vessel segmentation in 2-D X-ray images combining a per-pixel probability map from a local vessel enhancement filter with local directional information. The global segmentation provides additional connectivity and geometric shape information, which allows for sophisticated postprocessing techniques in order to refine the segmentation with respect to the requirements of the particular application.

Introduction

Minimally invasive cardiac interventions such as balloon angioplasty and coronary stenting for percutaneous recanalization are routinely guided under X-ray fluoroscopy. In particular, guidewire crossing of Chronic Total Occlusions (CTO) within Percutaneous Coronary Interventions (PCI) is challenging due to the lack of blood flow and contrast concentration in the occluded segment resulting in poor X-ray guidance. The operator steers and advances the interventional instruments essentially blindly in the diseased artery. For this reason, highly experienced cardiologists are required to reduce the high risk of hemorrhage, perforation or even dissection of the artery wall.

*This Diplomarbeit (Master Thesis) was supervised by Prof. Dr.-Ing. Joachim Hornegger (Pattern Recognition Lab, Friedrich-Alexander University Erlangen-Nuremberg, Germany), Hari Sundar, Ph.D., and Chenyang Xu, Ph.D. (both Siemens Corporate Research, Princeton, NJ, USA)

Integrating preoperative plans into the interventional surgery suite is a common way to assist the physician during the intervention. In the case of PCI for CTO, Computed Tomography (CT) guidance is able to provide useful visual information. A 3-D segmentation of the coronary arteries can be extracted from preoperative cardiac CT data [GT08] and then be overlaid on top of the intraprocedure fluoroscopic images [SKX⁺06]. In this way, the surgeon is provided additional anatomical information including the position of the vessel branches and the occlusion.

However, maintaining up-to-date information during the intervention requires continuous re-adjustment of the registration due to cardiac and respiratory motion. Such re-registrations of preoperative volumetric datasets with intraoperative images are challenging and computationally demanding due to the low quality and lack of contrast in intraprocedure fluoroscopic images.

Related Work

Various techniques have been proposed to overcome the problem of cardiac and respiratory motion in cardiac imaging. Errors due to cardiac motion can partly be avoided using electrocardiography (ECG) gating. Since the human heart rate is usually not smaller than 60 beats per minute, ECG gating yields a frame rate of about one frame per second. Similar approaches for respiratory gating in cardiac Magnetic Resonance Imaging (MRI) have been investigated, usually using displacement transducers to estimate the breathing phase [EMP⁺84]. Due to the relatively long breathing cycle of several seconds, the resulting imaging frequency of respiratory-gated angiograms is significantly reduced and hence not practical for interventional applications.

Another strategy for reducing respiratory motion is the breath-hold technique. Even if this method successfully reduces breathing motion by simple and natural means, it is significantly restricted by the patient's limited ability to perform a supervised breath-hold during the treatment or image acquisition [MY04].

A third class of strategies copes with the problem of respiratory motion correction by incorporating suitable motion models. Manke et al. [MRN⁺02] proposed a patient-specific affine motion model based on a diaphragmatic pencil-beam navigator for prospective respiratory motion correction in coronary MRI. In a follow-up study [MNB03], the authors extended the technique by multiple transducers at moving structures to achieve better correlation particularly for non-translational components of the affine parameter set.

Similarly, King et al. [KBR⁺09] proposed an affine motion model for respiratory motion correction of MRI-derived roadmaps for use in X-ray guided cardiac catheterization procedures. The model uses second order polynomials to relate the diaphragmatic position to the affine pose parameters. The intraprocedure diaphragm tracking requires manual landmark selection and relies on the diaphragm to be visible in the images. The authors report a model accuracy of 2–4 mm in 3-D and 3–4 mm in 2-D, respectively.

Shechter et al. [SSRB05] developed a parametric 3-D+time coronary model to correct X-ray images for both cardiac and respiratory motion. The diaphragm tracking again requires manual landmark selection. Another major drawback of the proposed navigator-based techniques is that they purely rely on the transducer data to “predict” the corre-

sponding pose parameters without considering any further image information. Therefore, intra-subject variabilities within the breathing cycle can not be accounted for.

Recently, Vandemeulebroucke et al. [VKCS09] proposed a technique for respiratory motion estimation from 2-D C-arm projections for radiotherapy of lung cancer. The authors introduce a parametric motion model that is able to cope with hysteresis effects and intrasubject variations. Even though the results of this approach are promising, it is not applicable for near-real-time motion compensation in image-guided cardiac interventions due to the computational complexity and the presence of cardiac motion, which makes it difficult to acquire proper training data for the model estimation.

Contribution¹

Studies on respiratory motion of the heart [MHA⁺02, SRM06] have concluded that breathing motion can be approximated as rigid transformations with a patient-specific profile. Even though this kind of motion is more complex than just pure 3-D translation, the profile is restricted in terms of both 3-D rotation and translation. In this thesis, we learn the patient-specific motion profile from a set of pose samples using Principal Component Analysis (PCA). The training data is obtained from automatic rigid 2-D/3-D registration [SKX⁺06] of a cardiac-gated biplane cineangiography at different breathing phases with the preoperative 3-D segmentation of the coronary arteries [GT08]. The PCA-based model is subsequently used to constrain re-registrations performed during interventions by restricting the search space to the most dominant modes of the motion profile.

The model's ability to recover respiratory motion has been validated on synthetic, phantom, and clinical datasets. The model-based registration greatly improves the registration accuracy and robustness with an average 3-D error of less than 2 mm even when performed using a monoplane image. For intraprocedure fluoroscopies without contrast agent, the model-constrained registration is based on 2-D guidewire tracking. The accuracy has been assessed for two clinical CTO cases resulting in an average 3-D error of 1.4–2.2 mm, which makes respiratory motion correction feasible during clinical procedures.

The accurate estimation of the cardiovascular pose at different breathing phases is required to robustly build the motion model. In this thesis, we use a feature-based rigid 2-D/3-D registration approach [SKX⁺06] to obtain the training data for the motion model. The quality of the registration result, however, strongly depends on the 2-D segmentation of the vascular structures in the X-ray images. In order to make the static registration more robust, we investigate a novel technique for automatic global vessel segmentation in 2-D X-ray angiograms using local directional information and vector field integration. In contrast to binary per-pixel segmentation, the global segmentation approach allows for sophisticated postprocessing techniques to tailor the segmentation to the requirements of the particular application. With respect to our needs, we propose different postprocessing techniques for hierarchical segmentation, centerline extraction, and catheter removal. We further demonstrate that the global segmentation approach yields better segmentation results and is more robust to noise compared to two conventional local Hessian-based approaches [FNVV98, SNS⁺98].

¹Parts of this thesis have been accepted for IEEE International Symposium on Biomedical Imaging (ISBI) 2010, and IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2010.

References

- [EMP⁺84] R. L. Ehman, M. T. McNamara, M. Pallack, H. Hricak, and C. B. Higgins. Magnetic resonance imaging with respiratory gating: techniques and advantages. *American Journal of Roentgenology*, 143(6):1175–1182, December 1984.
- [FNVV98] A. F. Frangi, W. J. Niessen, K. L. Vincken, and M. A. Viergever. Multiscale vessel enhancement filtering. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI’98*, volume 1496 of *LNCS*, pages 130–137. Springer, 1998.
- [GT08] M. A. Gülsün and H. Tek. Robust vessel tree modeling. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2008*, volume 5241 of *LNCS*, pages 602–611. Springer, 2008.
- [KBR⁺09] A.P. King, R. Boubertakh, K.S. Rhode, Y.L. Ma, P. Chinchapatnam, G. Gao, T. Tangcharoen, M. Ginks, M. Cooklin, J.S. Gill, D.J. Hawkes, R.S. Razavi, and T. Schaeffter. A subject-specific technique for respiratory motion correction in image-guided cardiac catheterisation procedures. *Medical Image Analysis*, 13(3):419–431, June 2009.
- [MHA⁺02] K. McLeish, D.L.G. Hill, D. Atkinson, J.M. Blackall, and R. Razavi. A study of the motion and deformation of the heart due to respiration. *IEEE Transactions on Medical Imaging*, 21(9):1142–1150, September 2002.
- [MNB03] D. Manke, K. Nehrke, and P. Börnert. Novel prospective respiratory motion correction approach for free-breathing coronary MR angiography using a patient-adapted affine motion model. *Magnetic Resonance in Medicine*, 50(1):122–31, July 2003.
- [MRN⁺02] D. Manke, P. Rösch, K. Nehrke, P. Börnert, and O. Dössel. Model evaluation and calibration for prospective respiratory motion correction in coronary MR angiography based on 3-D image registration. *IEEE Transactions on Medical Imaging*, 21(9):1132–1141, September 2002.
- [MY04] G. S. Mageras and E. Yorke. Deep inspiration breath hold and respiratory gating strategies for reducing organ motion in radiation treatment. *Seminars in Radiation Oncology*, 14(1):65–75, January 2004.
- [SKX⁺06] H. Sundar, A. Khamene, Ch. Xu, F. Sauer, and Ch. Davatzikos. A novel 2D-3D registration algorithm for aligning fluoroscopic images with pre-operative 3D images. In *Proceedings of SPIE*, volume 6141, San Diego, CA, USA, February 2006.
- [SNS⁺98] Y. Sato, Sh. Nakajima, N. Shiraga, H. Atsumi, Sh. Yoshida, T. Koller, G. Gerig, and R. Kikinis. 3D multi-scale line filter for segmentation and visualization of curvilinear structures in medical images. In *Medical Image Analysis*, volume 2, pages 143–168, June 1998.
- [SRM06] G. Shechter, J.R. Resar, and E.R. McVeigh. Displacement and velocity of the coronary arteries: cardiac and respiratory motion. *IEEE Transactions on Medical Imaging*, 25(3):369–375, March 2006.
- [SSRB05] G. Shechter, B. Shechter, J.R. Resar, and R. Beyar. Prospective motion correction of X-ray images for coronary interventions. *IEEE Transactions on Medical Imaging*, 24(4):441–450, April 2005.
- [VKCS09] J. Vandemeulebroucke, J. Kybic, P. Clarysse, and D. Sarrut. Respiratory motion estimation from cone-beam projections using a prior model. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2009*, volume 5761 of *LNCS*, pages 365–372. Springer, 2009.

Shader-Based-Image-Processor - Ein Framework für echtzeitfähige, grafikartenbasierte Bilderverarbeitung

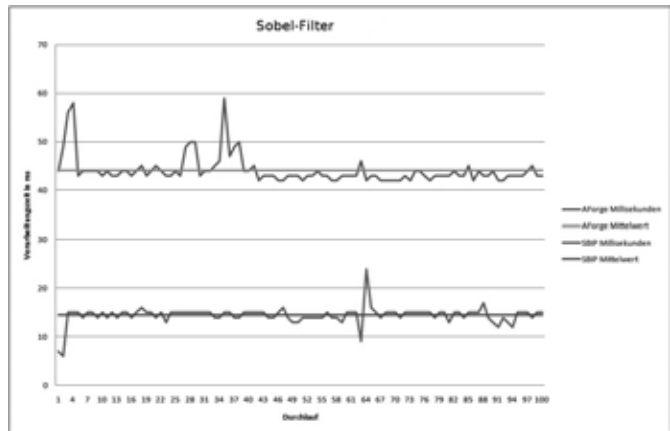
Martin Friedl martin.friedl@fh-erfurt.de

Frank Nagl frank.nagl@fh-erfurt.de

Andreas Schaefer andreas.schaefer@fh-erfurt.de

Andre Tschentscher andre.tschentscher@fh-erfurt.de

Abstract: In diesem Paper wird ein neuartiges Framework zur echtzeitfähigen Bildverarbeitung vorgestellt. Hierfür wird die Parallelisierung sequentieller Verarbeitungsschritte durch grafikartenbasierte Programme (Shader) genutzt. Das Framework stellt dazu eine Schnittstelle für die Entwicklung shaderbasierter Bildverarbeitungsfilter zur Verfügung. Zusätzlich wird vom Framework eine auf der Basis der 3D-API DirectX und der Shadertechnologie HLSL aufbauende Komponente bereitgestellt, welche das Anwenden von 20 Beispielfiltern ermöglicht.



(a) Original (oben) und Ergebnisbild des Sobel-Filters (unten) (b) Vergleich der Verarbeitungszeiten des Sobel-Filters zwischen AForge.NET und SBIP

1 Motivation

Die Bildverarbeitung ermöglicht Anwendern, digitale Fotos zu optimieren, zu manipulieren oder speziellen Wünschen anzupassen. Ungeachtet moderner Prozessortechnologien sind viele Anwendungen und Bibliotheken nicht in der Lage, Bilder in Echtzeit zu verarbeiten. Konventionelle Bildverarbeitungsbibliotheken arbeiten ausschließlich auf der

CPU. Die Parallelisierung von Arbeitsabläufen mit Hilfe der Grafikkarte und des Grafikkartenspeicher bleiben hierbei vollständig ungenutzt.

Unser Framework (im Folgenden Shader-Based-Image-Processor, kurz **SBIP**, genannt) bietet Softwareentwicklern eine Möglichkeit, Bildverarbeitungsmechanismen auf die GPU auszulagern. So können Bildverarbeitungsfilter in annähernder Echtzeit verarbeitet werden. Zusätzlich beinhaltet SBIP eine Komponente namens **HLSLFramework**. Diese stellt eine auf der High-Level-Shading-Language (HLSL) und der 3D-API DirectX basierende Sammlung von Bildfiltern bereit.

2 State-Of-The-Art

In der Bildverarbeitung ist das Programm Adobe Photoshop [ado09] der Marktführer. Die aktuelle Version beinhaltet im Funktionsumfang eine Unterstützung der CPU durch die vorhandene Grafikkarte[ado09]. Diese ermöglicht ein echtzeitfähiges, stufenloses Zoomen und Rotieren der Arbeitsfläche. Weitere Features sind in der Anwendung von Lichteffekten und Farbkonvertierung zu finden. Abgrenzend zu unserem Framework existiert hier keine Schnittstelle für Anwender und Entwickler zur Erweiterung des Funktionsumfangs durch eigene grafikartenbasierte Programme (**Shader**).

Die Bildverarbeitungsbibliotheken ImageJ [ima10], Image-Magick [mag10] und AForge.NET [Kir09] ermöglichen ausschließlich die Berechnung von Verarbeitungsschritten auf der CPU.

Wie der aufgezeigten Kurzfassung entnommen werden kann, ist zur Zeit kein einheitlicher und flexibler Lösungsansatz zu erkennen, welcher das Ziel verfolgt, die Parallelisierung von Verarbeitungsschritten einer Grafikkarte erschöpfend zu nutzen. Im nächsten Abschnitt werden das Konzept für die Erstellung des SBIP und eine umgesetzte Implementierung erläutert.

3 Konzept und Implementierung des Frameworks SBIP

Als Anforderung an das Framework ist die Bereitstellung einer Schnittstelle für die Echtzeitverarbeitung von Bildern gestellt. Hierbei sollen die GPU und der Grafikkartenspeicher unter Verwendung von Shadern genutzt werden. Um die Nutzung von SBIP durch eine große Zahl von Grafikkarten und Betriebssystemen zu ermöglichen, soll das Framework flexibel sein und unterschiedliche Shadertechnologien verwenden können.

Bilder werden durch so genannte Bildfilter verarbeitet. Dabei werden alle Pixel eines Bildes betrachtet und mithilfe der Nachbapixel und einer zuvor definierten Filterart neu bestimmt. In konventionellen Verfahren erfolgt dies sequentiell. Dies verhindert eine Echtzeitberechnung vollständig. SBIP ermöglicht das Verarbeiten von Bildfiltern auf der Grafikkarte. Mithilfe von Shadern werden die sequentiellen Verarbeitungsschritte eines Bildfilters parallelisiert und das Bild echtzeitfähig verarbeitet.

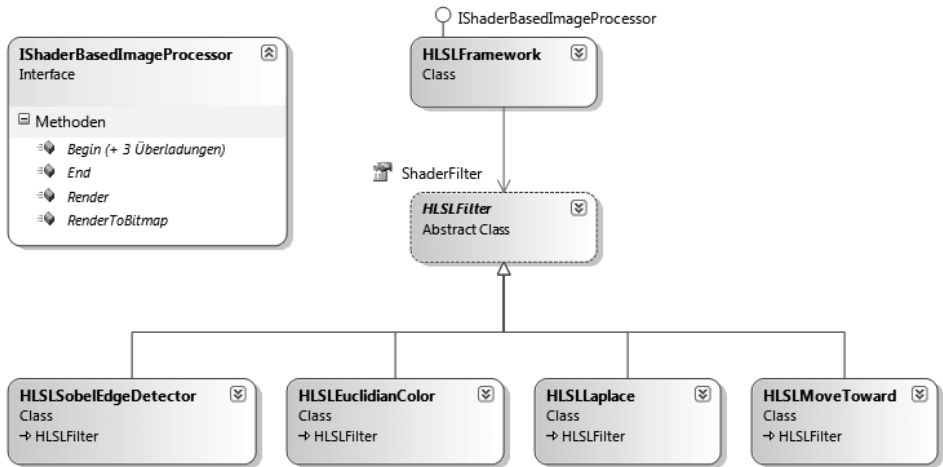


Abbildung 1: Architekturausschnitt von SBIP als UML-Klassendiagramm

SBIP stellt die Schnittstelle `IShaderBasedImageProcessor` bereit. Diese bietet folgende Methoden an:

- `Begin()` - Initialisierung für shaderbasierte Bildverarbeitung
- `End()` - Finalisierung der Verarbeitung und Bereinigen des Speichers
- `Render()` - Online-Rendering in ein Zeichenfenster
- `RenderToBitmap()` - Offline-Rendering in eine Bitmap

Die Schnittstelle erlaubt die Anwendung shaderbasierter Bildfilter. Hierfür muss eine Frameworkkomponente entwickelt werden, welche die Schnittstelle (in der gewünschten Shader-technologie) implementiert. Dabei ist es unerheblich, welche Grafikkarten- und Shader-technologien zum Einsatz kommen.

SBIP enthält zusätzlich eine beispielhafte Komponente namens `HLSLFramework` (mit 20 Beispielfiltern). Diese implementiert die Schnittstelle auf Basis der 3D-API DirectX und ermöglicht die Nutzung HLSL-basierender Bildfilter. Die Komponente enthält eine gleichnamige Klasse, welche die Schnittstelle `IShaderBasedImageProcessor` mithilfe von DirectX implementiert. Weiterführend enthält die `HLSLFramework`-Klasse eine Variable der abstrakten Klasse `HLSLFilter`. Innerhalb der `Render()`- sowie der `RenderToBitmap()`-Funktion wird diese Variable als der Lieferant des zu verarbeitenden Filters verwendet. Selbst entwickelte Shader werden durch das Ableiten von der Klasse `HLSLFilter` in das Framework eingebunden. Die `HLSLFilter`-Klasse besitzt ausschließlich die abstrakte Methode `void SetShader()`. Hier werden die Shader-Quelltextdatei, sowie die von der Anwendung an die Grafikkarte zu übergebenden Variablen gesetzt. Die Architektur von SBIP ist als UML-Klassendiagramm in Abbildung 1 dargestellt.

Tabelle 1: Verarbeitungszeiten verschiedener Filter von AForge.NET und SBIP

	AForge.NET in ms			SBIP in ms		
	Durchschnitt	Min	Max	Durchschnitt	Min	Max
Sobel-Filter	44,11	42	59	14,44	6	24
Convolution-Filter 3x3	419,04	399	605	14,89	12	23
Convolution-Filter 5x5	937,9	857	1067	20,23	15	32

4 Zusammenfassung, Ergebnisse und Ausblick

In diesem Paper wurde ein neuartiges Framework zur echtzeitfähigen, shaderbasierten Bildverarbeitung entwickelt und vorgestellt. Es stellt eine Schnittstelle bereit, mit der Entwickler ein leicht verwendbares Gerüst für echtzeitfähige Bildfilter implementieren können. Ein solches Gerüst wurde auf Basis der 3D-Grafik-API DirectX und der Shader-technologie HLSL umgesetzt. Dieses ermöglicht die Anwendung 20 vorgefertigter Bildfilter oder die Erweiterung des Frameworks um zusätzliche Filter. Für die Implementierung von SBIP wurde die .Net-Sprache C#, sowie Managed-DirectX 9.x verwendet.

Zur Auswertung der Bildverarbeitungszeit wurden die Filtertypen *Sobel-Filter*, *Convolution-Filter via 3x3 Matrix* und *Convolution-Filter via 5x5 Matrix* jeweils 100-mal mit der Bildverarbeitungsbibliothek AForge.NET und dem neu entwickelten SBIP angewandt und die Verarbeitungszeit gemessen. Als Testbild wurde ein 1024 x 768 Pixel großes RGB-Bild verwendet. Die Ergebnisse sind in Tabelle 1 aufgelistet. Sie zeigen deutlich den enormen Geschwindigkeitsvorteil von SBIP im Vergleich zur konventionellen Bildverarbeitung.

Als Ausblick kann die Erweiterung des Frameworks auf DirectX 11 gesehen werden. Dies würde die Nutzung neuer Shadertechnologien, wie dem *Compute-Shader* [com10], ermöglichen. So könnten komplexe, bisher auf Shader-Basis nicht implementierbare Bildfilter, realisiert werden.

Literatur

- [ado09] GPU and OpenGL Features and Limitations in Photoshop CS4. <http://kb2.adobe.com/cps/404/kb404898.html>, 2009.
- [com10] MS DirectX 11. <http://www.microsoft.com/games/en-US/aboutGFW/pages/directx.aspx>, 2010.
- [ima10] ImageJ Features. <http://rsbweb.nih.gov/ij/features.html>, 2010.
- [Kir09] Andrew Kirilov. AForge.NET Framework 2.0.0 BETA. <http://code.google.com/p/aforge/>, 2009.
- [mag10] ImageMagick: Convert, Edit, and Compose Images. <http://www.imagemagick.org/script/index.php>, 2010.

PhotoGet – Bilderbereitstellung einfach gemacht

Tobias Burggraf

Hochschule Deggendorf
Fakultät Betriebswirtschaft
tobias.burggraf@stud.fh-deggendorf.de

Art der Arbeit: Projektarbeit
Betreuer/in der Arbeit: Prof. Dr.-Ing. Herbert Fischer

Abstract: Der nachfolgende Text gibt einen kurzen Überblick über die Software PhotoGet, die eine Verbesserung des Workflows von Fotografen durch moderne Technologien erreicht.

1 Einleitung

PhotoGet ist eine Software, die es Fotografen ermöglicht angefertigte Bilder effizient ihren Kunden bereitzustellen.

2 Aufbau der Software

PhotoGet besteht aus zwei separaten Programmen, die entsprechend ihres Einsatzgebietes auf die Bedürfnisse der Kunden zugeschnitten sind.

2.1 Client- Programm für den Fotografen

Nach Eingabe der Zugangsdaten konfiguriert sich das Client- Programm automatisch und ist sofort einsatzbereit.

In einem intuitiven 3- Schritt Modus ermöglicht PhotoGet dem Fotografen die Bilder bereitzustellen. Nach Eingabe der allgemeinen Kundendaten (z. Bsp. Name, Mailadresse), Festlegung der Albeneigenschaften (z. Bsp. Skin der Weboberfläche, Zugänglichkeit) und Dauer der Verfügbarkeit der Bilder, kann der Unternehmer eine Festlegung der Preise treffen.

Diese werden in sogenannten Preissets gespeichert und können bei Bedarf für weitere Alben verwendet werden. Dabei werden zwei Rabattarten (Rabatt nach Stückzahl mit festen Preisen, Rabatt in Prozentangaben) unterstützt.

Im letzten Schritt können via Drag'n Drop die Bilder hinzugefügt und in einer Vorschau betrachtet werden. Anschließend erfolgt der Upload (siehe Abbildung 1) auf die Webseite des Fotografen und eine Benachrichtigung des Kunden per E- Mail. Manuelle Schritte, die vor Einsatz der Software notwendig waren (z. Bsp. Drehen, Verkleinern der Fotos), werden dabei automatisch von PhotoGet übernommen.

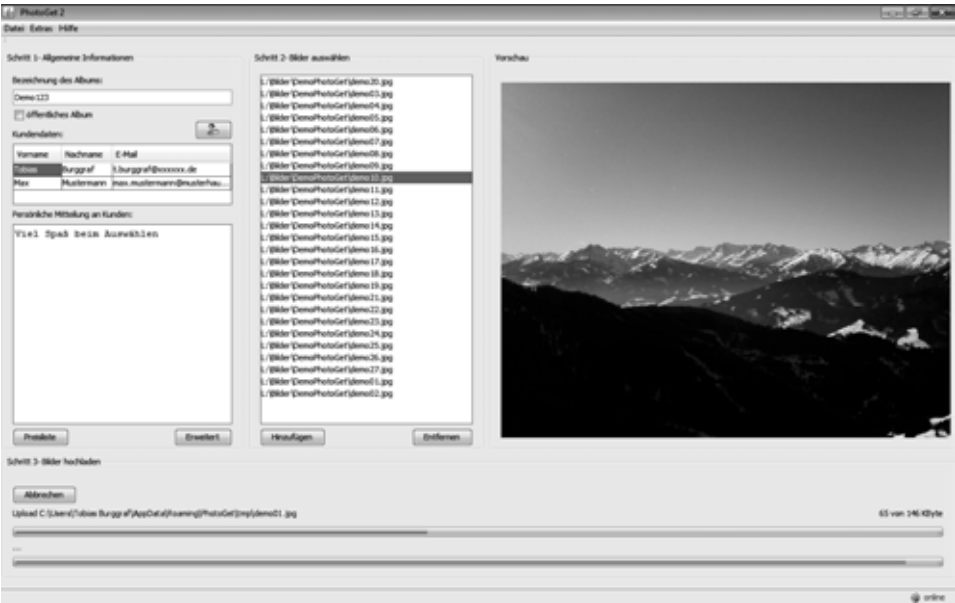


Abbildung 1: Programm zur Bereitstellung der Bilder

2.2 Weboberflächen für den Kunden des Unternehmers

Die Weboberfläche (siehe Abbildung 2) ist vollständig an das Corporate Design der jeweiligen Unternehmung anpassbar und ermöglicht es dem Kunden, nach erfolgreicher Authentifizierung, die für ihn bereitgestellten Bilder anzuschauen und für den Fotografen auszuwählen. Anhand eines simplen, mit SSL verschlüsselten drei Schritt- Systems, wird es dem Kunden ermöglicht die Fotos auszuwählen, Formate und Oberflächen festzulegen (mit Echtzeitberechnung des Preises), anschließend seine Daten einzugeben und zu verifizieren.

Nach Abschluss wird der Fotograf über die Auswahl seines Kunden per Mail informiert.



Abbildung 2: Weboberfläche für den Kunden des Fotografen

3 Technische Implementierung

Die Implementierung des Client- Programms erfolgt auf Grund der notwendigen Plattformunabhängigkeit (MacOS, Windows) in der Programmiersprache Java. Es unterstützt abhängig der Anzahl an CPU-Cores Multithreading, verschiedene Skalierungsverfahren für eine effektive Verkleinerung der Bilder und ist mehrsprachig verfügbar.

Die Weboberfläche wurde in PHP mit MySQL als Datenbankbackend implementiert. Zur Liveberechnung der Preise wird zudem AJAX eingesetzt. Um einen redundanten Datenbestand zu verhindern, wird das XML-RPC Protokoll zur Übertragung der notwendigen Daten zwischen der Weboberfläche und dem Clientprogramm eingesetzt (One-Click-Configuration, Anlegen des Albums).

Literaturverzeichnis

Chaffer, Jonathan, and Karl Swedberg. jQuery Reference Guide. Birmingham: Packt Publishing Ltd., 2007.

—. Learning jQuery. Birmingham: Packt Publishing Ltd., 2007.

Dumbill, Edd, Gaetano Giunta, Miles Lott, Justin R. Miller, und Andres Salomon. XML-RPC for PHP. 2001. <http://phpxmlrpc.sourceforge.net/doc-2/> (Zugriff am 10. 12 2009).

Ullenboom, Christian. Java ist auch eine Insel (8. Auflage). 2009. <http://openbook.galileocomputing.de/javainsel8/> (Zugriff am 15. 12 2009).

Welling, Luke, und Laura Thomson. PHP 5 & MySQL 5. München: Markt+ Technik Verlag, 2007.

Semantische Segmentierung

Björn Fröhlich

Lehrstuhl für Digitale Bildverarbeitung

Friedrich-Schiller-Universität Jena

bjoern.froehlich@uni-jena.de

<http://www.inf-cv.uni-jena.de>

Abstract: Das automatische Erlernen und Erkennen von Objektkategorien und deren Instanzen gehören zu den wichtigsten Aufgaben der digitalen Bildverarbeitung. Aufgrund der aktuell sehr weit fortgeschrittenen Verfahren, die diese Aufgaben beinahe so gut wie ein Mensch erfüllen können, hat sich der Schwerpunkt von einer groben auf eine genaue Lokalisierung der Objekte verlagert. In der vorliegenden Arbeit werden verschiedene Techniken für die Pixel genaue Klassifikation von Bildern, auch *semantischen Segmentierung* genannt, analysiert. Dieses relativ neue Gebiet erweitert die grobe Lokalisierung von Objekten in Bildern um eine punktgenaue Klassifikation. Der Schwerpunkt dieser Arbeit ist es, aktuelle Verfahren der *semantischen Segmentierung* zu vergleichen. Dabei werden verschiedene Methoden zur Ermittlung von Merkmalen, Merkmalstransformationen, Klassifikation und zur globalen Optimierung, wie zum Beispiel durch die Betrachtung von formbasierten Eigenschaften, vorgestellt. Abschließend werden die präsentierten Verfahren in umfangreichen Experimenten auf verschiedenen, frei zugänglichen Datensätzen verglichen und analysiert.

Art der Arbeit: Diplomarbeit
Betreuer: Dipl.-Inf. Erik Rodner
Prof. Dr.-Ing. Joachim Denzler

1 Einleitung

Die Aufteilung von Bildern in mehrere Bereiche und die Zuordnung dieser zu Objektklassen hat in den letzten Jahren stark an Bedeutung gewonnen. Noch vor zehn Jahren war es nahezu unmöglich, die Frage zu beantworten, ob ein Bild eine Instanz einer bestimmten Objektklasse enthält (Bildkategorisierung). Seit neue Methoden dieses lösen können, stellt sich die Frage, wo im Bild sich diese Objekte befinden (Objektlokalisierung). Das Ziel dieser Diplomarbeit war es, den nächsten Schritt in dieser Entwicklung anzugehen: die punktgenaue Lokalisierung von Objekten in Bildern. Hierbei wird für jeden Pixel die Zugehörigkeit zu einer bekannten Objektklasse bestimmt. Dieses Verfahren ist auch bekannt als *semantischen Segmentierung*. Ein Beispiel hierfür wird in Abbildung 1 dargestellt.

Den Inhalt von Bildern automatisch zu erkennen, ist ein weitgehend ungelöstes Problem. Nützlich könnten die Ergebnisse zum Beispiel für eine Online Bildersuche sein, um eine Datenbank von Bildern gezielt auf Objekte und Kombinationen von Objekten zu prüfen.



(a) Eingabebild

(b) Ergebnis der *semantischen Segmentierung*

Abbildung 1: Beispielergebnis der *semantischen Segmentierung*; jede Farbe steht für eine andere Klasse

Zur Zeit wird hierfür meist der Dateiname des Bildes bzw. der textuelle Inhalt der dazugehörigen Internetseite analysiert. Beides spiegelt nicht zwangsläufig den Inhalt eines Bildes wieder.

Ein weiteres Anwendungsgebiet für die *semantischen Segmentierung* ist die automatische Analyse und Auswertung von Satellitenaufnahmen. Für das Projekt OpenStreetMap¹ wäre es zum Beispiel sehr nützlich, wenn Flächen wie Wald, Straßen und Gebäude automatisch punktgenau lokalisiert und diese in eine digitale Karte umgewandelt werden. Zur Zeit wird dies durch die Mitglieder des Projekts manuell umgesetzt.

Weiterhin ist es sinnvoll, für 3-D-Modelle von Städten vorhandene Aufnahmen der Fassaden platzsparend zu speichern. Dies kann nur erreicht werden, wenn ein Modell der Gebäude erzeugt wird, welches wiederum das Wissen über die punktgenaue Lage von Fenstern, Türen und Wand in den Eingabebildern voraussetzt.

In der Diplomarbeit „Semantische Segmentierung“ werden verschiedene Verfahren analysiert. Es werden vorhandene Verfahren vorgestellt und mit neuen Ideen und Algorithmen kombiniert. Abschließend werden diese Methoden getestet, genau ausgewertet und verglichen.

2 Allgemeiner Ablauf

In diesem Abschnitt soll der abstrakte Ablauf eines Algorithmus zur *semantischen Segmentierung* mit lokalen Merkmalen dargestellt werden. Abbildung 2 zeigt die drei erforderlichen und weitere optionale Schritte. Als erstes werden für jedes Bild an verschiedenen Stellen lokale Merkmale bestimmt. Hierfür werden die sogenannten Opponent-SIFT-Merkmale von [vdSGS10] verwendet. Diese Merkmale zeichnen sich durch Invarianz gegenüber Skalierung, Rotation und Beleuchtungsänderungen aus. Anschließend können diese Merkmale in sogenannte High-Level-Merkmale transformiert werden. Csürka et al.

¹<http://www.openstreetmap.org/>

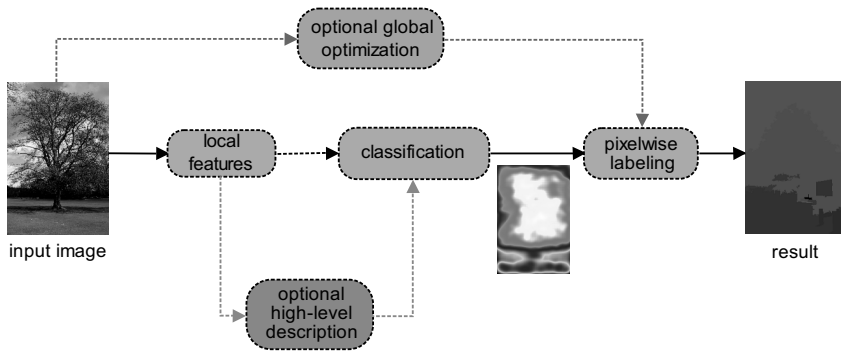


Abbildung 2: Ablaufschema für die *semantischen Segmentierung* mit lokalen Merkmalen ähnlich zu [CP08]

[CP08] schlägt hierfür zwei Verfahren vor: ein Bag-of-Words-Ansatz basierend auf einem Gaußschen Mischmodell (GMM) und dem Fisher-Kernel. Ein Ziel der Diplomarbeit war das Ermitteln eines robust und schnell geschätzten GMM. Zusätzlich wird noch ein alternativer Bag-of-Words Ansatz basierend auf k-Means vorgestellt und mit den beiden anderen Methoden verglichen. Diese High-Level-Merkmale bzw. die lokalen Merkmale werden anschließend klassifiziert. Dies geschieht entweder mit einem Wald aus Entscheidungsbäumen, die randomisiert angelernt werden [Bre01], oder einem Sparse-Logistic-Regression Klassifikator [KH05]. Somit wird für jede Klasse und für jeden Pixel eine Wahrscheinlichkeit ermittelt, welche in sogenannten Wahrscheinlichkeitskarten für jede Klasse dargestellt wird.

Ein weiterer Schwerpunkt der Diplomarbeit ist die Optimierung dieser Ergebnisse durch verschiedene Nachverarbeitungsschritte. Eine zusätzliche Verwendung einer klassischen Segmentierung [CM02] führt zum Beispiel zu signifikant besseren Erkennungsraten. Außerdem wird ein neues Verfahren vorgestellt, welches sich am klassischen Regionen-Wachstum [Ros98] orientiert und anhand von Momenten [Hu62] die Form von Objekten ausnutzt, um das Ergebnis zu verbessern. Um fehlerhafte Regionen zu beseitigen, wird zusätzlich ein auf Graph-Cut basierendes Verfahren [BK04, RVG⁺07] präsentiert. Weiterhin wird gezeigt, wie die typische relative Lage zweier Klassen zueinander (zum Beispiel ist Klasse „Himmel“ immer über Klasse „Gras“) zur Optimierung verwendet werden kann [GRC⁺08].

3 Zusammenfassung

Die verschiedenen Verfahren wurden unter anderem auf den eTRIMS Datensatz [KF09] getestet. Es konnte gezeigt werden, dass die Zufälligen Wälder bessere Erkennungsraten liefern, wenn die Merkmale nicht mit dem Bag-of-Words Ansatz transformiert werden und dass dies beim Sparse Logistic Regression Klassifikator genau anders herum ist. Insgesamt ist die Leistung der Kombination von Bag-of-Words mit Sparse Logistic Regression etwas

besser (ca. 65% auf eTRIMs), als die zufälligen Wälder ohne Bag-of-Words (64%). Dafür sind die zufälligen Wälder signifikant schneller beim Training und beim Testen.

Es konnte gezeigt werden, dass das Graph-Cut basierende globale Optimierungsverfahren Verbesserungen bringt (ca. 2%) in dem es übliche Nachbarschaften ausnutzt. In der Diplomarbeit konnten verschiedene Ansätze zur *semantischen Segmentierung* vorgestellt und weiterentwickelt werden. Es wurde gezeigt, dass unterschiedliche Methoden ähnliche Ergebnisse liefern und dass diese Ergebnisse durch weitere Optimierungsschritte verbessert werden können.

Literatur

- [BK04] Y. Boykov and V. Kolmogorov. An experimental comparison of min-cut/max- flow algorithms for energy minimization in vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(9):1124–1137, 2004.
- [Bre01] Leo Breiman. Random Forests. *Mach. Learn.*, 45(1):5–32, 2001.
- [CM02] D. Comaniciu and P. Meer. Mean shift: a robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5):603–619, 2002.
- [CP08] G. Csurka and F. Perronnin. A Simple High Performance Approach to Semantic Segmentation. In *British Machine Vision Conference*, pages 213–222, 2008.
- [GRC⁺08] Stephen Gould, Jim Rodgers, David Cohen, Gal Elidan, and Daphne Koller. Multi-Class Segmentation with Relative Location Prior. *Int. J. Comput. Vision*, 80(3):300–316, 2008.
- [Hu62] Ming K. Hu. Visual Pattern Recognition by Moment Invariants. *IRE Transactions on Information Theory*, IT-8:179–187, February 1962.
- [KF09] Filip Korč and Wolfgang Förstner. eTRIMS Image Database for Interpreting Images of Man-Made Scenes. Technical Report TR-IGG-P-2009-01, Dept. of Photogrammetry, University of Bonn, March 2009.
- [KH05] Balaji Krishnapuram and Alexander J. Hartemink. Sparse Multinomial Logistic Regression: Fast Algorithms and Generalization Bounds. *IEEE Trans. Pattern Anal. Mach. Intell.*, 27(6):957–968, June 2005.
- [Ros98] Paul L. Rosin. Refining Region Estimates. *International Journal of Pattern Recognition and Artificial Intelligence*, 12(6):841–866, 1998.
- [RVG⁺07] A. Rabinovich, A. Vedaldi, C. Galleguillos, E. Wiewiora, and S. Belongie. Objects in Context. In *Proc. IEEE 11th International Conference on Computer Vision ICCV 2007*, pages 1–8, 14–21 Oct. 2007.
- [vdSGS10] K. E. A. van de Sande, T. Gevers, and C. G. M. Snoek. Evaluating Color Descriptors for Object and Scene Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 99(in press), 2010.

Ein Bildequalizer zur Lehreunterstützung

Clemens Heppner

Uni Hamburg

Sheppner@informatik.uni-hamburg.de, ch510@cam.ac.uk

Abstract: Ein wichtiges Thema in der universitären Lehre der Bildverarbeitung ist die Frequenzraum-Darstellung und -Bearbeitung eines Bildes. Dies ist schwierig zu vermitteln, weil die Studenten sich zu Beginn des Themas nicht auf ihre Anschauung verlassen können und der Zusammenhang zwischen Frequenzraum, Ortsraum und Transferfunktion kompliziert ist. Mit dem in meiner Bachelorarbeit entwickelten Bildequalizer kann dieser Zusammenhang leichter erklärt und Anschauung mit Theorie verknüpft werden. Würden Bildequalizer in der Lehre eingesetzt, könnten die Studenten schneller eine Anschauung der Frequenzraum-Filterung entwickeln und in Eigenregie Frequenzraum-Filterung ausprobieren.

Darüber hinaus beschäftigt sich die Arbeit mit den allgemeinen theoretischen Eigenschaften zweidimensionaler Frequenzfilter, die sich aus Erweiterungs- und 1D-Transferfunktion zusammensetzen lassen. Dabei gelingt es, eine explizite Darstellung aller Erweiterungsfunktionen zu gewinnen, die das Kriterium der Tiefpasserhaltung erfüllen.

1 Einleitung

In der universitären Lehre der Bildverarbeitung stellt der Zusammenhang zwischen Fouriertransformation, Faltung und frequenzabhängiger Filterung meist einen Kernbereich der Vorlesungen und Lehrbücher dar ([McA03], [NH04], [SHB08]). Dieser Zusammenhang ist zu Beginn des Themas schwer zu erklären, da ein Bild in einer zunächst unbekannten Domäne (im Frequenzraum) auftritt und dort bearbeitet wird, sodass die Auswirkungen auf das Bild in der Ursprungsdomäne nicht intuitiv ersichtlich sind.

Um ein anschauliches Verständnis dieser Zusammenhänge zu fördern wäre es hilfreich, die Filterung interaktiv zu demonstrieren ([Bam94], [McA03]). Dazu ist die eine Software nötig, die die Eingabe theoretischer Parameter ermöglicht aber zugleich eine anschauliche Interpretation erlaubt. Vorhandene Software (z.B. [RLWB02], [IN08]) ist dazu nur eingeschränkt geeignet, da sie auf jeweils andere Anwendungsgebiete zugeschnitten.

Den Ausgangspunkt für meine Untersuchungen ([Hep09]) bildet die Oberfläche eines klassischen Audioequalizers, da diese sich beim (theoretischen) Verständnis von Klangfrequenzen als hilfreich erwiesen haben. Ich untersuchte praktisch und theoretisch, wie sich die Konzepte eines solchen Equalizers auf einen Bildequalizer übertragen lassen. Die entwickelte Anwendung lässt sich grundsätzlich auch für zweidimensionales Filterdesign und kreative Bildbearbeitung nutzen.

2 Bildequalizer

Ein Bildequalizer für die Lehre sollte das gefilterte Bild, die Transferfunktion und die Bildfrequenzen gleichzeitig darstellen, um deren Zusammenhang zu veranschaulichen. In meiner Arbeit werden darüber hinaus drei Anforderungen an den Bildequalizer formuliert, die sich bei Audioequalizern als hilfreich herausgestellt haben:

- Spektrum und Transferfunktion sollen im selben Koordinatensystem dargestellt werden, da dies die Zusammenhänge von Transferfunktion und Frequenz verdeutlicht.
- Die Parameter der Transferfunktion sollen sich direkt manipulieren lassen (z.B. mit „Anfassern“), da dies die Benutzung vereinfacht.
- Die Transferfunktion soll als Komposition verschiedener einfacher Filterbänder entworfen werden können, da diese jeweils einfach zu beschreibende Auswirkungen auf das Bild und eine mathematische Beschreibung haben.

Abbildung 1 zeigt zwei Bildequalizer, die in der Industrie¹ (siehe Abb. 1(a)) bzw. Wahrnehmungsforschung (siehe Abb. 1(b) [IN08]) entwickelt wurden. Im Unterschied zum Audioequalizer² in Abbildung 2(a) erfüllt keines der Programme die Anforderungen, obwohl alle relevanten Dinge visualisiert werden. Dass diese Anforderungen mit einer durchdachten Benutzeroberfläche relativ einfach umzusetzen sind, zeigt der Screenshot der von mir entwickelten Anwendung (siehe Abb. 2(b)): Unter dem Bild befindet sich die Anzeige des modifizierten Spektrums und der Transferfunktion mit „Anfassern“. Detaillierte Einstellungen der Filterbänder lassen sich rechts neben dem Bild vornehmen.

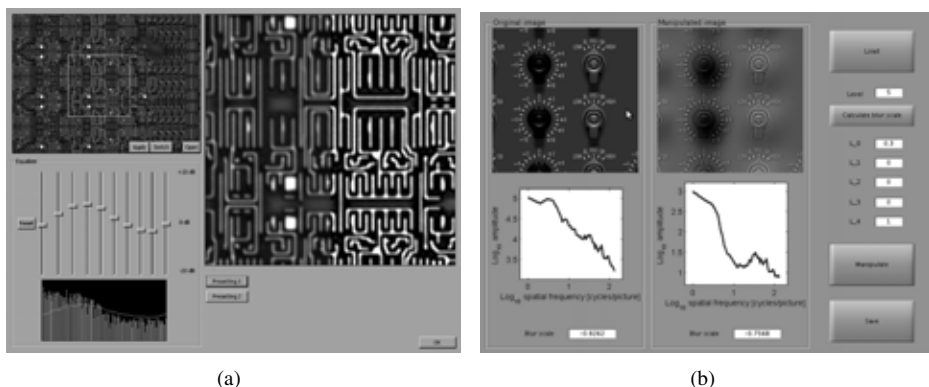


Abbildung 1: Zwei Bildequalizer, die nicht für die Lehre konzipiert wurden.

¹Quelle: <http://www.jentech.de>, 7. Januar 2010.

²Quelle: <http://www.aixcoustic.com/>, 7. Januar 2010.

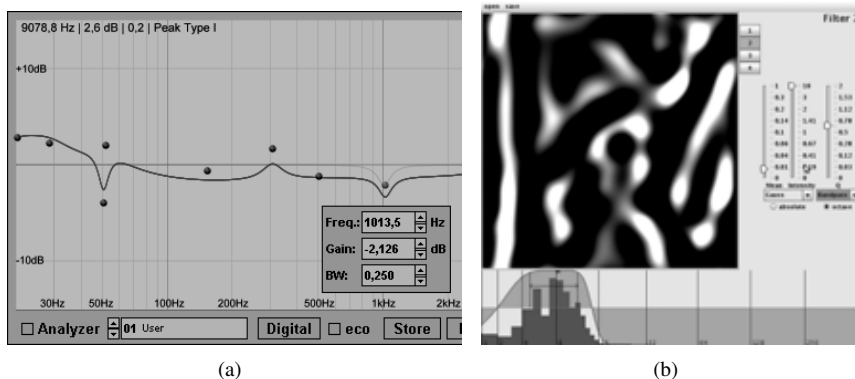


Abbildung 2: Der entwickelte Bildequalizer und ein Audioequalizer.

3 Einsatz in der Lehre

Der entwickelte Bildequalizer lässt sich in der Lehre vielfältig einsetzen, was von verschiedenen Lehrenden bestätigt wurde. Empirische Untersuchungen dazu waren jedoch nicht Gegenstand meiner Arbeit.

Ich liefere jedoch einige Ideen für „Experimente“, die mit dem Equalizer durchgeführt werden können. Diese können verwendet werden, um verschiedene Eigenschaften der zweidimensionalen Fouriertransformation mit Bezug auf die Bildbearbeitung zu demonstrieren. Sie bilden damit eine Anleitung für den praktischen Einsatz in der Lehre.

Diese Experimente können sowohl vom Dozenten, als auch von Studenten (z.B. in Übungen) durchgeführt werden und folgen in idealisierter Form dem Schema von Durchführung, Beobachtung und Interpretation, das aus anderen Naturwissenschaften bekannt ist. Das folgende Beispiel verdeutlicht dies anhand der Unschärferelation:

Der Dozent behauptet, dass Orts- und Frequenzauflösung insofern beschränkt sind, dass eine große Ortsauflösung nur bei einer geringen Frequenzauflösung möglich ist. Zur Veranschaulichung wählt er einen Bandpass und variiert dessen Bandbreite. Eine gute Lokalisation im Frequenzbereich führt automatisch zu einer schwächeren Lokalisation im Bild: Es lässt sich bei guter Frequenzselektion nicht mehr so genau sagen, an welchen Orten die Frequenzen im Bild auftreten.

4 Theoretische Ergebnisse

Da in dem Bildequalizer nur eine eindimensionale Transferfunktion dargestellt wird, muss diese zunächst auf eine zweidimensionale Transferfunktion erweitert werden, um auf ein Bild angewendet werden zu können.

Eine Möglichkeit dafür besteht darin, eine richtungsunabhängige 2D-Transferfunktion di-

rekt aus der 1D-Transferfunktion zu erzeugen: $T_{2D}(u, v) = T(\sqrt{u^2 + v^2})$. Die Impulsantwort des zweidimensionalen Filters ergibt sich dann mit Hilfe der Hankel-Transformation ([Wei09]): $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{i(ux+vy)} T_{2D}(u, v) dv du = \int_0^{\infty} r T(r) J_0(r \sqrt{x^2 + y^2}) dr$ (J_0 ist die Besselfunktion nullter Ordnung). Der Bezug zur Hankel-Transformation zeigt, dass viele Eigenschaften des eindimensionalen Filters auf T_{2D} übertragbar sind.

Allgemein kann eine Erweiterungsfunktion B verwendet werden, sodass $T_{2D}(u, v) = T(B(u, v))$ gilt. Da nicht jede Erweiterungsfunktion anschauliche Eigenschaften der eindimensionalen Transferfunktion erhält, untersuchte ich das allgemeine Verhalten von Erweiterungsfunktionen anhand verschiedener Kriterien. Für das Kriterium der Tiefpasserhaltung³ konnte ich die Lösungsmenge in einer einfachen Gleichung formulieren: $B(u, v) = \beta_1 u^2 + \beta_2 v^2 + f(v/u)$ (für beliebige Funktionen f). Mit dieser Formel lassen sich viele der üblichen Filterformen (z.B. radialsymmetrisch, elliptisch und fanförmig) darstellen und anschaulich aus der eindimensionalen Funktion ableiten.

Literatur

- [Bam94] R.H. Bamberger. Portable tools for image processing instruction. In *Image Processing, 1994. Proceedings. ICIP-94., IEEE International Conference*, volume 1, pages 525–529 vol.1, Nov 1994.
- [Hep09] Clemens Heppner. Ein Bild-Equalizer zur Unterstützung der Lehre der Bildverarbeitung. Bachelor's thesis, Uni-Hamburg, September 2009.
- [IN08] Yugo Imazumi and Shigeki Nakauchi. Blur adaptation to spatial frequency structure. *Journal of Vision*, 8(17):60–60, 12 2008.
- [McA03] Alasdair McAndrew. Teaching image processing using minimal mathematics. In *ACE '03: Proceedings of the fifth Australasian conference on Computing education*, pages 15–23, Darlinghurst, Australia, 2003. Australian Computer Society, Inc.
- [NH04] Alfred Nischwitz and Peter Haberäcker. *Masterkurs Computergrafik und Bildverarbeitung*. Vieweg Verlag, 1. edition, 2004.
- [RLWB02] Alexander Roth, Hans-Gerd Lipinski, Martin Wiemann, and Dieter Bingmann. Ein Image-Equalizer für die Bearbeitung von histologischen Bilddaten. In *Bildverarbeitung für die Medizin*, pages 235–238, 2002.
- [SHB08] Milan Sonka, Vaclav Hlavac, and Roger Boyle. *Image Processing, Analysis, and Machine Vision*. Thomson-Learning, 3. edition, 2008.
- [Wei09] E. W. Weisstein. Hankel Transform. MathWorld – A Wolfram Web Resource, <http://mathworld.wolfram.com/HankelTransform.html>, 2009.

³Informell: Jeder 1D Tiefpass- wird auf einen 2D-Tiefpassfilter abgebildet.

Umsetzung eines 3D-Action-Adventures mit der Open Source 3D-Grafik-Engine Irrlicht

Maximilian Hönig

Martin-Luther-Universität Halle-Wittenberg
Naturwissenschaftliche Fakultät III – Institut für Informatik

squarefox@gmx.net

Art der Arbeit: Bachelorarbeit
Betreuer der Arbeit: Dr. Peter Schenzel

Abstract: Diese Arbeit befasst sich mit der Umsetzung eines Computerspiels, welches dem Genre des 3D-Action-Adventures angehört. Der Schwerpunkt dieser Arbeit liegt im Bereich Computergrafik. Kriterien wie Echtzeit, Detailtreue und Interaktionsmechanismen sind im besonderen Fokus.

1 Einleitung

Computerspiele sind das neue Unterhaltungsmedium des 21. Jahrhunderts. Im Gegensatz zu anderen Medien vor ihnen wird die erzählte Geschichte hier aktiv erlebt. Bei der Entwicklung eines Computerspiels sollte, wie auch bei anderen Softwareprojekten, eine gut strukturierte Vorgehensweise verwendet werden, damit es im späteren Projektverlauf nicht zu schwer überwindbaren Problemen kommt.

Bei diesem Projekt wurde als Vorgehensmodell ein angepasstes Wasserfallmodell nach [B01] verwendet, wie es in Abbildung 1 zu sehen ist. Der Vorteil dieses Modells liegt an einem einfach verständlichen und übersichtlichen Top-Down-Vorgehen bei der Planung und Realisierung.

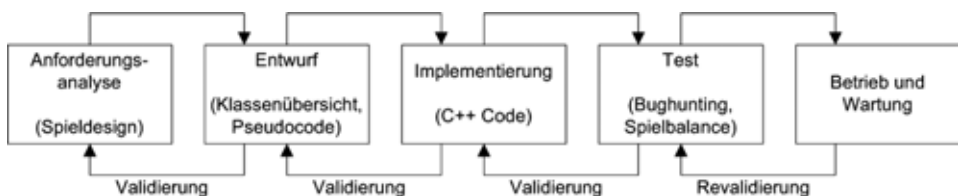


Abbildung 1: Das Wasserfallmodell für die Computerspielentwicklung

2 Spieldesign

Das Spieldesign wird typischerweise in einem Spieldesigndokument festgehalten, welches einer Art Pflichtenheft für Computerspiele entspricht. In diesem wird das Spiel steuerungs-, handlungs- und spielmechanisch definiert.

Das Spiel, welches durch diese Arbeit entstand, trägt den Namen „Moonork Adventures“. Es handelt sich um ein 3D-Action-Adventure [S01] aus einer Third-Person-Perspektive, wie sie in Abbildung 2 zu sehen ist.

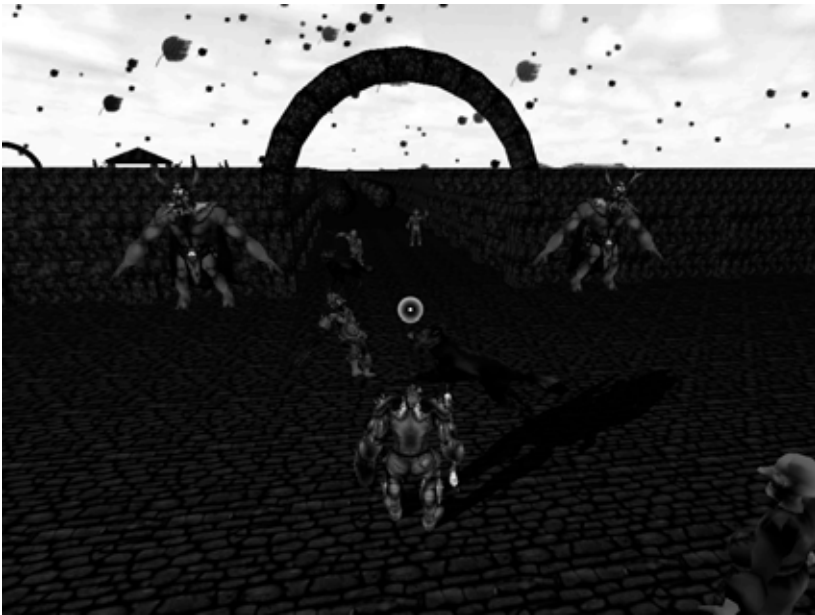


Abbildung 2: Die Third-Person-Perspektive in „Moonork Adventures“

Die Steuerung erfolgt mittels Maus und Tastatur. Ziel des Spieles ist es, den Stein der Nacht zu finden und zu bannen. Dieser befindet sich im uralten Tempel der Zeit. Die Schwierigkeit liegt darin, die zahlreichen Fallen im Tempel zu überwinden, um zum Stein der Nacht zu gelangen.

3 Umsetzung

3.1 Die Werkzeuge

Bei der Entwicklung kamen, neben der Programmierumgebung Visual C++ 2005, drei wesentliche Werkzeuge zum Einsatz: Irrlicht, IrrEdit und IrrKlang. Bei Irrlicht [IrL01] handelt es sich um eine Open Source 3D-Grafik-Engine. Diese ist hervorragend doku-

mentiert, besitzt eine große Community und beherrscht alle nötigen State-of-the-Art Funktionen, wie z. B. einfaches Szenenmanagement oder Kollisionsabfragen. Weiterhin verfügt Irrlicht über viele grafische Effekte sowie eine große Palette an verwendbaren 2D- und 3D-Grafikformaten. IrrEdit [IrE01] ist ein kostenloser Szeneneditor für in Irrlicht einlesbare 3D-Szenen. Das Erstellen von Szenen funktioniert hier ähnlich wie bei einem Leveleditor oder einem 3D-Modelling-Programm.

IrrKlang [IrK01] ist eine für nichtkommerzielle Nutzung kostenlose Sound-Engine, die viele Soundformate versteht und 2D- sowie 3D-Sound abspielen kann. Zusätzlich ist der Einsatz verschiedener Soundeffekte möglich. Der Einsatz von IrrKlang beschränkt sich bei diesem Projekt allerdings auf das Abspielen von Hintergrundmusik im .mp3-Format.

3.2 Programmstruktur

Das “Moonork Adventures”-Programm wurde streng objektorientiert programmiert und ist in mehrere Layer unterteilt. Objekte des Layer $n+1$ werden stets von Objekten des Layer n verwaltet, soweit $n \geq 0$. Diese Layer sind:

Layer 0: (Programmkern)	Hauptroutine, Eventverarbeitung, Optionen
Layer 1: (Levelverwaltung)	Levelobjekt, welches alle Objekt in der Spielwelt verwaltet
Layer 2: (Manager)	Einzelne Objektmanager für die Spielwelt, wie z. B. Spielerobjekt (Steuerung und Animation der Hauptfigur und Kamera), Felskugelfallenmanager (verwaltet Felskugelfallen-Objekte)
Layer 3: (Einzelobjekte)	Einzelne Objekte der Spielwelt, z. B. Felskugelfalle, Stachelfalle, Blitzfalle

Um das Kriterium der Echtzeit, trotz der großen Anzahl von Interaktionsobjekten, zu gewährleisten, ist eine sinnvolle Verwaltung dieser Objekte und ihrer Aktionen nötig. Die Spielwelt wird dazu linear in Gebiete unterteilt. Bei dieser Unterteilung wird darauf geachtet, dass man nur in jeweils benachbarte Gebiete sehen kann. Die Gebiete werden aufsteigend durchnummeriert. Jedes Interaktionsobjekt bestimmt nun per Fallunterscheidung, welche Aktionen auszuführen sind.

Dabei gibt es drei Aktionslevel. Globale Aktionen sind Aktionen, die immer ausgeführt werden. Hierzu gehört z. B. die Aktualisierung von Count-Down-Timern. Benachbarte Aktionen werden ausgeführt, wenn sich die Hauptfigur in einem benachbarten Gebiet befindet. Dazu zählen beispielsweise Animationen. Lokale Aktionen werden nur ausgeführt, wenn sich die Hauptfigur im gleichen Gebiet wie das Interaktionsobjekt befindet. Dies betrifft hauptsächlich Kollisionsabfragen.

3.3 Grafische Effekte

Irrlicht bietet eine hohe Anzahl grafischer Effekte. Viele davon werden im „Moonork Adventures“-Programm genutzt. Dadurch kann der Grad der Detailtreue erhöht werden. Dabei muss allerdings darauf geachtet werden, dass Detailtreue und Echtzeit prinzipiell im Widerspruch zueinander stehen. Als 3D-Grafik-Engine bietet Irrlicht bereits eingebaute Beschleunigungsstrukturen, die automatisch verwendet werden. Allerdings ist es trotzdem erforderlich, beim Einsatz der Effekte auf die Performanz zu achten.

Zu den verwendeten Effekten gehören beispielsweise Partikeleffekte für verschiedene Arten des Feedbacks. Auch animierte Texturen und animierte Oberflächenstrukturen werden genutzt, z. B. für die Darstellung von Portalen. Um bestimmte Strukturen realistischer darzustellen, wird Parallax Mapping verwendet, welches glatten Flächen eine unebene, raue Oberfläche verleiht. Die Hauptfigur ist mit einem Stencil Schatten versehen. Für die restlichen Objekte der Szene wird allerdings auf Grund der Performance auf den Einsatz solcher Schatten verzichtet.

3.4 Datenablage

Die große Anzahl von Objekten bringt auch das Problem der Datenablage mit sich. Die Daten im Quellcode abzulegen, wäre ineffizient. Deshalb werden die Objektdaten direkt aus der in IrrEdit erstellten Szenendatei gewonnen. Jedem Objekt ist eine ID-Nummer zugeordnet. Mit Hilfe dieser wird das Objekt in der Szene wiedergefunden. Danach kann mit seinen Daten, wie z. B. der Position, das konkrete Programmobjekt erstellt werden.

Literaturverzeichnis

- [S01] Saltzmann, Marc: Game Design – Die Geheimnisse der Profis. Markt & Technik Verlag, 1999, ISBN 3-8272-9075-9.
- [B01] Balzert, Helmut: Lehrbuch der Softwaretechnik. Spektrum Akademischer Verlag, 1998, ISBN 978-3827403018.
- [IrL01] Irrlicht Team: Irrlicht Engine – Features.
Version: 20.02.2010
<http://irrlicht.sourceforge.net/features.html>
- [IrE01] Ambiera: IrrEdit Features.
Version: 20.02.2010
<http://www.ambiera.com/irredit/features.html>
- [IrK01] Ambiera: IrrKlang Features.
Version: 20.02.2010
<http://www.ambiera.com/irrklang/features.html>

Modellbasierte Poseschätzung aus dichten Volumendaten

Benjamin Knopp

bknopp@uni-koblenz.de

Kurzbeschreibung

Die Zahl der Szenarien, in welchen die Beziehung Mensch - Computer nicht mehr auf die eines Instruierenden und einer ausführenden Maschine beschränkt ist, steigt innerhalb der letzten Jahre kontinuierlich.

In unterschiedlichsten Einsatzgebieten ist zudem ein genaues Erfassen und Verstehen des menschlichen Bewegungsapparates (Motion-Capturing) erforderlich. So werden im medizinischen Kontext häufig anhand von Videosequenzen aufgezeichnete Bewegungsabläufe auf Gelenkfehlstellungen hin untersucht, während im Augmented-Reality Bereich sowie in der Robotik die Übertragbarkeit menschlicher Bewegungen auf virtuelle Modelle oder Roboter im Vordergrund steht. Auch im Bereich der Mensch-Maschine Interaktion spielt Gesten- und Bewegungserkennung eine zunehmend bedeutendere Rolle. Im Rahmen dieser Arbeit wird ein Verfahren zur Rekonstruktion der menschlichen Pose¹ vorgestellt, das auf einem modellbasierten, probabilistischen Ansatz beruht. Hierbei finden volumetrische Datensätze Verwendung, welche die zu rekonstruierende Position über die Zeit enthalten.

keywords: *model-based tracking, probalistic reconstruction, particle-filter, partitioned sampling, annealing, unconstrained motion, volumetric representation*

Stand der Technik

In vielen der eingangs erwähnten Fällen entspricht die Verwendung von Marker basierten Systemen dem Stand der Technik. Hierbei werden leicht interpretierbare Messpunkte am Körper befestigt, welche z.B. optisch oder elektrisch ausgelesen werden können. Dieses Vorgehen beschränkt die Anwendbarkeit auf im Vorfeld entsprechend präparierte Aufnahmesequenzen und führt zu einer Abhängigkeit von komplexen Hardware / Software-Systemen. Weiterhin sind die erzielten Ergebnisse von der exakten Positionierung der verwendeten Marker abhängig. Steigende Rechenleistung und die zunehmende Erschließung von Szenarien, in welchen die Anbringung von Markern nicht erwünscht oder technisch umsetzbar ist, lassen Verfahren zum markerlosen Tracking attraktiver erscheinen. Das allgemeine Vorgehen besteht nun darin, ein parametrisierbares Modell des menschlichen

¹Pose = Position + Orientierung

Körpers mit Beobachtungen im Bildraum derart zur Übereinstimmung zu bringen, dass die Unterschiede zwischen Modell und Beobachtung minimiert werden. Für die Modellierung des menschlichen Körpers wird häufig auf die Simulation von Starrkörperketten zurück gegriffen, was sich zum Quasi-Standard in der hier beschriebenen Domäne entwickelt hat. Die erzielte Genauigkeit hängt hierbei im Wesentlichen davon ab, wie präzise die gewählte Starrkörper-Repräsentation das zugehörige Körperteil annähert.

Sowohl in [ZH04] als auch [Gav99] werden Taxonomien erstellt, welche bestehende Tracking-Algorithmen anhand der von ihnen gemachten Annahmen sowie der zugrunde liegenden Modelle unterscheiden. Prinzipiell lassen sich die Methoden zu Pose-Rekonstruktion sowohl nach ihrem zugrunde liegenden Körper- sowie Vorhersagemodellen und der Dimensionalität selbiger Modelle unterscheiden. Zunehmend werden Verfahren entwickelt, welche verschiedene Ansätze kombinieren und somit versuchen verfahrensspezifische Schwächen zu eliminieren.

Eigener Ansatz: Modellbasiert, Volumetrisch, Probabilistisch

Der in dieser Arbeit gewählte Ansatz verfolgt das Ziel sowohl die Pose-Schätzung als auch die zu verwendende Abstandsfunktion vollständig im dreidimensionalen Raum anzusiedeln.

Die hierfür notwendigen 3D-Volumendaten können auf verschiedenste Art und Weise generiert werden. Während die verwendeten Testdaten durch ein Mehrkamerasystem gewonnen werden, ist auch der Einsatz von Laserscanner und den immer populärer werdenden *time-of-flight* Kameras möglich. Da auch bei entsprechend hoher zeitlicher Auflösung eine deterministische Rekonstruktion komplexer Bewegungsabläufe - ohne den Bewegungsablauf einschränkende Annahmen - kaum zu gewährleisten ist [DBR09], wird im Rahmen dieser Arbeit ein probabilistischer Ansatz verfolgt.

Das verwendete Körpermodell basiert auf [Ste09] und erlaubt eine genaue anatomisch korrekte Rekonstruktion, sowohl der auftretenden Gelenkwinkel, als auch der hieraus resultierenden Kräfte. Es liefert eine wirklichkeitsgetreue Modellierung der verschiedenen menschlichen Gelenktypen (Drehgelenke, Kugelgelenke etc.), welche die zugehörigen Bewegungsräume exakt abbildet und ermöglicht daher eine intuitive räumliche Ausrichtung. Um das Modell mit einem Volumendatensatz vergleichen zu können, ist zusätzlich eine Beschreibung der Volumina einzelner Starrkörper notwendig. Üblicherweise werden hierfür einfache geometrische Beschreibungen wie Zylinder oder polygonale Netze verwendet. Durch die dichte Rekonstruktion² ist es möglich, auf dreidimensionale Überlappungsmaße zurück zu greifen und die volumetrische Überschneidung zu messen. Neben der etablierten Repräsentation einzelner Körpersegmente durch geometrische Primitive soll eine Segmentierung der Voxeldaten³ zu einer genaueren Annäherung der Körpersegmente verwendet werden. Ein möglichst allgemeingültiges Menschmodell, welches die Freiheitsgrade der entsprechenden Körperteile wirklichkeitsgetreu abbildet, besitzt eine derart hohe Dimensionalität, dass eine globale Betrachtung des hieraus resul-

²im Vergleich z.B. zur konvexen Hülle, wird der Körper als durchgehend gefüllt betrachtet

³dreidimensionale Entsprechung eines Pixels

tierenden Suchraumes wenig plausibel erscheint. Da die Aufenthaltswahrscheinlichkeit nicht für jede denkbare Position ermittelt werden kann, werden die Modi der zugrundeliegenden Verteilungsfunktion mittels eines Partikelfilters [IB98] ermittelt. Die Anzahl der benötigten Partikel und die Dimension des betrachteten Suchraums stehen in einem direkten Zusammenhang, so dass im konkreten Fall eine harte Partitionierung des eigentlichen Raumes in Subräume niedriger Dimensionalität vorgenommen wird, was eine deutliche Reduktion der benötigten Partikel ermöglicht [BB08]. Hieraus resultiert eine hierarchische Suche der optimalen Modellposition, wobei der Einfluss unterschiedlicher Traversierungsreihenfolgen zu untersuchen ist. (So führt z.B. eine unzureichende Ausrichtung des Rumpfes zwangsläufig auch zu einer fehlerhaften Ausrichtung der Arme etc.) Innerhalb der erzeugten Partitionen werden die einzelnen Hypothesen weiter verbessert, in dem relevante Partikel mittels geeigneter Optimierungsfunktion [Gal05] [GPS⁺06] in Richtung der Maxima der Aufenthalts-Wahrscheinlichkeitsfunktion bewegt werden.

Die Übereinstimmung zwischen Modell und Messwerten, welche sich aus der jeweiligen Hypothese ergibt, wird durch den Schnitt der zugehörigen Volumina ermittelt. Die ermittelten Ergebnisse werden durch *simulierte Abkühlung* (simulated annealing) [DBR09] [KGV07] weiter optimiert.

Zum einen bietet der gewählte Ansatz den Vorteil, dass Ambiguitäten, wie sie durch die Auswertung in einer geringen Anzahl von 2D Ansichten auftreten können [CH04], vermieden werden, zum anderen bietet der Vergleich dichter volumetrischer Daten die Möglichkeit, auf die fehleranfällige Betrachtung von Silhouetten oder Kanten zur Ähnlichkeitsbestimmung verzichten. Sogar eine individuelle Gewichtung einzelner Regionen, z.B. an Gelenkpunkten, ist durch die einfache Faltung der Segmentrepräsentationen mit einer entsprechenden Gewichtungsmatrix umsetzbar.

Evaluation und Ausblick

Zur Evaluation der rekonstruierten Bewegungsabläufe wird auf Aufnahmesequenzen zurückgegriffen, welche unter zusätzlicher Verwendung optischer Marker angefertigt wurden. Während die hierdurch erhaltenen Zusatzinformationen für den eigentlichen Rekonstruktionsprozess unberücksichtigt bleiben, bilden die mittels eines VICON©-Systems ausgewerteten Markerpositionen den Goldstandard. Die genaue Kenntnis der absoluten Gelenkpositionen sowie auch der jeweiligen Gelenkwinkel erlaubt eine fundierte abschließende Beurteilung der erzielten Ergebnisse. Neben dem Aspekt, inwiefern sich der Suchaufwand durch die Integration von Vorwissen reduzieren lässt, werden auch die Auswirkung verschiedener Abstandsmaße und Volumenmodelle auf die geometrische Genauigkeit der Ergebnisse betrachtet.

Literatur

- [BB08] Jan Bandouch und Michael Beetz. Tracking Humans Interacting with the Environment Using Efficient Hierarchical Sampling and Layered Observation Models, 2008.
- [CH04] Fabrice Caillette und Toby Howard. Real-Time Markerless Human Body Tracking Using Colored Voxels and 3-D Blobs. Bericht, Advanced Interfaces Group, University of Manchester, 2004.
- [DBR09] Jonathan Deutscher, Andrew Blake und Ian Reid. Articulated Body Motion Capture by Annealed Particle Filtering. Bericht, University of Oxford, Microsoft Research, 2009.
- [Gal05] Jürgen Gall. Generalised Annealed Particle Filter - Mathematical Framework, Algorithms and Applications. Diplomarbeit, Universität Mannheim, 2005.
- [Gav99] D. M. Gavrilu. The Visual Analysis of Human Movement: A Survey. *Computer Vision and Image Understanding*, 73:82–98, 1999.
- [GPS⁺06] Jürgen Gall, Jürgen Pothoff, Christoph Schnörr, Bodo Rosenhahn und Hans-Peter Seidel. Interacting and Annealing Particle Filters: Mathematics and a Recipe for Applications. Bericht, Max-Planck-Institut Saarbrücken, Universität Mannheim, 2006.
- [IB98] Michael Isard und Andrew Blake. CONDENSATION—Conditional Density Propagation for Visual Tracking. *International Journal of Computer Vision*, 29:5–28, 1998.
- [KGV07] S. Kirkpatrick, C. D. Gelatt und M. P. Vecchi. Optimization by Simulated Annealing. *Science, New Series*, 220:671–680, 2007.
- [Ste09] Günther Stelzner. Zur Modellierung und Simulation biomechanischer Mehrkörpersysteme. *Schriftenreihe des Instituts für Technische Mechanik*, 11, 2009.
- [ZH04] Huiyu Zhou und Huosheng Hu. A Survey - Human Movement Tracking and Stroke Rehabilitation. Bericht, University of Essex, 2004. ISSN 1744 - 8050.

Efficient Game State Management für Massive Multiplayer Online Games

Max Marklein

marklein@informatik.uni-bonn.de

Abstract: Mit der steigender Anzahl von Teilnehmern wachsen die Anforderungen an das zentrale Game State Management von Multiplayer Online Games wie z.B. Second Life. Dieser Artikel skizziert eine exemplarische Serverarchitektur und beschreibt darauf aufbauend Ansätze zur Optimierung des Game State Management.

1 Einleitung

Massive Multiplayer Online Games, *MMOGs*, sind Online-Spiele für eine große Anzahl an Benutzern. Um so eine große Anzahl an Teilnehmern zu verwalten, müssen die knappen Hardware- und Netzwerkressourcen effizient genutzt werden.

Die erfolgreiche Umsetzung eines solchen Spiels ist mit Second Life gelungen, dort sind im Januar 2010 Besitzer von mehr als 1.000.000 Accounts [SL10] in einer virtuellen Welt aktiv, gleichzeitig miteinander spielen bis zu 60.000 Clients. Die Faszination des Spiels ergibt sich daraus, dass der Benutzer die Möglichkeit hat sich mit seinem *Resident* in einer virtuellen Umgebung bewegen und interagieren zu können, die zumindest in Teilbereichen der Komplexität der realen Welt nahe kommt. Spieler haben die Möglichkeit Objekte und Skripte zu erstellen, darüber hinaus können auch eigene Texturen in das Spiel geladen werden. Welche Auswirkungen und Herausforderungen dies mit sich bringt, wird im nächsten Abschnitt thematisiert.

2 Problemstellungen bei dem Design von MMOGs

Bei vielen Spielen, die über einen Mehrspielermodus verfügen, wird die gesamte Szene in einem lokalen Szenengraph vorgehalten, welcher von der Grafikengine gerendert wird. Wird ein solches online miteinander gespielt, dann werden die Eingaben zunächst in die lokale Szene eingebaut und anschließend mit den Mitspielern synchronisiert. [AS09]. Oftmals ist aufgrund eines hohen Spieltempos der Zeitfaktor entscheidend, so dass an dieser Stelle eine möglichst unmittelbare Kommunikation zwischen den Clients gefordert ist.

Ein solches Netzwerkdesign ist im Kontext eines MMOGs, aus zweierlei Gründen nicht möglich. Denn aufgrund des dadurch entstehenden Daten-Aufkommens ist direkte Kom-

munikation bei hohen Benutzerzahlen für den einzelnen Benutzer unmöglich. Zweitens, würde der Synchronisations-Ansatz lokaler Szenengraphen aufgrund der viel größeren virtuellen Welt scheitern.

Bei MMOGs werden daher Eingaben über einen Server geleitet, der den aktuellen Spielstand verwaltet. Hier können Benutzereingaben in die Simulation eingebunden werden und die daraus resultierenden Änderungen werden nur an die daran interessierten Spieler verschickt. Die limitierende Größe in diesem Entwurf ist oftmals das Netzwerk, das mit hohen Benutzerzahlen schnell stark beansprucht wird.

Darüber hinaus bringt die Vielfalt an von Nutzern gestalteten Objekten mit sich, dass sich Inhalte mit einer lokalen Installation nicht vorab bereit stellen lassen. Die für jeweils relevanten Objekte der Welt, müssen dem Client demnach kurzfristig über das Netzwerk gesendet werden.

An dieser Stelle sei bemerkt, dass die Vorstellung von *einem* Server im Kontext von MMOGs stark vereinfachend ist. Bei zehntausenden Benutzern wird der Dienst von einem Rechnerverbund bearbeitet, womit sich bei deren Synchronisierung auch eine Vielzahl von Problemstellungen ergeben, diese sollen hier aber vernachlässigt werden. Damit wird in Abschnitt 3 eine Architektur vorgestellt, die sämtliche notwendige Module zur Bereitstellung einer virtuellen Welt auf einem einzelnen Rechner bereitstellt.

3 Serverseitige Architektur

Im Folgenden werden die einzelnen Module eines Servers zur Bereitstellung eines Second Life ähnlichen Dienstes kurz vorgestellt (siehe Abb.1, vgl. [A⁺03] [AW⁺05]). Die Verbindung mit den Spielern wird in dem Netzwerkmodul vorgehalten. Nachrichten zur Spielsteuerung und zu der Clientverbindung werden über ein spezielles Protokoll ausgetauscht. Hier wird bspw. unter Auswertung der Verbindungsinformationen eine Priorisierung von Daten vorgenommen, um die Netzwerklast der Kapazität optimal anzupassen.

Das User Management genannte Modul repräsentiert die eingeloggten Benutzer in einer räumlichen Datenstruktur und schafft die Verbindung von Client im Netzwerk und dessen räumlicher Repräsentation in der Simulation.

Die Datenbank wahrt die Persistenz des Spielstandes. Im Falle von Serverabstürzen muss die Wiederaufnahme des Spiels möglich sein, ohne sämtliche Spielstände zu verlieren.

Das Game State Management ist das zentrale Modul, hier werden Benutzer Befehle verarbeitet und an die Physik weitergegeben, sowie Spielstandsänderungen persistent gemacht und die Update-Nachricht an den Spieler initiiert. Die Physik Engine ist das Modul, in dem sämtliche physikalisch reagierenden Objekte der Welt verwaltet werden. Steuerungsbefehle werden hier als Kräfte auf die Objekte angewendet und das Zusammenspiel von Kräften, Reibung Kollisionen wird unter Einfluss der Schwerkraft simuliert. Sämtliche Veränderungen und Bewegungen in der Welt werden an das Game State Management zurückgegeben.

Im Messaging Modul kommen Veränderungen aus der Physik, also des Spielstandes, über

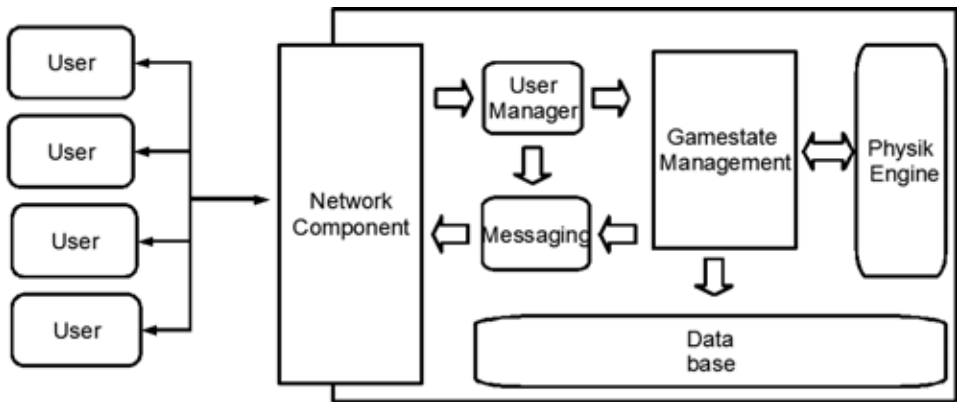


Abbildung 1: Entwurf eines Servers einer virtuellen Welt

das Gamestate Management an. Die „interessierten“ Adressaten werden aus dem User Management erfragt. Eine zu grobe Filterung der Ziel-Clients in diesem Modul führt schnell zu einer von der Netzwerkkomponente nicht zu bewältigenden Informationsflut.

Die auf Basis dieser Architektur angestellten Optimierungsansätze werden in Kap.4 vorgestellt .

4 Game State Management

Das Game State Management ist das zentrale Modul des Servers und hat Zugriff auf die meisten Komponenten, so dass hier eine Vielzahl von Möglichkeiten entsteht die Effizienz des Servers zu verbessern. Da eine Physik Engine ihre Objekte intern auch in einer dem Szenengraph ähnlichen Struktur aufbaut, kann das Game State Management deren Aufwand stark vereinfachen. Denn durch die Reduzierung der Objekte auf ihre Bounding Volumes, werden alle relativen Bewegungen von Objektknoten zu ihren Vaterknoten serverseitig ignoriert. Des Weiteren verringert sich auch die Anzahl der von der Physik verwalteten Objekte und damit der Aufwand für die Physiksimulation. Da solche Bewegungen für Menschen kaum Parallelen zur realen Welt erkennen lassen, werden die verloren gegangenen Informationen durch Animationsskripte auf Client-Seite ausgeglichen.

Für ein realistisches Spielerlebnis sind nicht alle Bereiche der virtuellen Welt gleich wichtig. Durch ein gewichtetes Physikupdate können Objekte in Gebieten von großem „Interesse“ bevorzugt werden. Entscheidend ist dabei, dass das Game State Management sicher stellt, dass Spieler mit wenig Aktivität nicht zu stark vernachlässigt werden und ihr Spielstand konsistent gehalten wird.

Literatur

- [A⁺03] Thor Alexander et al. *Massively Multiplayer Game Development*. Charles River Media, 2003.
- [AS09] Manuel Fradinho Oliveira Anthony Steed. *Networked Graphics - Building Networked Games and Virtual Environments*. Morgan Kaufmann, 2009.
- [AW⁺05] Thor Alexander, Gordon Walton et al. *Massively Multiplayer Game Development 2*. Charles River Media, 2005.
- [SL10] Linden Lbs Second Life. Economic Statistics (Raw Data Files), January 2010.

GPU-basiertes Volumen-Rendering mit Hilfe von Octree gestütztem Raycasting

Mathias Meissner

Martin-Luther-Universität Halle-Wittenberg
Naturwissenschaftliche Fakultät III – Institut für Informatik
mathias_me@gmx.de

Art der Arbeit: Diplomarbeit

Betreuer/in der Arbeit: Dr. Peter Schenzel

1 Motivation und Zielsetzung

Motivation Für die meisten 3D-Echtzeitdarstellungen von Volumeneffekten wie Nebel, Rauch, Feuer etc. werden heutzutage Billboards benutzt. Diese erzeugen allerdings häufig Artefakte beim Schnitt mit der Szenerie und tragen merklich dazu bei, dass ein gerendertes Bild nicht realitätsnah aussieht (Abb. 1). Eine typische Methode diese Artefakte zu eliminieren, sind die “Spherical Billboards” [Mei07a]. Mit ihnen lassen sich jedoch keine “eingeschlossenen” Volumen darstellen (wie zum Beispiel Eis oder Wachs). Eine zweite Möglichkeit ist es, den Z-Buffer für das Volumen-Sampling in konvexen Polyedern zu benutzen. Daraus ergab sich die Frage, wie man das Rendering sinnvoll vereinheitlichen kann. Eine Antwort darauf ist GPU-basiertes Raycasting. Raycasting bietet den Vorzug, dass es die Restriktion auf konvexe Volumen aufhebt und damit die Grundidee Volumen-Rendering in Polygon begrenzten Volumen universell einsetzbar macht. Außerdem werden mit Raycasting neue Effekte möglich, zum Beispiel diverse Beleuchtungseffekte, die sich konventionell nur schwer umsetzen ließen.

Aktuelle GPUs (Graphic Processing Unit) haben inzwischen genügend Rechenleistung, so dass die Implementierung eines Echtzeit-Raycasters durchaus erfolgversprechend erscheint. Vor allem weil für dieses Anwendungsgebiet im Falle zu geringer Grafikleistung, problemlos die Genauigkeit reduziert werden kann, ohne dass das Bild sofort entfremdet wird. Es sei noch erwähnt, dass es im Moment noch keinen Polygon-basierten Raycaster gibt, der auf einem gewöhnlichen Heim-PC in Echtzeit läuft, und dass absichtlich auf Hardwarehersteller gebundene Sprachen wie CUDA von Nvidia verzichtet wurde, um eine generelle Lauffähigkeit auf allen DirectX 10 Systemen garantieren zu können.

Zielsetzung Im Rahmen der Diplomarbeit wurde GPU-Ray, ein artefaktfreier, universell einsetzbarer, GPU-beschleunigter Volumen-Renderer auf der Basis eines Raycasters zur realistischen Echtzeit-Darstellung aller Arten transparenter Volumenobjekte wie zum Beispiel Nebel, Wolken, Rauch, Feuer, Eis, Wachs aber auch den “Röntgenblick” in einen 3-dimensionalen CT-Scan, entwickelt. Dazu wurden das GPU-basierte Raycasting / Raytracing sowie das für das Volumen-Rendering nötige Volumen-Sampling genauer betrachtet.

Um das Volumen-Rendering in adäquater Umgebung testen zu können, wird eine Szene generiert und für selbige eine Rendering-Engine erstellt, die Phong-Shading, Normal-Mapping und (Uniform-) Shadow-Mapping beherrscht. Zur Steigerung der Ausgabequalität des Volumen-Renderers wird als Postprocessing-Effekt ein Gaußscher Weichzeichner eingesetzt.

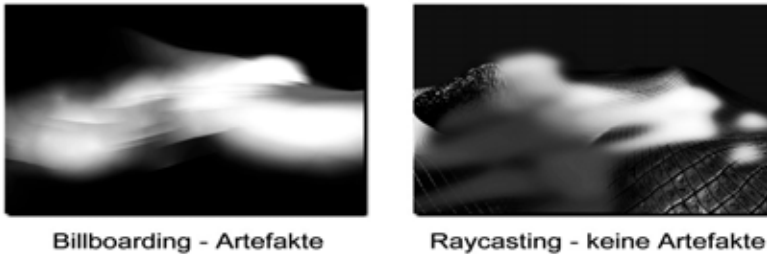


Abbildung 1: Billboarding vs. Raycasting

2 Echtzeit Raycasting und Volumen-Rendering

Raycasting Das Raycasting läuft in zwei Schritten ab. Im ersten Schritt werden, mit Hilfe einer Octree-Datenstruktur, die Dreiecke bestimmt, die für eine Schnittpunktberechnung in Betracht kommen. Im zweiten Schritt wird für jedes der im ersten Schritt bestimmten Dreiecke getestet, ob sie sich mit dem Strahl des Raycastings schneiden und gegebenenfalls wird dann der Schnittpunkt berechnet. Für die Schnittpunktberechnung kommt das "Fast, Minimum Storage Ray/Triangle Intersection" Verfahren von T. Möller und B. Trumbore [MT03] zum Einsatz. Die beiden Schritte müssen für jeden Strahl des Raycastings jeweils einmal ausgeführt werden, das bedeutet für jeden Punkt der Bildebene. Dabei hilft eine Raycasting-Maske, die mit dem Scanline-Renderer im Preprocessing erstellt wird, die Anzahl der betreffenden Bildpunkte zu reduzieren. Damit wird das Raycasting nur dort durchgeführt, wo das Volumen sichtbar ist und nicht verdeckt wird.

Beschleunigung Das Raycasting-Verfahren prüft bei der Bilderzeugung für jeden Strahl, welche Dreiecke er schneidet. Daraus ergibt sich ein Berechnungsaufwand von "Auflösung des Bildes" • "Anzahl der Dreiecke". Daher ist es zwingend notwendig, eine geeignete Datenstruktur zur Bestimmung der Dreiecke für den Schnitttest zu benutzen und so die Anzahl der Berechnungen und Speicherzugriffe erheblich zu reduzieren. Typische Datenstrukturen für solche Anwendungen sind BSP-Bäume und Octrees. Da von GPU-Ray alle Berechnungen, die für eine Datenstruktur notwendig sind, in den Pixels-Shadern durchgeführt werden müssen, entstehen einige Restriktionen, die nur den Octree als Datenstruktur zulassen. Um die Parallelisierbarkeit auf der GPU möglichst einfach zu halten, können und sollten Pixel-Shader-Programme nicht untereinander kommunizieren. So ist gewährleistet, dass jeder Shader-Prozessor sein Kernel/Programm ohne Seiteneffekte ausführen kann. Damit sich die Shader-Prozessoren auch beim Schreiben des Ergebnisses nicht blockieren, existiert hier die Vorgabe, dass die Speicherposition schon zu Beginn eines Shader-Programms feststeht. Daraus ergibt sich die Forderung ein statisches Sortiernetz zur Octree-Generierung zu nutzen. Hier bietet

sich Bitonic Sort [Bat68] an. Dies besitzt zwar nur eine asymptotische Laufzeit von $\mathcal{O}(n \log(n)^2)$, ist aber trotzdem das schnellste Sortiervorgehen, welches sich auf der GPU parallelisieren lässt.

Volumen-Rendering Für das Volumen-Sampling mit Lichtabsorption (Abb. 3) wird bei jedem Sampling-Schritt beachtet, wie viel Licht von dieser Position noch zum Augpunkt vordringen kann. Dies entspricht der folgenden Reihe (S = Ergebnis des Samplings, x, t = Farbwert, Transparenzwert am Sampling-Punkt, n = Anzahl der Sampling-Schritte):

$$S = \sum_{i=0}^n x_i \cdot t_i \cdot \prod_{j=0}^{i-1} 1 - t_j$$

Zur Schattierung kommt Volume-Shadow-Mapping zum Einsatz. Notwendig wird dies, da heutzutage die Rechenleistung von GPUs im Performance-Sektor nicht ausreicht, um die Schatten auch durch den Raycaster zu berechnen. Dazu wird für jede Lichtquelle eine uniforme Shadow-Map erzeugt und in das Volumen projiziert. GPU-Ray unterstützt dank des Raycasters auch Subsurface-Rendering und Refraktion an der Volumenoberfläche.

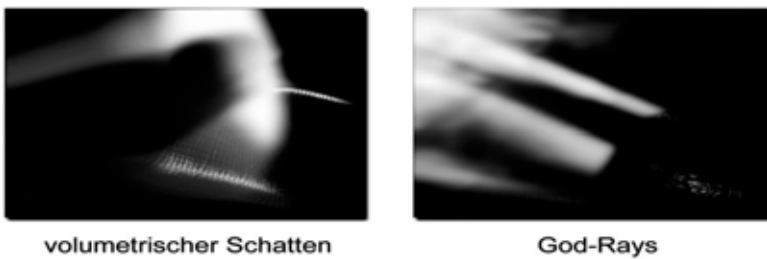


Abbildung 2: Verschiedene Volumen-Rendering-Effekte

3 Ergebnisse

Als Ergebnis dieser Arbeit unterstützt die GPU-Ray-Engine Volumen-Rendering, Echtzeit-Volumen-Schatten, farbigen Schattenwurf durch das Volumen auf die Szenerie, X-Ray-Rendering, Subsurface-Rendering und Echtzeit-Raycasting (in niedriger Auflösung). GPU-Ray ist dank seiner Architektur gut erweiterbar und kann, sobald die Rechenleistung der GPUs es zulässt, weitere Effekte nutzbar machen.

Auswertung Im Rahmen der Arbeit wurde GPU-Ray umfangreich auf verschiedenen Systemen und Grafikkarten getestet, um die Abhängigkeit der Rendering-Leistung von GPU-Ray und der Anzahl der Shader-Prozessoren, deren Takt sowie der Datentransferrate des Grafikspeichers zu untersuchen. Die Benchmark-Ergebnisse zeigen, dass GPU-Ray mit der Anzahl und dem Takt der Shader-Prozessoren hervorragend, meist linear, skaliert. Da das Raycasting mit vielen Speicherzugriffen verbunden ist, ist hier eine höhere Datentransferrate und Speicherbandbreite von Vorteil.

Eine Nvidia Geforce GTX 285 erreichte 60 fps bei einer Auflösung des Raycaster von $96 \cdot 96$ und einer Größe der Octree-Textur von $128 \cdot 128$, einer Octree-Tiefe von 4 sowie der Nutzung von 2 Octrees und 256 Volumen-Sampling-Schritten.

4 Ausblick

Einsatz in der Medizin Durch die einfache Steigerung der Präzision lässt sich das Volumen-Rendering aus GPU-Ray auch zur medizinischen Visualisierung nutzen (Abb. 3), um in Echtzeit einen 3-dimensionalen Blick in einen CT- oder MRT-Scan zu erhalten. Diese Präzisionssteigerung (Auflösung) auf zum Beispiel $512 \cdot 512$ Pixel schafft GPU-Ray in einem Multi-GPU-Verbund heutiger High End Grafikkarten mit Alternate-Frame-Rendering.

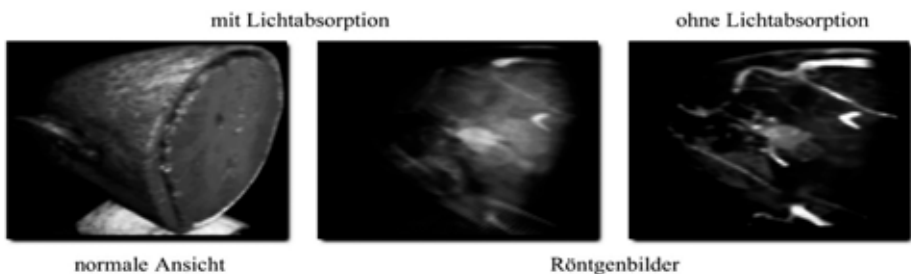


Abbildung 3: Schädel-CT mit verschiedenen Sampling-Verfahren

Weiterentwicklung Die DirectX 11 Compute-Shader sowie Hardware-Architekturen (Cache, etc.) die besser für Raytracing geeignet sind, werden in den nächsten Jahren für einen erheblichen Bedeutungsgewinn von Raytracing in Echtzeitanwendungen sowie Volumen-Rendering sorgen. In diesem Zusammenhang wird es nötig werden, die dazu verwendeten Techniken zu verbessern.

Literaturverzeichnis

- [MT03] T. Möller und B. Trumbore: Fast, Minimum Storage Ray/Triangle Intersection, 2003, <http://www.cs.virginia.edu/~gfx/Courses/2003/ImageSynthesis/papers/Acceleration/Fast%20MinimumStorage%20RayTriangle%20Intersection.pdf>
- [Bat68] K. E. Batcher: Sorting networks and their applications, 1968, <http://www.cs.kent.edu/~parallel/papers/sort.pdf>
- [ERC02] C. Everitt, A. Rege, and C. Cebenoyan: Hardware shadow mapping, 2002, <http://developer.nvidia.com/attach/8456>
- [Mei07] M. Meissner: Vortrag "Spherical Billboards" und Projektarbeit "Volumetric Rendering", 2007, Universität Halle
- [Mic09] Microsoft: Microsoft Developer Network - DirectX 10 und HLSL, 2009, [http://msdn.microsoft.com/en-us/library/ee663301\(VS.85\).aspx](http://msdn.microsoft.com/en-us/library/ee663301(VS.85).aspx)

Alle Abbildungen wurden mit GPU-Ray erstellt.

Konzeption und Umsetzung eines 3D-Bildbetrachters

Frank Nagl, frank.nagl@fh-erfurt.de
Jonas Röse, j10@j10.eu

Fachhochschule Erfurt
Gebäudetechnik und Informatik

www.xanko.de

Abstract: Diese Arbeit befasst sich mit der Konzeption und Umsetzung eines dreidimensionalen Bildbetrachters. Eine Recherche hat ergeben, dass viele Bildbetrachter Nachteile aufweisen. Aufbauend auf dieser Recherche wurde ein dreidimensionaler Bildbetrachter konzipiert und entwickelt, der die Vorteile anderer Betrachter vereinen soll.

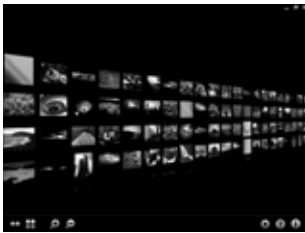


Abbildung 1: Vorschaumodus



Abbildung 2: Vorschaumodus
mit gekrümmter Bilderwand



Abbildung 3: Einzelbildmodus

1 Motivation

Die digitale Technik wird verwendet, um Bilder zu erstellen, aufzubewahren und zu betrachten. Ein wichtiges Programm, welches dabei zum Einsatz kommt, ist der verwendete Bildbetrachter, der dem Benutzer auf einfache, unkomplizierte und angenehme Art, die Bilder präsentieren soll. Es gibt bereits ein großes Angebot an solchen Programmen, aber viele haben kleine bis große Nachteile. Bei großen Bildmengen geht schnell der Überblick verloren. Außerdem fehlen bei einigen Programmen Funktionen, wie das Beschriften oder das verlustfreie Drehen von Bildern. Viele Funktionen können nicht auf mehrere Bilder gleichzeitig angewendet werden. Manche Programme haben Einschränkungen in der Bedienung, so dass es beispielsweise nicht möglich ist, während der automatischen Bildwiedergabe Bilder zu drehen oder zu zoomen. Andere Programme sind wiederum zu umfangreich und dadurch nicht so einfach bedienbar für den Benutzer.

2 Ziel

Das Ziel dieser Arbeit ist die Konzeption und Umsetzung eines übersichtlichen und benutzerfreundlichen Bildbetrachters. Dieses Programm soll dem Anwender die wichtigsten Funktionen zum Betrachten, Präsentieren, Auswählen und Bearbeiten von lokal gespeicherten Bildern zur Verfügung stellen. Zu diesen Funktionen zählen das Beschriften, Zoomen, Drehen und Aussortieren von Bildern. Außerdem soll durch eine dreidimensionale Darstellung der Bilder die Übersicht bei großen Bildmengen nicht verloren gehen. Zur Präsentation der Bilder dient eine automatische Bildwiedergabe.

3 Stand der Technik

Mit der Entwicklung digitaler Bilder entstanden auch eine Vielzahl von Bildbetrachtern mit unterschiedlichen Funktionalitäten. Aufgrund der immer größer werdenden Datenflut von Bildern ergeben sich neue Anforderungen an solche Programme, wie in Kapitel 2 beschrieben.

Existierende Bildbetrachter, wie Windows Fotogalerie [M06], ACDSee Foto-Manager [A09], IrfanView [S09], Cooliris [C09] erfüllen nicht all diese Anforderungen. Einige sind zu umfangreich und dadurch für den unerfahrenen Benutzer ungeeignet. Andere sind zwar benutzerfreundlich, aber mit einigen Einschränkungen in der Funktionalität. Außer Cooliris hat auch keiner dieser Bildbetrachter eine dreidimensionale Galerie zur guten Übersicht bei großen Bildmengen.

4 Konzept

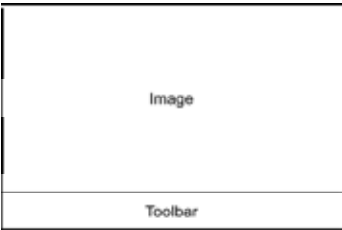


Abbildung 4: Einzelbildmodus

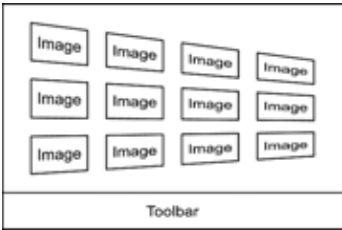


Abbildung 5: Vorschaumodus

Der Bildbetrachter ist grundsätzlich in zwei Modi aufgeteilt, das sind zum einen der Vorschaumodus und zum anderen der Einzelbildmodus. Die Bilder werden in mehreren Qualitätsstufen geladen, um die Ladezeiten zu verkürzen. Außerdem steht dem Benutzer ein Einstellungsfenster zur Verfügung, mit dem das Verhalten des Programms an einigen Stellen beeinflusst werden kann. In einem Hilfefenster werden alle Funktionen mit Tastatur- und Mausbelegung aufgelistet.

4.1 Einzelbildmodus

Der Einzelbildmodus dient zur direkten Betrachtung eines bestimmten Bildes. Hier stehen alle Funktionen zur Verfügung, die auf ein einzelnes Bild anwendbar sind. Zu diesen gehören das Beschriften, Zoomen und Drehen von Bildern. Der Benutzer kann zum vorherigen und zum nachfolgenden Bild wechseln. Eine automatische Bildwiedergabe ist von jedem Bild aus startbar. Die Funktionen sind zum einen durch Buttons in einer Toolbar und zum anderen durch Tastenkürzel auslösbar.

Abbildung 4 zeigt, wie die Benutzeroberfläche im Einzelbildmodus aufgeteilt wird. Den größten Teil nimmt dabei das Bild selbst ein. Falls ein Bild beschriftet wurde erscheint der Text direkt unter dem Bild. Im unteren Bereich der Benutzeroberfläche befindet sich die Toolbar mit Buttons. Der Benutzer kann mit einem Button der Toolbar oder mit einem Klick auf das Bild zum Vorschaumodus wechseln.

4.2 Vorschaumodus

Der Vorschaumodus dient dazu, um einen schnellen Überblick aller Bilder eines Ordners zu bekommen. Alle Vorschaubilder werden im Hintergrund parallel zu den Benutzerinteraktionen geladen werden. Der Benutzer kann das Programm währenddessen uneingeschränkt bedienen. Gerade bei großen Bildmengen muss der Benutzer dadurch nicht auf das Laden vieler Bilder warten. Es werden alle Bilder des Ordners geladen, der das ausgewählte Bild enthält.

Abbildung 5 zeigt wie der Vorschaumodus aufgebaut ist. Der größte Teil der Benutzeroberfläche wird durch eine dreidimensionale Galerie eingenommen. Im unteren Teil befindet sich, wie auch im Einzelbildmodus, eine Toolbar. Diese enthält die wichtigsten Funktionen, die für den Vorschaumodus relevant sind, wie das Zoomen oder das Einstellen der Bildzeilen. Außerdem können Bilder aussortiert werden. Mit einem Klick auf ein beliebiges Bild kann zum Einzelbildmodus gewechselt werden.

5 Umsetzung

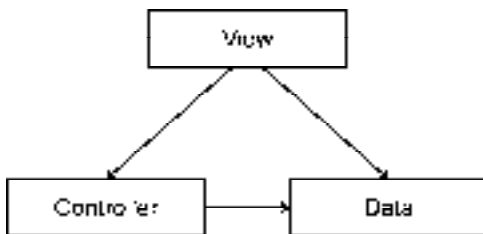


Abbildung 6: Vereinfachtes Klassendiagramm

Der Bildbetrachter wird in der Programmiersprache C# [M09a] geschrieben. Als Entwicklungsumgebung dient Microsoft Visual Studio 2008 [M08]. Das Programm soll eine dreidimensionale Oberfläche bekommen, deswegen dient als Programmierschnittstelle zur Grafikkarte Microsoft DirectX [M09b].

Abbildung 6 zeigt den vereinfachten Aufbau des Programms. Der Bildbetrachter besteht grundsätzlich aus einer View-, einer Controller- und einer Data-Class. Die View-Class dient zur grafischen Ausgabe der Benutzeroberfläche und nimmt die Benutzerinteraktionen entgegen. In der Controller-Class werden durch Methodenaufrufe bestimmte Funktionen ausgeführt, wie beispielsweise das Zoomen oder Drehen von Bildern. Die Data-Class enthält Daten und Zustände des Bildbetrachters.

6 Ergebnis

Abbildung 1 bis Abbildung 3 zeigen den finalen Bildbetrachter. Dieser bietet nun durch die dreidimensionale Bildergalerie einen guten Überblick bei großen Bildmengen. Außerdem können Bilder einfach beschriftet, gedreht und aussortiert werden. Durch eine einfach aufgebaute Toolbar ist der Bildbetrachter auch benutzerfreundlich und für unerfahrene leicht zu bedienen.

Der Bildbetrachter heißt Diashow. Auf der Webseite www.xanko.de befinden sich eine detaillierte Liste des Funktionsumfangs, weitere Abbildungen und die ausführbaren Programmdateien zum herunterladen.

7 Ausblick

Zukünftig kann der Bildbetrachter weiterentwickelt werden, um den Benutzer Funktionen, wie Komprimierung und Beschneidung von Bildern, Farbanpassungen, Entfernen von roten Augen, Mehrfachumbenennung, Abspielen von Videos oder Anzeigen von Metainformationen zu ermöglichen.

Literaturverzeichnis

- [A09] ACD Systems International: ACDSee Foto-Software, <http://de.acdsee.com>, Zugriffsdatum: 01.01.2010
- [C09] Cooliris, <http://www.cooliris.com>, Zugriffsdatum: 01.01.2010
- [M06] Microsoft: Windows Vista, <http://www.microsoft.com/germany/windows/windows-vista/>, Zugriffsdatum: 01.01.2010
- [M08] Microsoft: Visual Studio 2008, <http://www.microsoft.com/germany/visualstudio/>, Zugriffsdatum: 01.01.2010
- [M09a] Microsoft: C#, <http://www.microsoft.com/vcsharp/>, Zugriffsdatum: 01.01.2010
- [M09b] Microsoft: DirectX, <http://www.microsoft.com/directx/>, Zugriffsdatum: 01.01.2010
- [S09] Irfan Skiljan: IrfanView - Freeware Graphic Viewer, <http://www.irfanview.de>, Zugriffsdatum: 01.01.2010

Negolook: Eine Integration clientseitiger Verhandlungsunterstützung in betriebliche Kommunikationssysteme

Robert Elsler, Malte Horstmann, Michael Körner

Universität Hohenheim
Lehrstuhl für Wirtschaftsinformatik I
robert.elsler@wil.uni-hohenheim.de
malte.horstmann@wil.uni-hohenheim.de
michael.koerner@wil.uni-hohenheim.de

Art der Arbeit: Bachelor Thesis

Betreuer/in der Arbeit: Prof. Dr. Mareike Schoop

Abstract: Elektronische Transaktionen im Kontext betrieblicher Strukturen gewinnen stetig an Bedeutung. Im Zeitalter des E-Business stellt sich die Frage, wie solche Transaktionen elektronisch unterstützt werden können. Negolook ist ein erster Ansatz, Unterstützung bei der Verhandlungsführung mittels eines in bestehende Geschäftsprozesse integrierten Systems zu bieten. Durch die Integration von verhandlungsunterstützender Funktionalität in Microsoft Outlook, werden Zugangsbarrieren, die bisherige Ansätze aufweisen, entschärft. Somit soll die Diskrepanz zwischen der von Praktikern geäußerten Bedarf dieser Systeme und deren tatsächlichem Einsatz in einem praktischen Umfeld vermindert werden.

1 Motivation und Ziel der Arbeit

Die elektronische Abwicklung zwischenbetrieblicher Transaktionen gewinnt im Zeitalter des E-Business stetig an Bedeutung. Diese Entwicklung ist begleitet von einem häufigeren und intensiveren Einsatz von Informationssystemen, mit denen die Abwicklung solcher Transaktionen unterstützt werden kann [La07]. Wie Lau [La07] ausführt, besteht speziell im Bereich elektronischer Verhandlungsabwicklung großer Bedarf nach derartiger Unterstützung. Verschiedene Forschergruppen beschäftigen sich seit Jahren mit der Entwicklung und der experimentellen Evaluation von Verhandlungsunterstützungssystemen (vgl. hierzu entsprechende Publikationen z.B. in [GDN92]).

Allerdings werden diese Systeme in der Praxis kaum verwendet, obwohl sie – basierend auf empirischen Forschungsergebnissen (bspw. [DL08]) – von Praktikern befürwortet werden. Stattdessen verhandeln 55% der Betriebe, die elektronisch verhandeln – so die Ergebnisse einer Umfrage von Herbst et al. [HKS08] - lediglich per E-Mail.

Es befindet sich demnach eine Diskrepanz zwischen dem geäußerten Bedarf an Verhandlungsunterstützungssystemen und deren tatsächlichem Einsatz. Grundlegendes Ziel dieser Arbeit ist es, einen Ansatz zu finden, mit dem diese Lücke geschlossen werden kann.

Durch eine Integration von verhandlungsunterstützender Funktionalität in bereits vorhandene Anwendungssysteme soll eine höhere Akzeptanz erzeugt werden, die in einer tatsächlichen praktischen Verwendung von Verhandlungsunterstützungssystemen resultiert. Prototypisch wurde hierzu ein Add-In für Microsoft Outlook 2007 entwickelt, das die Funktionalitäten eines Verhandlungsunterstützungssystems bereitstellt – der bestehende Kommunikationskanal per E-Mail wird also auf diese Weise angereichert.

2 Erläuterung des Ansatzes

Die bisherigen Ansätze der Entwicklung von Verhandlungsunterstützungssystemen, wie Inspire, CBSS oder Negoisst, basieren auf einer unabhängigen Lösung eines serverseitigen Systems. Der im Folgenden vorgestellte Ansatz des Negolook-Prototyp (*Negotiation & Outlook*) basiert auf einer integrierten clientseitigen Lösung, dessen Funktionen sich im Wesentlichen an Negoisst orientieren. Als Testsystem wurde Outlook 2007 aufgrund seiner hohen Verbreitung [Fi09] gewählt.

Ein Negotiation Support System (NSS) basiert im Wesentlichen auf drei Komponenten: der Entscheidungsunterstützung, der Kommunikationsunterstützung und dem Dokumentenmanagement.

Mittels der entscheidungsunterstützenden Komponente wird der Nutzen der Angebote in den Verhandlungsnachrichten berechnet. Als gängiges Verfahren zur Berechnung des Nutzens aus einer Präferenzstruktur hat sich die multiattributive Nutzentheorie [KR76] erwiesen. Zur Berechnung ist die Gewichtung der einzelnen Elemente der Agenda und damit die Generierung eines verhandlungsspezifischen Präferenzmodells nötig. Diese Gewichtung wird durch den Anwender direkt in Outlook vorgenommen. Die daraus erzeugte Bewertung der jeweiligen Agenda ist für den Anwender bereits beim Erstellen und Empfangen einer neuen Nachricht sichtbar.

Die kommunikationsunterstützende Komponente umfasst die Unterscheidung von Nachrichtentypen und einen protokollbasierten Handlungsablauf. Basierend auf der Sprechakttheorie von Searle [S69] und der Theorie des kommunikativen Handelns von Habermas [H81] entstand die Language/Action Perspective (LAP). Diese bildet die Grundlage für die von Jertila et al. [JLS03] abgeleiteten sieben Nachrichtentypen. Sie bieten die Möglichkeit einer Nachricht eine Intention und dem Verhandlungspartner somit eine eindeutige Handlungsaufforderung mitzusenden. Jede Nachricht, die mit Hilfe des Negolooks-Add-In versendet wird, muss mit einem dieser Nachrichtentypen versehen werden. Aus der LAP und dem damit verbundenen Ansatz des strukturierten Nachrichtenaustausches geht desweiteren hervor, dass auch die Nachrichtenabfolge strukturiert werden kann. Quix und Schoop [QS01] vertreten die Idee eines strikt alternierenden Verhandlungsprotokolls. So wissen Verhandlungsteilnehmer zu jeder Zeit der Verhandlung, ob sie verpflichtet sind, selbst zu agieren oder auf eine Nachricht ihres Partners warten müssen. Dieser Ansatz wurde im Prototyp in Form eines bilateralen Protokolls implementiert.

Verhandlungen werden mitunter auch als das iterative Austauschen von spezifizierten Agenden beschrieben. Damit jeweils die richtige Version des Vertrages an die passende Nachricht gekoppelt ist, weisen Quix und Schoop [QS01] auf ein Dokumentenmanagement hin. Dieses bietet dem Nutzer einen chronologischen Überblick seiner Verhandlung. Um das bindende Dokument, eine mit Werten versehene Agenda, einer konkreten Nachricht zuzuweisen, wurde ein Framework entwickelt, das an das von Quix und Schoop [QS01] entwickelte DOC.COM Framework angelehnt ist.

3 Ausblick

Ziel der Arbeit war es, Potenziale und Grenzen eines Verhandlungsunterstützungssystems aufzuzeigen, das in ein dem Benutzer bekanntes Medium integriert ist. Der aktuelle Stand der Arbeit stellt eine theoretische Evaluation des Ansatzes anhand der Unified Theory of Acceptance and Use of Technology (UTAUT) [MV03] dar. Die Evaluation beschränkt sich innerhalb des UTAUT auf die ersten beiden Kriterien, Performance Expectancy und Effort Expectancy.

Im Hinblick auf die Performance Expectancy kann der Prototyp Negolook dem Benutzer einen deutlichen Mehrwert im Vergleich zur konventionellen E-Mail Kommunikation bieten. Die verhandlungsunterstützenden Elemente helfen dem Benutzer dabei einen stringenten und leicht nachvollziehbaren Verhandlungsverlauf zu generieren. Dadurch wird der Benutzer in die Lage versetzt, durch einen einfachen Zugang auf die für ihn entscheidungsrelevanten Informationen, effizient und auf einer qualitativ hochwertigen Entscheidungsgrundlage seine Verhandlung zu führen. Im Idealfall führt dies dazu, dass weniger Nachverhandlungen notwendig sind, und somit die Transaktionskosten sinken.

Desweiteren wird der Benutzer durch die Integration in Outlook 2007 nicht dazu gezwungen, sich mit einem neuen System vertraut zu machen. Durch die Integration ist der Schulungsaufwand für die Mitarbeiter geringer, da auf bereits vorhandenem Wissen aufgebaut wird.

Grenzen sind in der grundlegenden Client-Client-Architektur des Prototypen zu sehen. Die Folge dieser Architekturwahl ist, dass es innerhalb einer Verhandlung keine Trusted Third Party gibt [JLS03]. Dennoch kann eine nicht manipulierte Verhandlung gewährleistet werden, da sobald eine Manipulation an verhandlungsspezifischen Dateien durchgeführt werden sollte, der Prototyp nicht mehr in der Lage ist die Verhandlung fortzusetzen. Die Architekturwahl bietet zudem den Vorteil, dass zur Nutzung des Negolook Add-In die vorhandene Infrastruktur weiterverwendet werden kann.

Die bisher genannten Faktoren sind Gründe, weshalb das Negolook Add-In für Unternehmen von hohem Interesse ist. Die Kommunikation ist genauso sicher, wie der reguläre E-Mail-Versand. Eine empirische Untersuchung in Form von Verhandlungsexperimenten stellt nun die konsequente Fortsetzung dieser Arbeit dar, um somit einen direkten Vergleich zwischen einer reinen E-Mail Verhandlung und einer Verhandlung mit Negolook zu ermöglichen.

Literaturverzeichnis

- [DL08] Doong, H.-S., Lai, H.: Exploring usage continuance of e-negotiation systems: expectation and disconfirmation approach. In (Shakun, M.F., Hrsg.): Group Decision and Negotiation, Springer Netherlands, Dordrecht, 17 (2), 2008. S. 111-126.
- [Fi09] <http://www.finanznachrichten.de/nachrichten-2009-05/13949074-mailprogramm-umfrage-2009-outlook-trotz-rueckgaengen-weiterhin-marktfuehrer-mit-27-thunderbird-mit-14-auf-platz-zwei-microsoft-google-appl-001.htm>, letzter Aufruf: 9.1.2010.
- [GDN92] Group Decision and Negotiation Journal, Shakun, M.F. (Hrsg.), Springer Netherlands, Dordrecht, 1992-heute
- [H81] Habermas, J.: Theorie des kommunikativen Handelns (Bd. 1). Suhrkamp, Frankfurt am Main, 1981.
- [HKS08] Herbst, U., Köhne, F., Schoop, M., Staskiewicz, D., Voeth, M.: The antecedants of renegotiations in practice – An exploratory analysis. In (Shakun, M.F., Hrsg.): Group Decision and Negotiation, Springer Netherlands, Dordrecht, 17(2), 2007. S. 127-139
- [JLS03] Jertila, A., List, T., Schoop, M.: Negoisst: a negotiation support system for electronic business-to-business negotiations in e-commerce. In (Chen, P.P., Hrsg.): Data & Knowledge Engineering, Elsevier, Amsterdam, 47 (3), S. 371-401.
- [KR76] Keeney, R. L., Raiffa, H.: Decisions with multiple objectives: preferences and value trade-offs. J. Wiley & Sons, New York, 1976.
- [La07] Lau, R.Y.K.: Towards a web services and intelligent agents-based negotiation system for B2B eCommerce. In (Kauffman, R., Hrsg.): Electronic Commerce Research and Applications, Elsevier, Amsterdam, 6 (3), 2007. S.260-173.
- [Li03] Lim, J.: A conceptual framework on the adoption of negotiation support systems. In (Wohlin, C., Hrsg.): Information and Software Technology, Elsevier, Amsterdam, 45 (8), 2003. S. 469-477.
- [MV03] Morris, M. G., Davis, G. B., Davis, F. D., Venkatesh, V.: User Acceptance of Information Technology: Toward Unified View. In (Gupta, A., Hrsg.) MIS Quarterly, Minnesota, 27 (3), 2003. S. 425-478.
- [QS01] Quix, C., Schoop, M.: DOC.COM: a framework for effective negotiation support in electronic marketplaces. In (Akyildiz, I.F., Hrsg.): Computer Networks, Elsevier, Amsterdam, 37 (2), 2001. S. 153-170.
- [S69] Searle, J. R.: Speech Acts: An Essay in the Philosophy of Language. Cambridge University Press, Cambridge, 1969.

Serviceorientiertes-Modell eines TOGAF basierten Portals für Enterprise Architecture Management (EAM)

Matthias Gutbrod, Felix Schiele, Christian Zinsenhofer

Hochschule Reutlingen / Reutlingen University
Wirtschaftsinformatik Master

Matthias.Gutbrod@Student.Reutlingen-University.de
Felix.Schiele@Student.Reutlingen-University.de
Christian.Zinsenhofer@Student.Reutlingen-University.de

Betreuer: Prof. Dr. rer. nat. Alfred Zimmermann

Abstract: Das Konzept behandelt eine Metamodell gestützte Integration der Fachlichkeiten eines Enterprise Architecture Management Portals, welches mit Hilfe einer serviceorientierten Architektur umgesetzt wurde. Undurchsichtige, komplexe IT-Landschaften mit dezentraler Verwaltung stellen eine Herausforderung für die Unternehmen dar. Dieser kann durch die Einführung einer Enterprise Architecture begegnet werden. Bei immer kürzer werdenden Anpassungszyklen benötigen Unternehmen in ihrer Architektur einen hohen Grad an Agilität. Ein serviceorientiertes zentralisiertes EAM-Portal bietet die nötige Transparenz und Flexibilität. Einen solcher Ansatz bieten aktuelle Systeme wie PlanningIT, essential und iteratec noch nicht. In diesem Kontext stellt das beschriebene serviceorientierte EAM-Portal eine Innovation dar. Das EAM-Portal bietet dem Nutzer einen zentralen, personalisierbaren Einstiegspunkt in unternehmensweiten Architekturen.

1 Problemformulierung

Durch steigende Anforderungen sind in den Unternehmen komplexe IT-Landschaften mit dezentraler Verwaltung und mangelnder Transparenz entstanden. Redundante Systeme, Applikationen, Prozesse und steigender Verwaltungsaufwand führen zu hohen Betriebskosten für die IT. Diese gewachsene IT-Landschaft zeichnet sich durch starre Strukturen aus, was zu mangelnder Flexibilität führt. Laut M. Kütz [Kü09] wird die IT in den Unternehmen oft als reiner Kostenfaktor gesehen, den es zu minimieren gilt.

Änderungen der Business- und IT-Strategie erfolgen oft ohne ausreichende Planung und ohne Beachtung der Abhängigkeiten zwischen den Bereichen. Es existiert kein Wissen darüber, wie sich Änderungen auf der Fachseite und der IT auswirken. Wechselwirkungen und Konflikte können nicht frühzeitig identifiziert werden, was eine optimale Abstimmung und Problemlösung verhindert.

Für die zentrale Verwaltung der Unternehmensarchitektur bietet sich Enterprise Architecture Management an. Wichtig ist hierbei eine flexible Struktur, so dass die EAM Lösung schnell angepasst werden kann. Aktuell verfügbare EAM Lösungen von IDS Scheer, Alfabet, iteratec und EAS bieten diese Flexibilität mit Hilfe von serviceorientierten Integrationen noch nicht.

2 Einführung

IT-Abteilungen sind häufig auf Konzeption, Entwicklung und Einsatz von Systemen, deren Architekturen und Technologien fokussiert. Neuere Architekturkonzepte erweitern diese zu einer Perspektive auf das Gesamtunternehmen und müssen daher Geschäftsarchitekturen, mit zugehörigen IT-Strategien und Betriebsarchitekturen mit einschließen. In diesem Kontext wird von einer Unternehmensarchitektur bzw. von einer Enterprise Architecture [To09] gesprochen. Dieser erweiterte Blickwinkel hilft dabei Probleme zu lösen, welche mit einem eng gesetzten Fokus nicht sinnvoll zu lösen sind [Sc09]. Durch die ganzheitliche Sicht auf das Unternehmen kann die Lücke zwischen IT und Business bereits bei der Erstellung der Architektur verringert werden.

Die Enterprise Architecture setzt sich aus verschiedenen Teilarchitekturen zusammen. Im einfachsten Fall wird dabei die Geschäfts- und die IT-Architektur unterschieden. Die Geschäftsarchitektur umfasst dabei alle nicht-technischen Bausteine der Architektur. Die IT-Architektur umfasst hingegen alle technischen Bestandteile der Enterprise Architecture [Sc09]. Diese Zweiteilung ist in der Praxis nicht ausreichend, deshalb hat sich nach I. Hanschke [Ha09] eine Vierteilung der Unternehmensarchitektur als Best Practice erwiesen.

TOGAF [To09] stützt sich als wesentlicher international anerkannter Standard für EAM auf das TAFIM (Technical Architecture Framework for Information Management) des „Department of Defense“ (DoD) und wird seit 1995 ständig weiterentwickelt [Ha09]. TOGAF ist ein umfangreiches Enterprise Architecture Framework, das verschiedene Methoden und Werkzeuge zur Entwicklung einer Enterprise Architecture zur Verfügung stellt. Dazu wird ein iteratives Prozessmodell, welches auf Basis von Best Practices erstellt wurde verwendet. Im Kern von TOGAF befinden sich vier Teilarchitekturen. Die Business Architektur definiert die Business Strategie, Steuerung, Organisation, und die Key Business Prozesse. Über die Data Architecture werden für das Unternehmen wichtigen Datentypen und Datenquellen definiert. In der Application Architecture werden wesentliche Anwendungen definiert, welche für die Verarbeitung der Daten und die Unterstützung der Geschäftstätigkeit benötigt werden. Die Technology Architecture beschreibt die benötigten logischen Software- und Hardwareressourcen, die zur Unterstützung von Business, Daten und Applikationen benötigt werden. Wird eine Enterprise Architecture im Unternehmen eingeführt, muss sie fortlaufend gepflegt und weiterentwickelt werden. Dies geschieht unter Anwendung eines Enterprise Architecture Managements.

3 Das serviceorientierte EAM-Portal

Die Hauptidee des präsentierten Ansatzes ist das Konzept eines serviceorientierten EAM- Portals. Das EAM-Portal stellt den zentralen Einstiegspunkt für den Benutzer in das Enterprise Architecture Management dar. Ein Single Sign On ermöglicht den Zugriff auf verschiedene Applikationen. Das EAM-Portal bündelt durch eine konsistente, einheitliche und personalisierte Interaktionsmöglichkeit den ergonomischen Zugang für eine Vielfalt von angebotenen Funktionalitäten. Die Verwendung eines Rollenmanagements erlaubt die unternehmensweite Umsetzung von Sicherheitsaspekten.

Die vier Teilarchitekturen des TOGAF Frameworks umfassen die Business-, Daten-, Applikations- und Technologiearchitektur und bestimmen das Metamodell für die Modellierung der Enterprise Architecture. Ein Architecture Repository ermöglicht es alle wesentlichen durch das Metamodell definierten Artefakte konsistent abzulegen und diese Architekturbestände zu verwalten. Dadurch ermöglicht das EAM-Portal einen Gesamtüberblick über die ausbaubare Enterprise Architecture. Es bietet dem Benutzer eine zentralisierte Verwaltung aller Teilarchitekturen, wodurch die Transparenz im Unternehmen erhöht wird. Dies führt dazu, dass Redundanzen bei Geschäftsprozessen, Anwendungssystemen und den eingesetzten Technologien schneller identifiziert werden können. Darüber hinaus eröffnet die verwaltete Bestandsinformation der Enterprise Architecture die Simulation künftiger Ausbauszenarien der Informationssysteme bereits vor der Realisierung und zu einem frühen Zeitpunkt der Konzeption. Definierte und aufeinander ausgerichtete Teilarchitekturen der Enterprise Architecture ermöglichen ein konsequentes Business-IT-Alignment und damit konsistent und effizient abgestimmte Fachlichkeiten, Anwendungssysteme und IT-Infrastrukturen. Um die Unternehmensarchitekturen schnell an neue Anforderungen anpassen zu können, benötigt das EAM-Portal eine hohe Agilität. Eine serviceorientierte Realisierung des EAM-Portals bietet hierfür die notwendigen Eigenschaften zur effizienten Adaption der Funktionalitäten sowie zur flexiblen Integration der Bestandsdaten. Die Implementierung des EAM-Portals soll durch Web Services mit der Java Enterprise Edition (Java EE) realisiert werden.



Abbildung 1: Serviceorientierte Architektur des EAM-Portals

In Abbildung 1 werden die Schichten des serviceorientierten EAM-Portals dargestellt. In der Darstellungsschicht befinden sich die Interaction-Services, welche die Interaktion mit dem Benutzer ermöglichen.

Für die Benutzungsoberfläche wird das Java Server Faces (JSF) Framework verwendet, welches das Model View Controller (MVC) Pattern implementiert [St09]. Die Prozess-Services in der Prozessschicht übernehmen die Steuerung und Kontrolle der Prozesse des EAM-Portals. In diesen Prozessen werden die Task-Services der Logikschicht orchestriert. In der Logikschicht befinden sich die einzelnen Task-Services, welche die verschiedenen Grundfunktionen des EAM-Portals ermöglichen. Für den persistenten Datenzugriff werden spezifische Entity-Services eingesetzt. Dabei enthalten die Entity-Services neben den CRUD Funktionen auch entitätsspezifische Kernfunktionalitäten. Monitoring-Services unterstützen den Portalbenutzer durch Überwachungs- und Analysefunktionen von aktiven Services und liefern damit eine zusätzliche Transparenzfunktionalität über integrierte Systemlandschaften sowie konkret ausgeprägte Bearbeitungsfälle und Geschäftsprozesse. Utility-Services helfen in der Gesamtverarbeitung der Services durch allgemeine und mehrfach benötigte Basisfunktionen, beispielsweise für die Protokollierung von Ereignissen [Er08] oder die Behandlung von fachlichen und technologischen Ausnahmen. Der Enterprise Service Bus (ESB) ist für die Kommunikation zwischen unterschiedlich konzipierten Services zuständig. Er koordiniert den Ablauf der Kommunikation und ermöglicht diese auch bei heterogenen Services und Landschaften indem er eine Kapselung der Nachrichten vornimmt [St07].

4 Schlussfolgerung und Ausblick

Im Rahmen dieses Projekts wurde ein serviceorientierter Entwurf eines Metamodell-gestützten EAM-Portals erstellt. Der Einsatz eines serviceorientierten EAM-Portals bietet dem Unternehmen die notwendige Flexibilität um die Unternehmensarchitektur effizient neuen Gegebenheiten anzupassen. Durch die Serviceorientierung verfügt das Portal über eine optimierte Anpassbarkeit, so lassen sich Änderungen mit geringem Aufwand in das EAM-Portal integrieren. Das Ergebnis der abgeschlossenen Konzeptions- und Entwurfsphase ist das beschriebene serviceorientierte Modell des EAM-Portals. Das Modell soll in weiteren Schritten durch Prototypen validiert und in einem Pilotprojekt erprobt werden. Gemäß durchgeführter Evaluierungen von alternativen Technologien und Infrastrukturen kann eine Realisierung auf Open Source Basis unter Verwendung des GlassFish Portfolios – der Referenzimplementierung von SUN für Java Technologien - erfolgen.

Literaturverzeichnis

- [Er08] Erl T.: SOA, Addison-Wesley Verlag, München, 2008
- [Ha09] Hanschke, I.: Strategisches Management der IT Landschaft, Hanser 2009
- [Kü09] Kütz, M.: Kennzahlen in der IT. dpunkt Verlag, Heidelberg, 2009
- [Sc09] Schwarzer, B.: Einführung in das Enterprise Architecture Management – Verstehen-Planen-Umsetzen. Books on Demand GmbH, Norderstedt, 2009
- [St07] Starke, G.; Tilkov, S.: SOA-Expertenwissen. dpunkt Verlag, Heidelberg, 2007
- [St09] Stark, T.: Java EE 5. Pearson Studium, München, 2009
- [To09] The Open Group: TOGAF™ Version 9. Van Haren Publishing, 2009

Retooling in der Automobilindustrie: betriebswirtschaftliche Vorteile durch den Einsatz einer Software

Heike Johannes

FH Deggendorf

Fakultät Betriebswirtschaftslehre / Wirtschaftsinformatik
heike.johannes@csp-sw.de

Art der Arbeit: Bachelorarbeit
Betreuer der Arbeit: Prof. Dr. Herbert Fischer

Abstract: Die Wirtschaftskrise beutelte in 2009 vor allem die Automobilindustrie in Deutschland: hochwertige und dementsprechend hochpreisige Fahrzeuge wurden von den Verbrauchern weniger nachgefragt als in den Vorjahren. Deshalb lautet die Devise in dieser Branche: Sparen!

Der daraus resultierende Trend hin zur Wiederverwendung von Anlagen – im Allgemeinen als Retooling bekannt – führt derzeit zu neuen Arbeitsabläufen im Bereich der Montageplanung bei den Automobilherstellern. Die Arbeit soll klären, welche betriebswirtschaftlichen Vorteile der Einsatz einer Software in diesem Bereich bietet.

1 Notwendigkeit eines definierten Prozesses

Für ein geplantes und kosteneffizientes Retooling ist ein definierter Prozess erforderlich, um eine aussagekräftige Entscheidungsgrundlage für Wiederverwendung oder Neuinvestition zu erhalten.

Deshalb müssen bei einer Retooling-Maßnahme verschiedene Abteilungen zusammen arbeiten: zunächst steht die Aufnahme der Ist-Situation an, die meist Personen aus der Instandhaltung übernehmen. Die Anforderungen an Equipments für neue Baureihen werden in der Regel von Montageplanern erfasst. Aus diesen beiden Informationsquellen müssen Auswertungen erstellt werden, ob eine Wiederverwendung unter Kostenaspekten sinnvoll ist oder ob eine Neuinvestition durchgeführt werden soll.

Jedes Unternehmen muss also zunächst einen Arbeitsablauf festlegen, um transparente und nachvollziehbare Entscheidungen herbeiführen zu können.

2 Softwareeinsatz bei einem Retooling-Prozess

Neue Prozesse in Unternehmen bergen immer das Risiko, dass sie in der Anfangsphase instabil werden. Um dieser Gefahr entgegenzuwirken, sollte eine Applikation den neuen Arbeitsablauf vereinfachen und ihn damit langfristig im Unternehmen etablieren.

2.1 Anforderungen an eine Software

Unabhängig vom Einsatzzweck der Software gelten bestimmte Kriterien, die eine Applikation erfüllen soll.

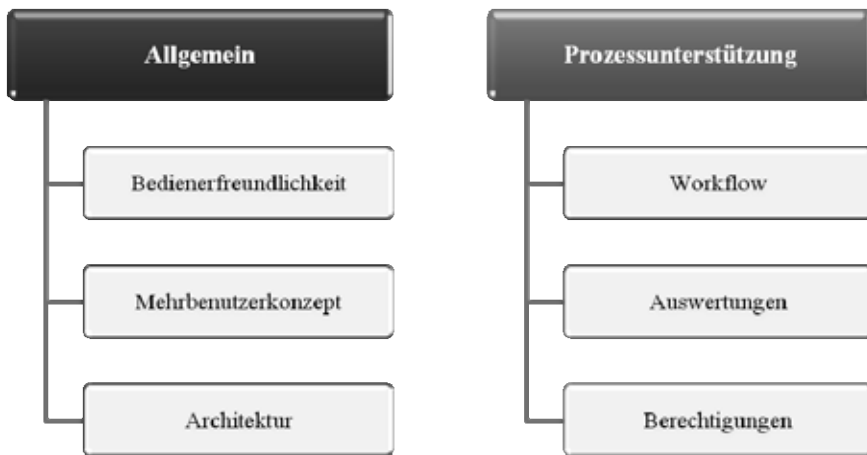


Abbildung 1: Kriterien für eine Software

Diese gelten natürlich auch für eine Applikation zur Durchführung von Retooling-Maßnahmen und sind unter „Allgemein“ in Abbildung 1 aufgeführt. Gerade im Hinblick auf Bedienerfreundlichkeit muss die Software ausgereift sein, da andernfalls die Effizienz und Effektivität sinkt oder die Arbeitsergebnisse sich verschlechtern können, wie [Her00 S. 15] beschreibt. Unter Hinblick auf das Mehrbenutzerkonzept empfiehlt sich die sogenannte Drei-Schichten-Architektur. Deren Vorteile sind laut [Sch00 S. 23f] und [Div00]:

- Paralleles Arbeiten mehrerer Anwender
- Einfache Skalierbarkeit der Applikation
- Möglicher redundanter Aufbau zur Erhöhung von Ausfallsicherheit und Verfügbarkeit
- Leichtere Wartbarkeit des Systems

Darüber hinaus muss eine Software weitere Aspekte abdecken, um den komplexen Prozess rund um Retooling abzubilden. Diese sind unter „Prozessunterstützung“ in Abbildung 1 aufgelistet.

2.2 Betriebswirtschaftliche Vorteile durch Einsatz einer Software

Es ist durchaus denkbar, dass der Prozess für eine Retooling-Entscheidung auch ohne Software abzubilden ist. Allerdings fehlt bei einer solchen Vorgehensweise meist die Dokumentation darüber, welche Faktoren zur finalen Entscheidung führten.

Um Vergleiche zwischen Equipments vom selben Typ zu ermöglichen, müssen die Strukturen der Informationen vorgegeben sein. Beispielsweise sind Inventarnummer und Neupreis bei allen Komponenten vorhanden und als Pflichtfelder zu definieren. Andere Informationen können aber typspezifisch sein. Damit ergibt sich die Anforderung, dass die Software Pflichtfelder nicht nur global und damit Typ-unabhängig zulässt. Vielmehr ist beispielsweise für ein Hubshuttle wichtig, wie weit es Bauteile maximal transportieren kann. Somit lassen sich die benötigten Informationen kaum mehr mit Standardsoftware wie einer Tabellenkalkulation verarbeiten, da dort Abweichungen in der Informationsstruktur nicht oder nur unter bestimmten Auflagen (z. B. verschiedene Arbeitsblätter) möglich sind und bei großen Datenmengen unübersichtlich werden.

Eine speziell für Retooling-Maßnahmen entwickelte Software unterstützt den Prozess auch dadurch, dass mehrere Personen die Informationen bereitstellen können. Entsprechend der verschiedenen Aufgabenbereiche empfiehlt es sich, dass die Software unterschiedliche Ansichten bereitstellt, um ein effizientes Arbeiten zu ermöglichen. Beispielsweise erleichtert die exklusive Anzeige der zu befüllenden Felder die Arbeit für den Erfasser von vorhandenen Anlagen, da für ihn nicht relevante Eigenschaften nicht sichtbar sind.

Ein umfassendes Berichtswesen sorgt außerdem dafür, dass Beschlüsse jederzeit nachvollzogen werden können. Die Auswertungen sollten dabei auf betriebswirtschaftlichen Verfahren aufsetzen. Sinnvoll sind im Zusammenhang mit Retooling die Verfahren Total Cost of Ownership (TCO) und Life Cycle Costing (LCC), wie sie [Div01] beschreibt.

Wichtig für die Berechnung des Gesamtaufwandes sind hierbei die indirekten Kosten, die beispielsweise durch Mitarbeiterschulungen entstehen. Beim LCC-Verfahren werden zusätzlich die Kosten berücksichtigt, die während der Laufzeit der Anlage anfallen.

3 Zusammenfassung und Ausblick

Abschließend bleibt festzuhalten, dass die noch junge Thematik um Retooling von Altanlagen das Potential in sich trägt, in den nächsten Jahren der Automobilindustrie immense Kosten zu sparen. Allerdings sollten die Unternehmen dieses Thema nicht überstürzt angehen, sondern vorab ein schlüssiges Konzept entwickeln, um unvorhergesehene Kosten zu verhindern. Andernfalls birgt die vermeintliche Aufwandssenkung durch Retooling das Risiko von schlecht kalkulierbaren nachträglichen Folgekosten.

Literaturverzeichnis

- [Div00]. **Diverse**. Schichtenarchitektur. *Wikipedia*. [Online] 16. August 2009. [Zitat vom: 17. August 2009.] <http://de.wikipedia.org/wiki/Schichtenarchitektur>.
- [Div01]. **Diverse**. Total Cost of Ownership. *Wikipedia*. [Online] 17. August 2009. [Zitat vom: 19. August 2009.] http://de.wikipedia.org/wiki/Total_Cost_of_Ownership.
- [Her00]. **Herczeg, Michael**. *Software-Ergonomie: Grundlagen der Mensch-Computer-Kommunikation*. München : Oldenbourg, 2005.
- [Sch00]. **Schill, Alexander und Springer, Thomas**. *Verteilte Systeme*. Berlin [u.a.] : Springer, 2007.

PhotoBird – Entwicklung eines Corporate Designs anhand einer multimedialen Anwendung

Benedikt Langner

Fachhochschule Brandenburg
Informatik und Medien
langner@fh-brandenburg

Art der Arbeit: Bachelorarbeit

Betreuer/in der Arbeit: Prof. Alexander Urban, Marco W. Linke

Abstract: Die Zielsetzung dieser Arbeit ist die Sensibilisierung von kleinen Unternehmen für Corporate Design (CD). Dazu werden Grundbestandteile des CDs erläutert und für das Start-Up Unternehmen PhotoBird beispielhaft umgesetzt. Besonders das User-Centered Design Prinzip wird für die Anwendungsentwicklung hervorgehoben.

1 Einleitung

Unternehmen nutzen Visitenkarten, animierte Anwendungen und viele weitere Maßnahmen zur Kundenwerbung. In einigen Fällen verfügen diese Werbemaßnahmen über kaum konstante Strukturen, die eine Wiedererkennung in der Öffentlichkeit zulässt. Besonders kleine Unternehmen sind davon betroffen. Diese versuchen die schrumpfende Konsumentenzahl durch weitere Werbemaßnahmen entgegen zu wirken, anstatt sich mit dem Urproblem – mangelnde Selbstbetrachtung – zu beschäftigen (vgl. [We10]).

Hier setzt die Arbeit an und versucht kleine Unternehmen für Corporate Identity (CI), speziell Corporate Design (CD) zu sensibilisieren, indem es die Vorteile und Nutzen von CD aufzeigt. Zur Verdeutlichung der Ergebnisse wird ein CD für das Start-Up Unternehmen PhotoBird entwickelt und anhand einer multimedialen Anwendung und einer reduzierten Geschäftsausstattung umgesetzt.

Die Zielsituation für das Unternehmen PhotoBird ist der Markteintritt, die Positionierung des Produktes PhotoBird (digitaler Photorahmen) und die Nutzung der durch CD erreichbaren Synergieeffekte. Die multimediale Anwendung ist hier das tragende Marketingvorhaben für den Markteintritt und dient zur Information über das Produkt PhotoBird.

2 Corporate Design

CD bietet die Stabilisierung und Steigerung des Unternehmenswertes in der Öffentlichkeit (vgl. [He09 : 43ff.]). Besonders große Unternehmen wie Capgemini sd&m AG haben für sich CD entdeckt, nun müssen auch kleine Unternehmen nachziehen um die vorteilhaften Wettbewerbsfaktoren nutzen zu können. Dazu gehört eine stetige und konstante Präsenz und Erinnerung am Markt, Kostenreduktion durch einfache Reproduzierbarkeit von CD-Elementen, wie Flyer und Präsentationen sowie die Nutzung abteilungsübergreifender Synergieeffekte. Um diese Potenziale auszuschöpfen, müssen die Unternehmensleitung und das Corporate Identity Management (Legislative) mit den Mitarbeitern (Exekutive) eng zusammenarbeiten um den Konsumenten (Judikative) das CD fassbar zu machen (siehe Abbildung 1).

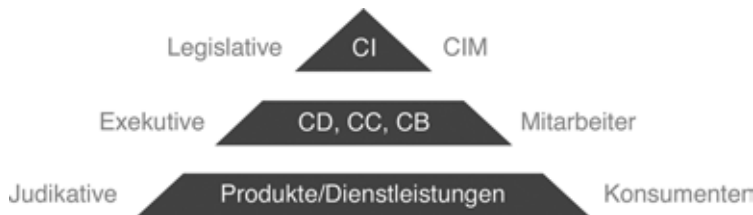


Abbildung 1: Gewaltenteilung eines Corporate Designs, angelehnt an Dunkl [Du98 : 14]

Hauptbestandteile eines CDs sind das Logo, die Typografie, die Farben und ein Gestaltungsraster, welches die Elemente für alle Medien anordnet und somit die Konstanz des CDs bewahrt.

Nach Herbst und Quednau (vgl. [Qu05 : 62f], [He09 : 18ff.]) sollte das Logo eines Unternehmens unter anderem intern wie extern repräsentieren, einprägsam und merkfähig sein sowie beim Betrachter einen Vertrauensvorschuss erzeugen, um zukünftigen Produkten einen leichteren Markteintritt zu ermöglichen. Eine Schriftart enthält neben der Informationsebene auch eine emotionale Ebene, die im Idealfall mit der Information übereinstimmt und somit eine einheitliche Stimmung erzeugt (vgl. [He09 : 38]). Gleiches gilt für die Hausfarben des Unternehmens.

3 Umsetzung

Aufbauend auf der Zielsetzung des Unternehmens lässt sich die multimediale Anwendung in dem Informationssektor (vgl. [Wa08]) mit zusätzlichen Promotion Charakter einordnen. Aufgrund der Parallelisierung von Design-, Konzeptions- und Programmierungsphase wurde (vgl. [vG09 : 20-29]) für die Anwendungsumsetzung die agile Softwareentwicklung in Verbindung mit dem User-Centered Design Prinzip (vgl. [W309]) verwendet. Zentraler Ansatzpunkt war hier die Vorgehensweise nach Garret [Ga02]. Dabei werden unter anderem User Models zur Modellierung von Use Cases genutzt. Hinzu kamen Anwendungskontrollen auf Usability und Accessibility durch Expertentest sowie Prototyping. Insbesondere Wireframes und Papier Prototyping boten

neben dem klassischen Pflichtenheft und Ablaufdiagrammen (vgl. [BBS06 : 127]) eine gute Möglichkeit zur frühen Kontrolle und Visualisierung der Anforderungen.

Technikaffine Personen zwischen 15 und 30 Jahren sind die Hauptzielgruppe. Diese verfügen über ein durchschnittliches Einkommen, was ihnen in bestimmten Abschnitten eine riskante Investition (z.B. neue Technologie, neue technische Spielerei) erlaubt. Somit gehören sie zu der Gruppe der Early Adopters. Zusätzlich besitzt die Zielgruppe aufgrund hohem Interesse an Innovationen über eine ausgeprägte Erfahrung mit dem Umgang neuer Medien. Dies führt zu einer hohen Erwartungshaltung gegenüber dem PhotoBird, dem CD und der Anwendung.

Die genutzte Technologie zur Anwendungsentwicklung ist das Adobe Flash Framework mit dem Adobe Flash Player. Die Entscheidung basiert auf der sehr großen Verbreitung (vgl. [Ad09], [Rs09]) des Players, der Zukunftssicherheit (vgl. [We09 : 24-32]) und einem hohen Standardisierungsgrad (Skalierbarkeit, Erweiterbarkeit, OOP) des Frameworks. Das Top-Down Verfahren zur Festlegung der Informationsarchitektur (vgl. [RM07 : 4]) erschien als das Geeignetste, da es neben der klaren Kategorisierung der Inhalte auch die Möglichkeit der hierarchischen Navigationsstruktur anbot. Aufbauend auf den Anforderungen nach Herbst und Quednau zeigt Abbildung 2 einen Teil der Umsetzungen zum PhotoBird.

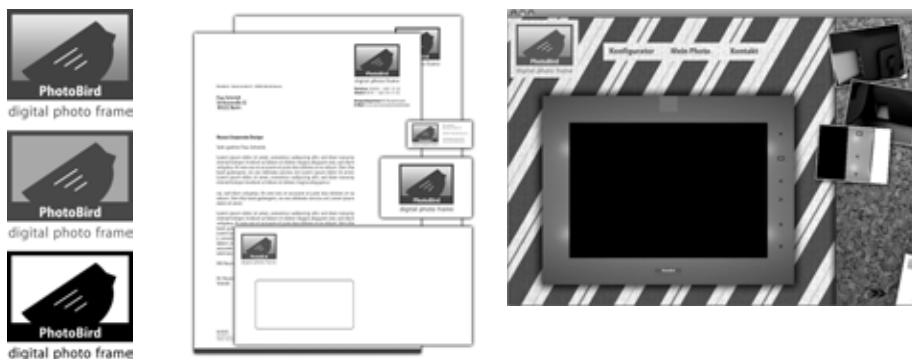


Abbildung 2: v.l.n.r. Logo (CMYK, Vollton, Schwarz/Weiß), Teil der Geschäftsausstattung, multimediale Anwendung

4 Fazit

Entscheidend für die Annahme eines Unternehmens und deren Marketingmaßnahmen in der Öffentlichkeit sind eine breite Anzahl von verschiedenen Eindrücken und deren emotionale Wirkung, die zu einem ganzheitlichen Bild (Corporate Image) von den Konsumenten unbewusst zusammengefügt werden – dies zeigt die Umsetzung zu PhotoBird. Jedoch wird deutlich, dass es auch für kleine Unternehmen möglich ist eine konstante und differenzierbare CD-Strategie zu entwickeln und umzusetzen. Zusammenfassend kann eine gut geplante und strukturierte CI/CD-Strategie eine Investition in die Zukunft sein, solange sich das Unternehmen selbst treu bleibt.

Literaturverzeichnis

- [Ad09] Adobe Flash: Content for Flash Player reaches 99% of Internet-enabled desktops in: Adobe – Flash Player Statistics, 2009, http://www.adobe.com/products/player_census/flashplayer/, 22. Dezember 2009.
- [BBS06] Böhringer, J., Bühler, P., Schlaich, P.: Kompendium der Mediengestaltung für Digital- und Printmedien. Berlin Heidelberg: Springer-Verlag 2006.
- [Du98] Dunkl, M.: Corporate Design Praxis. Das Handbuch der visuellen Identität von Unternehmen. Wien: Druckerei Robitschek & Co. Ges.m.b.H. 1998.
- [Ga02] Garrett, J. J.: The Elements of User Experience. User-Centered Design for the Web. Berkeley: New Riders 2002
- [He09] Herbst, D.: Corporate Identity. Aufbau einer einzigartigen Unternehmensidentität Leitbild und Unternehmenskultur Image messen, gestalten und überprüfen. Berlin: Cornelsen Verlag Scriptor GmbH & Co. KG 2009.
- [Qu05] Quednau, A.: Corporate Design – Ein Kursbuch. Norderstedt: Books on Demand GmbH 2005.
- [RM07] Rosenfeld, L.; Morville, P.: Information Architecture for the World Wide Web. Sebastopol: O'Reilly Media 2007.
- [Rs09] Riastats (Hrsg.): Rich Internet Application Statistics, in: riastats, 2009, <http://www.riastats.com>, 14. Januar 2010.
- [vG09] von Gager, S.: Konzeptionstechniken Wireframes. In: Page. Ideen und Know-how für Design, Werbung, Medien 09.2009, Ebner Verlag GmbH & Co. KG, 2009.
- [Wa08] Walder, P.: Multimedia Development. Vorlesungsskript zur gleichnamigen Lehrveranstaltung an der University of the West of Scotland, Paisely, 2008/2009.
- [W309] W3C: Notes on User Centered Design Process (UCD), 2004, <http://www.w3.org/WAI/EO/2003/ucd>, 22. Dezember 2009.
- [We09] weave. Interactive design, konzeption & development, 05.09, Page Verlag, 2009.
- [We10] Gedächtnisprotokoll zum Interview mit Ina Jahnel-Werner und Marco Mehlhorn, Geschäftsführer der werbefactory, vom 11. Januar 2010

Realisierung einer Anbindung von SAP ERP zu SAP EWM auf Basis SAP SCM und Dokumentation von Technik, Systemen und Schnittstellen zur Abbildung eines ausgewählten Geschäftsprozesses

Michael Urmann

Hochschule Deggendorf
Fakultät Betriebswirtschaft und Wirtschaftsinformatik
michaelurmann@gmx.de

Art der Arbeit: Diplomarbeit
Betreuer der Arbeit: Prof. Dr.-Ing. Herbert Fischer

Abstract:

Im Rahmen der Diplomarbeit wird ein Systemverbund von SAP-ERP und SAP-SCM realisiert, in dem die Funktionalität des SAP Extended Warehouse Managements zur Verfügung gestellt wird. Neben der praktischen Umsetzung der Systemanbindung werden die technischen sowie betriebswirtschaftlichen Hintergründe, die Voraussetzungen und die Anforderungen dokumentiert, sowie an einem betriebswirtschaftlichen Beispielprozess veranschaulicht.

1 IT und Logistik

Die Definition von Logistik ist weitläufig bekannt: Es ist die wissenschaftliche Lehre der Planung, Steuerung und Überwachung der Material-, Personen-, Energie- und Informationsflüsse in Systemen. Dabei besteht die logistische Aufgabe darin, die richtige Menge der richtigen Objekte als Gegenstände der Logistik (Güter, Personen, Energie, Informationen) am richtigen Ort im System (Quelle, Senke) zum richtigen Zeitpunkt in der richtigen Qualität zu den richtigen Kosten zur Verfügung zu stellen.

Um diesen komplexen Anforderungen gerecht zu werden und um im Zeitalter der Globalisierung und immer rascher wachsender Volkswirtschaften bestehen zu können, wird es für Unternehmen immer wichtiger, ihre logistischen Prozesse auszubauen und zu optimieren. Die Logistik eines Unternehmens beschränkt sich also längst nicht mehr nur auf die reine Lagerhaltung und den Gütertransport von A nach B. Vielmehr werden die logistischen Prozesse immer komplexer und vielseitiger, so dass eine starke Vernetzung innerhalb der ganzheitlichen Unternehmensprozesse notwendig wird.

Eine derartige Vernetzung ist allerdings nicht nur intern von großer Bedeutung, sondern auch unternehmensübergreifende, integrierte Logistikprozesse, in denen das eigene Unternehmen Hand in Hand mit den externen Geschäftspartnern wie Lieferanten und Kunden zusammenarbeitet, können Wettbewerbsvorteile schaffen und erheblich zum Erfolg des Unternehmens beitragen.

Während in der letzten Zeit das Hauptaugenmerk in der Optimierung der zwischenbetrieblichen Abläufe und der Bedarfs- und Absatzplanung lag, gerät nun wieder zunehmend die eigene Lagerverwaltung ins Blickfeld. Zwar wird schon immer versucht, in der Lagerlogistik die Prozesse zu optimieren und dies wird auch weitläufig umgesetzt: chaotische Lagerhaltung, dynamische Bestandsführung und Kommissionierung sind nur wenige Bestandteile aktueller Systeme, mit denen Unternehmen heute ihre Läger verwalten. Durch neue wirtschaftliche Anforderungen und den technischen Entwicklungen wird nun weiteres Verbesserungspotential in der Lagerverwaltung gesehen, welches erst mit entsprechender IT-Unterstützung bestmöglich ausgeschöpft werden kann.

Beispiele für derartige Erweiterungen und Optimierungsansätze sind wegeoptimale Lagerraumdisposition, Ressourcen- und Arbeitsmanagement, flexibles Handling Unit Management, Materialflusssysteme oder die Nutzung von mobilen Endgeräten.

Vorraussetzung für den reibungslosen Ablauf derartiger komplexer Prozesse ist allerdings eine entsprechende, leistungsfähige IT-Systemlandschaft, die größtmögliche Flexibilität und Schnelligkeit bei relativ geringen Kosten bietet. Doch gerade bei großen und umfangreichen IT-Systemen besteht die Gefahr, dass zahlreiche Insellösungen geschaffen werden. Dabei sind die einzelnen Systeme zwar optimal für den jeweiligen Geschäftsprozess ausgerichtet, in ganzheitlicher Betrachtung allerdings gehen die damit gewonnenen, betriebswirtschaftlichen Vorteile durch die steigenden Implementierungs- und Wartungskosten der Schnittstellen wieder verloren. Eine einheitliche IT-Strategie mit einer homogenen Systemlandschaft und klar definierten und standardisierten Schnittstellen trägt dazu bei, den Wartungsaufwand und damit Kosten zu senken.

2 Gründe und Anforderungen zur Verbindung von IT-Systemen

Die Geschwindigkeit in den weltweiten Wirtschaftsabläufen steigt. Immer kürzer werdende Innovationszyklen, zunehmender Wettbewerb und Globalisierung sowie zahlreiche Fusionen zwingen Unternehmen, ihre Geschäftsmodelle in immer kürzeren Zeitabständen zu überdenken und anzupassen. Diese veränderten Geschäftsanforderungen müssen mithilfe adäquater IT-Lösungen einfach, schnell und zu marktgerechten Preisen umgesetzt werden. Deshalb ist es wichtig, die Unterstützung durch IT-Systeme so zu gestalten, dass diese ebenso flexibel und universell einsetzbar sind, wie es die Geschäftsprozesse erfordern. Jedoch führen diese permanenten Veränderungen der fachlichen Anforderungen zu heterogenen und komplexen IT-Landschaften, deren Wartung zunehmend schwieriger und kostenintensiver wird.

Genau dies stellt einen nicht zu unterschätzenden Zielkonflikt dar: einerseits sollen IT-Systeme alle Besonderheiten und Spezialfälle des Unternehmens abdecken, um die einzelnen Geschäftsprozesse kostenoptimal abzuwickeln, wodurch allerdings auch die IT-Kosten steigen. Andererseits sind standardisierte Lösungen wesentlich kostengünstiger, bieten aber geringere Individualität und Flexibilität. Hier einen geeigneten Mittelweg zu finden, der die notwendige Flexibilität bietet und dennoch kostengünstig ist, das ist die Herausforderung, der sich die Verantwortlichen in den Unternehmen stellen müssen.

Doch auch die System- und Produktanbieter haben diesen Missstand erkannt und entwickeln ihre Angebote dahingehend, dass ihre Standardprodukte flexibel und individuell anpassbar werden. Dies war zwar bereits seit langem möglich, doch bisher waren dazu immer zeit- und kostenintensive Eigenentwicklungen notwendig, um das gewünschte Ziel zu erreichen. Kurzfristig ist das Ergebnis durchaus akzeptabel, langfristig gesehen führt dies allerdings häufig zu Problemen: Die IT-Landschaft ist im Laufe von Jahrzehnten gewachsen. Systeme, die funktionieren, werden nach dem Grundsatz „Never touch a running system“ nur verändert, wenn es unbedingt notwendig ist, Erweiterungen werden nur hinzugefügt, anstatt sie zu integrieren und neue Technologien werden implementiert, ohne dass ältere Technologien abgebaut werden. Dies führt zu einer unüberschaubaren Vielzahl von Schnittstellen und einer technologischen Vielfalt, die mit zunehmendem Alter immer schwieriger zu warten, zu erweitern oder gar abzulösen sind.

Mittlerweile stellen fast alle namhaften Hersteller Produkte und Lösungen bereit, die an diesem Problem ansetzen. Anstatt viele verschiedene, in sich abgeschlossene Systeme zu führen, werden die Produkte durch einzelne Module offener und flexibler. Definierte und standardisierte Schnittstellen helfen zudem, die Verzahnung der Systeme zu verbessern und vor allem die Wartung und die Integration neuer Komponenten zu vereinfachen.

Somit besteht die Möglichkeit für die Unternehmen, eine homogene IT-Landschaft aufzubauen und zu betreiben, die zwar viele verschiedene, spezialisierte Systeme besitzt, alles in allem aber einen gemeinsamen Systemverbund darstellt.

3 SAP Extended Warehouse Management

Im Rahmen einer praktischen Implementierung von SAP Extended Warehouse Management beschreibt diese Diplomarbeit die Konzipierung und Integration in die Systemlandschaft. Der Aufbau der Arbeit ist dabei an die Vorgehensweise angelehnt wie bei einer Implementierung vorzugehen ist, sodass in einer Art Handbuch die notwendigen Schritte und die dazugehörigen Hintergrundinformationen vorliegen, um ein Einführungsprojekt zügig und kostengünstig abwickeln zu können.

Ablauf der Systemverbindung:

1. Verbindung der Systeme über RFC
2. liveCache Aktivierung
3. Technische Konfiguration
4. Betriebswirtschaftliche Konfiguration
5. Core Interface und Integrationsmodell
6. Customizing im EWM
7. Monitoring und Fehleranalyse

Nach Abschluss der Anbindung der Systeme und des Customizings werden anhand eines betriebswirtschaftlichen Beispielprozesses, die grundlegenden Funktionen und deren Abwicklung dargestellt. Dazu wird ein typischer Prozess abgebildet, der im Produktiveinsatz eines derartigen Systemverbundes alltäglich ist: der Wareneingang mit Einlagerung ins Lager und danach die erneute Auslagerung mit Warenausgang.

Betriebswirtschaftlicher Beispielprozess

1. Wareneingang
 - 1.1. Bestellung im ERP
 - 1.2. Anlieferung im ERP
 - 1.3. Anlieferungsbenachrichtigung im EWM
 - 1.4. Anlieferung im EWM
 - 1.5. Umlagerung
2. Warenausgang
 - 2.1. Kundenauftrag im ERP
 - 2.2. Auslieferung im ERP
 - 2.3. Auslieferungsauftrag im EWM

Literaturverzeichnis

- [Plü03] Plümer, T.: Logistik und Produktion. Oldenbourg Wissenschaftsverlag, München, 2003.
- [Han09] Hanschke, I.: Strategisches Management der IT-Landschaft. Carl Hanser Verlag, München, 2009.
- [Bar02] Bartsch, H.; Bieckenbach, P.: Supply Chain Management mit SAP APO. Galileo Press, Bonn, 2002.
- [Sap09] SAP Help Portal: SAP Dokumentation EWM 7.0
http://help.sap.com/saphelp_ewm70/helpdata/de/58/c69040bca2ef4ae10000000a1550b0/frameset.htm, Zugriffsdatum: Juni 2009.

Fachliche Modellierung von multidimensionalen Datenstrukturen im Business Intelligence Umfeld einer öffentlichen Verwaltung

Doreen Wehr

Hochschule Deggendorf
Fakultät Betriebswirtschaft und Wirtschaftsinformatik
doreen.wehr@gmx.de

Art der Arbeit: Masterarbeit

Betreuer/in der Arbeit: Prof. Dr.- Ing. Herbert Fischer

Hochwertige Entscheidungen von Unternehmen setzen eine qualitativ gute Informationsbasis voraus, auf der notwendige Entscheidungen getroffen werden können. Nicht zuletzt hängt davon der Unternehmenserfolg bzw. Unternehmensmisserfolg ab. „Es werden Informationen benötigt, welche die Entwicklung des Unternehmens über einen bestimmten Zeitraum [in der Vergangenheit] repräsentieren und zur Erkennung und Untersuchung von Trends [für die Zukunft] herangezogen werden können.“ [Böhn01, S. 1]

Trotz großer Datenflut und großem Volumen an Daten herrscht in Unternehmen oft ein Informationsdefizit vor. Hier sind Lösungen gefragt, welche in der Lage sind, die wachsende Datenflut zu strukturieren, kanalisieren und entsprechend bereitzustellen [Böhn01, S. 2]. Ein möglicher und viel diskutierter Ansatz zur Informationsversorgung stellt dabei das Data Warehouse Konzept dar [Inmo96, S. 33]. Wesentlicher Erfolgsfaktor eines solchen Data Warehouses ist hierbei die Spezifikation der fachlichen Anforderungen [Böhn01, S. 2]. Es existiert bereits eine Vielzahl an grafischen Modellierungsansätzen, oft als sogenannte semantische Modellierung bezeichnet, welche die Informationen entsprechend abbilden kann.

Data Warehouses, multidimensionale Datenbanken und Online Analytical Processing (OLAP) Anwendungen werden dazu verwendet, wichtige und auch kritische Geschäftsinformationen in Form von Kennzahlen bereitzustellen. Wie lassen sich multidimensionale Datenstrukturen in einem Business Intelligence Umfeld geeignet fachlich so modellieren, dass sie sowohl für verschiedene Fachbereiche, die IT als auch für koordinierende Stellen, z. B. für das Requirement Engineering, eine hilfreiche Informationsquelle darstellen?

Dazu werden in dieser Masterarbeit bereits bestehende Ansätze und Methodiken der multidimensionalen Modellierung vorgestellt. Wobei angemerkt werden muss, dass sich derzeit immer noch kein Standard zur Modellierung in diesem Bereich etabliert hat [Rizz09, S. 2]. Es wird zunächst auf die Notwendigkeit fachlicher Modellierung näher eingegangen, bevor ein eigener Modellierungsansatz vorgestellt wird, der unter den besonderen Gegebenheiten einer öffentlichen Verwaltung entwickelt wurde und unter anderem die Anforderung zu Governancezwecken (Steuerung durch vordefinierte Richtlinien) im Bereich Business Intelligence erfüllt. Der Modellierungsansatz wird an einem Beispiel aus verschiedenen Blickwinkeln näher erläutert und vorgestellt.

Nicht alle Ansätze der Modellierung sind immer gleichermaßen für alle Gegebenheiten gleichgut geeignet. Dies führte zur Entwicklung eines eigenen Modellierungsansatzes, des Semantischen Modells (SeMo). Dieser Ansatz basiert auf drei Diagrammtypen, die ausgehend von den strukturellen Zusammenhängen der einzelnen Modellelemente hin zu einer schrittweisen Verfeinerung des jeweiligen Sachverhaltes (Dimensionsstruktur, Kennzahlentransformation) in entsprechenden Diagrammtypen erfasst wird. Diese Sichtenbildung trägt explizit zur Komplexitätsbeherrschung bei. Jeder Diagrammtyp besitzt dabei einen spezifischen Fokus auf das Gesamtdatenmodell. Innerhalb eines Dimensions-diagramms wird der Aufbau der Strukturen einer Dimension definiert. Das Strukturdiagramm stellt den Zusammenhang zwischen Fakten und Dimensionen dar. Im dritten Diagrammtyp, dem Transformationsdiagramm wird die Herleitung der im SeMo aufgezeigten Kennzahlen abgebildet.

Der resultierende Nutzen des entwickelten Modellansatzes kann, in drei wesentliche Kategorien eingeteilt und zusammengefasst werden: als Werkzeug, zur Wissensvermittlung und um Prozessnutzen zu erzielen.

Als Werkzeug arbeitet es zum einen dadurch, dass durch eine formalisierte Darstellung fachlicher Inhalte in SeMo Notation Interpretationsspielräume vermieden und Missverständnisse entgegen gewirkt werden kann (Präzisierung). Ein weiterer Nutzenaspekt stellt die Standardisierung durch die Vereinheitlichung von unterschiedlichen Ausprägungen dar. Durch eine Ist-Aufnahme in der Notation können z. B. Alterszusammenfassungen konsolidiert und vereinheitlicht werden. Eine Validierung, beispielsweise die Prüfung auf Existenz bzw. korrekte Verknüpfungen oder Prüfung auf korrekte Granularitäten, kann auf Grundlage des SeMo durchgeführt werden. Außerdem kann in dieser Rubrik das Argument der Wiederverwendung angeführt werden (z. B. Hierarchien).

Bezüglich des Themas der Wissensvermittlung leistet es hinsichtlich der transparenten Darstellung der fachlichen Inhalte einen wesentlichen Beitrag, da so die Analyse von übergreifenden Zusammenhängen ermöglicht wird. Darüberhinaus werden sogenannte „Kopfmonopole“ durch diese Art der Dokumentierung aufgelöst und damit eine schnellere Einarbeitung bei internen und externen Mitarbeitern ermöglicht. Momentan ist die Einarbeitungszeit auf Grund der speziellen Thematik und Komplexität im BI-Umfeld vergleichsweise sehr hoch. Hier kann durch den Einsatz des SeMo eine Reduzierung herbeigeführt werden. Ein weiterer Punkt, bei dem das SeMo nutzenbringend eingesetzt werden kann, ist bei der Reduktion von Analyse von Request for Changes (RfCs). Möglicherweise machen die im SeMo dargestellten Inhalte eine ausführliche Analyse überflüssig, weil beispielsweise die Granularität der vorliegenden Daten durch den Fachbereich selbst oder die koordinierenden Stellen schnell und einfach ermittelt werden kann.

Aber letztlich sind es Punkte wie das Sprechen und Verwenden einer gemeinsamen Sprache, welches zu einer effektiveren Kommunikation führt und das leichtere Erkennen und damit Vermeiden von Redundanzen (Harmonisierung), welches die einleuchtenden Argumente darstellen, die für das SeMo und seinen Einsatz sprechen. Langfristig gesehen, trägt das zu einer schnelleren Abwicklung von RfCs und auch zu einer höheren Qualität und höherem Durchsatz bei, weil der Prozess durch transparente, übersichtliche und verständliche Darstellung der Fachlichkeit erheblich verbessert werden kann. Das SeMo stellt außerdem einen Ordnungsrahmen zur Ausrichtung der BI-Governance dar.

Am Ende dieser Arbeit findet eine kritische Betrachtung der entwickelten Modellierungsmethodik, ihrer Etablierung in den Fachbereichen und in der IT statt. Welche Erfolgsfaktoren, Risiken und Empfehlungen lassen sich daraus notwendigerweise ableiten? Welche Weiterentwicklungen sollten ggf. noch vorgenommen werden?

Literaturverzeichnis

Hinweis: In diesem Verzeichnis wurden sowohl Print- wie Online-Ressourcen berücksichtigt.

- [Abel02] Abello, A.: YAM2: A Multidimensional Conceptual Model, PhD Thesis, Departament de Llenguatges i Sistemes Informàtics Universitat Politècnica de Catalunya, 2002.
- [AbSS02] Abello, A.; Samos, J.; Saltor, F.; YAM2: a multidimensional conceptual model extending UML, article in press, 2002, www.sciencedirect.com, Zugriffsdatum: 05.05.2009.
- [BaGü04] Bauer, A., Günzel, H.: Data-Warehouse-Systeme, Architektur – Entwicklung – Anwendung, 2., überarbeitete und aktualisierte Auflage, dpunkt Verlag, Heidelberg, 2004.
- [BASc07] Bundesagentur für Arbeit, Schulungsunterlage „Data Warehouse für Anwender“, Interne Quelle, Stand Juni 2007.
- [Biss00] Bisseck, P.: Modellierung und Verwaltung zeitbezogener Informationen in relationalen Datenbanken: Konzepte und Sprachen: Dissertation, Rheinische

- Friedrich-Wilhelms-Universität Bonn, 2000.
- [Böhn01] Böhnlein, M.: Konstruktion semantischer Data-Warehouse-Schemata, Deutscher Universitäts Verlag, 2001.
 - [ChGl06] Chamoni, P., Gluchowski, P.; Analytische Informationssysteme – Business Intelligence-Technologien und –Anwendungen, 3. Auflage, Berlin – Heidelberg, Springer Verlag, 2006.
 - [DoLe05] Dombrowski, E.; Lechtenbörger, J.: Evaluation objektorientierter Ansätze zur Data-Warehouse-Modellierung, in Datenbank-Spektrum 15/2005.
 - [Enzy09] Enzyklopädie der Wirtschaftsinformatik – Online Lexikon: URL: <http://www.oldenbourg.de:8080/wi-enzyklopaedie/lexikon/daten-wissen>: Zugriffsdatum: 22.03.09, Oldenbourg Wissenschaftsverlag.
 - [GaGl98] Gabriel, R./ Gluchowski, P. (1998): Grafische Notationen für die semantische Modellierung multidimensionaler Datenstrukturen; in: Management Support Systemen; in Wirtschaftsinformatik, Jahrgang 40, Heft 6, 1998, S. 493-502.
 - [Goek06] Goeken, M.: Entwicklung von Data-Warehouse-Systemen – Anforderungsmanagement, Modellierung, Implementierung, Deutscher Universitäts Verlag, 2006.
 - [GoRi98] M. Golfarelli, D. Maio, S. Rizzi. Conceptual Design of Data Warehouses from E/R Schemes. Proceedings 31st Hawaii International Conference on System Sciences (HICSS-31), vol. VII, Kona, Hawaii, pp. 334-343, 1998, In: URL: <http://www-db.deis.unibo.it/~srizzi/PDF/hicss98.pdf>, Zugriffsdatum: 29.03.2009.
 - [Hahn05] Hahne, M.: SAP Business Information Warehouse – Mehrdimensionale Datenmodellierung, Springer-Verlag, Berlin, 2005.
 - [Hett02] Hettler, D.: Semantische Datenmodellierung für Data Warehouse-Systeme aus Sicht des Controlling – ein Vergleich ausgewählter Modellierungsansätze, Diplomarbeit, Universität Mannheim, 2002.
 - [Holt99] Holthuis, J. (1999): Der Aufbau von Data Warehouse-Systemen: Konzeption – Datenmodellierung – Vorgehen, 2. Auflage, Deutscher Universitäts-Verlag und Gabler Verlag, Wiesbaden, 1999.
 - [HuWi05] Humm, B.; Wietek, F.: Architektur von Data Warehouses und Business Intelligence Systemen, Informatik Spektrum 23.02.2005, URL: http://www.de.capgemini-sdm.com/web4archiv/objects/download/fachartikel/sdm_pub_is_humm_dwh.pdf; Zugriffsdatum: 27.07.09.
 - [Inmo96] Inmon, W. H.: Building the Data Warehouse. Second Edition, John Wiley & Sons, New York, 1996.
 - [Kenn99] Kennel, A.: Data Warehousing, Wie viele Dimensionen hat ein Würfel? Der Einsatz von Datamarts als Analysehilfen in einem Data Warehouse, 1999, http://www.infokennel.ch/Publikationen/Grundbegriffe_1999.pdf, Zugriffsdatum: 02.06.2009.
 - [KiRo02] Kimball, R.; Ross, M.: The Data Warehouse Toolkit, The Complete Guide to Dimensional Modeling, Second Edition, Wiley Computer Publishing, John Wiley & Sons, Inc., New York, 2002.
 - [Lehm01] Lehmann, P.: Meta-Datenmanagement in Data-Warehouse-Systemen – Rekonstruierte Fachbegriffe als Grundlage einer konstruktiven, konzeptionellen Modellierung, Dissertation, Otto-von-Guericke-Universität Magdeburg, 2001.
 - [Lehn03] Lehner, W.: Datenbanktechnologie für Data-Warehouse-Systeme; Konzepte und Methoden, dpunkt.verlag, 2003.
 - [Maie98] Maier, R.: Nutzen und Qualität der Datenmodellierung – Ergebnisse einer empirischen Studie, in Wirtschaftsinformatik 2/1998, S. 130 – 149.
 - [MaZi06] Malinowski, E.; Zimanyi, E. (2006): Hierarchies in a Multidimensional Model: URL: <http://code.ulb.ac.be/dbfiles/MalZim2006article.pdf>; Zugriffsdatum: 11. April 2009.

Ein intelligentes Benachrichtigungssystem für das Incident Management

Birol Yildiz

Hochschule Bonn-Rhein-Sieg
Fachbereich Informatik
birol.yildiz@smail.inf.h-brs.de

Art der Arbeit: Master Thesis

Betreuer der Arbeit: Prof. Dr. Rudolf Berrendorf

Abstract: Diese Arbeit beschreibt die Architektur und Auswahlalgorithmen für ein intelligentes Benachrichtigungssystem für das Incident Management. Es handelt sich dabei um einen zentralen Benachrichtigungsdienst, welcher bidirektionale Kommunikation unterstützt und automatisch Alerts von verschiedenen Anwendungen an die Person weiterleitet, die am besten zur Lösung des spezifischen Problems geeignet ist. Hierzu werden dynamische Benachrichtigungsalgorithmen vorgestellt, die anhand von Kontext-Informationen (wie z.B. Fehlerart, Kenntnisse, Ort, Zeit, Verfügbarkeit, ...) die am besten geeignete Person zur Lösung des Problems bestimmen. Eine weitere Komponente des Systems iteriert gegebenenfalls dieses Verfahren, bis eine Person verantwortlich die Behebung des Problems akzeptiert hat.

1 Einleitung

Für Unternehmen, die ihren Kunden IT-Dienstleistungen anbieten, ist es wesentlich, über den aktuellen Status ihrer angebotenen Dienste in Kenntnis zu sein, unabhängig davon, wo sich das Personal gerade befindet. Monitoring-Tools wurden dazu konzipiert, den aktuellen Zustand der IT-Infrastruktur zu überwachen und Störungen bzw. den Ausfall von Diensten und Hardware-Komponenten zu detektieren. Informationen über den Zustand werden üblicherweise über web-basierte Reports mitgeteilt. Ein Pull-Mechanismus, über den Administratoren periodisch den Status abrufen und entsprechend reagieren, oder die permanente Überwachung über gebundenes Personal ist oft sehr ineffizient. Stattdessen bieten sich Push-Mechanismen an, die bei geeigneter Auslegung der Gesamtsystemarchitektur gezielt im Problemfall Personal zur Fehlerbehebung benachrichtigen, falls eine maschinelle Fehlerbehebung zum Beispiel aufgrund der Komplexität des Fehlers nicht möglich ist. Strategien, die zur Benachrichtigung eingesetzt werden, sind Massen-Benachrichtigung (jede in Frage kommende Person wird benachrichtigt), Einzel-Benachrichtigung und Benachrichtigung anhand von statischen Regeln (z.B. nach Dienstplänen). Diese Benachrichtigungsstrategien sind aus den folgenden Gründen meistens ineffektiv und ineffizient [3],[6],[7]: Es ist nicht immer

möglich, eine Person an einem fixen Arbeitsplatz zu erreichen, wenn ein Fehlerfall eintritt, oder die Benachrichtigung wurde nicht in Kenntnis genommen (z.B. während eines Meetings). Benachrichtigungssysteme müssen das Personal auf ihren mobilen Geräten erreichen können und bidirektionale Kommunikation unterstützen, so dass der Incident automatisch eskaliert wird, wenn die benachrichtigte Person untätig bleibt. Weiterhin sollte ein intelligentes Benachrichtigungssystem sicherstellen, dass die am besten geeignete Person benachrichtigt wird. D.h. die Person, die die richtigen Kenntnisse zur Lösung des Problems aufweist und wahrscheinlich am schnellsten reagieren kann. Dies optimiert den Einsatz und Zeitaufwand des Personals. In der Master Thesis [8] wird ein intelligentes Benachrichtigungssystem vorgestellt, welches die genannten Probleme berücksichtigt. Solch ein Benachrichtigungssystem schließt die Lücke zwischen Monitoring-Systemen und dem Support-Personal, das zeitnah auf Incidents reagieren muss. Der vorliegende Abstract geht aus Platzgründen nicht auf die Konzeption und Gesamtarchitektur des Benachrichtigungssystems ein, sondern erläutert im Wesentlichen einen der dynamischen Benachrichtigungsalgorithmen als Lösung der Kernproblematik.

2 Ein Auswahl- und Benachrichtigungsalgorithmus

Der Auswahl- und Benachrichtigungsalgorithmus ist ein wesentlicher Bestandteil des Benachrichtigungssystems. Er stellt sicher, dass im Falle eines Incidents der am besten geeignete Mitarbeiter benachrichtigt wird. Hierzu wird der aktuelle Kontext eines jeden in Frage kommenden Mitarbeiters ausgewertet und daraus jeweils ein Prioritätswert errechnet. Anhand der Prioritätswerte wird entschieden, wer benachrichtigt werden soll. Die derzeit für das Incident Management ermittelten relevanten Kontextelemente sind Ort, Zeit, Kenntnisstand, Verfügbarkeit und die aktuelle Auslastung eines Mitarbeiters, weitere wären flexibel integrierbar. Bei Kontextinformationen ist nicht der explizite Wert von Interesse (z.B. wo genau sich der Mitarbeiter gerade befindet), sondern lediglich, ob diese Information günstig für die Eignung des Mitarbeiters ist das Problem zu lösen oder sich eher negativ darauf auswirkt. Die Auswirkung einer Kontext-Information auf die Eignung eines Mitarbeiters wird über den sogenannten *Context Gain* ausgedrückt. Der Context Gain entspricht der mathematischen Repräsentation einer Kontext-Information und drückt seinen Effekt auf die Eignung eines Mitarbeiters aus, wegen eines Incidents benachrichtigt zu werden. Folgende Definition charakterisiert den Context Gain:

Definition (Context Gain):

Sei A ein Alert, e ein Mitarbeiter und c_{Ae} ein Kontext-Element bzgl. A und e . Der Context Gain von c_{Ae} ist eine reelle Zahl $g \in [-1,1]$ und drückt den Effekt von c_{Ae} auf die Eignung von e aus, um das Problem A zu lösen. Er hat die folgenden Eigenschaften:

- Der Effekt von c_{Ae} ist positiv, wenn $g > 0$
- Der Effekt von c_{Ae} ist negativ, wenn $g < 0$
- c_{Ae} hat keinen Effekt, wenn $g = 0$ ist. In diesem Fall hat c_{Ae} keinen Einfluss auf den Benachrichtigungsprozess.

Die Berechnung des Context Gains erfolgt spezifisch für jedes Kontext-Element unter Einbeziehung von Domänen-Wissen. Abbildung 1 zeigt beispielhaft den Verlauf eines Zeitkontext Gains.



Abbildung 1: Beispiel zum Verlauf eines Zeitkontext Gains

Die x-Achse stellt die Uhrzeit dar (0 – 24 Uhr) und auf der y-Achse ist der Context Gain aufgetragen. Die relevanten Zeitabschnitte sind die Arbeitszeit (9 – 18 Uhr), die Abendstunden (18 – 23 Uhr) und die Nachtzeit (23 – 8 Uhr). Ersteres hat einen hohen Context Gain Wert, da während der Arbeitszeit ein Mitarbeiter eher in der Lage ist sich um einen Incident zu kümmern als z.B. in der Nachtzeit. In den Übergängen der Zeitabschnitte verläuft der Context Gain linear.

Der Benachrichtigungsalgorithmus erwartet als Eingabe einen Alert und eine Menge von Mitarbeitern, die potentiell für die Lösung des Alerts in Frage kommen und bestimmt aus diesen Parametern den Mitarbeiter, der am besten zur Lösung des Problems geeignet ist. Dabei wird für jeden Mitarbeiter eine Priorität ausgerechnet, welcher die Eignung widerspiegeln soll. Diese Priorität ist wie folgt definiert:

Definition (Bedingte Benachrichtigungspriorität):

Sei A ein Alert, e ein Mitarbeiter, sowie $C_{Ae} = \{c_0, c_1, \dots, c_{n-1}\}$ n Kontextelemente bezüglich A und e . Weiterhin seien g_0, g_1, \dots, g_{n-1} die zugehörigen Context Gain Werte für C_{Ae} , wobei $g_i, 0 \leq i < n$, den Context Gain von $c_i, 0 \leq i < n$, bezeichnet. Die bedingte Benachrichtigungspriorität p_{Ae} bezüglich e und A ist wie folgt definiert:

$$p_{Ae} = \begin{cases} \frac{\sum_{i=0}^{n-1} a_i \cdot g_i}{\sum_{i=0}^{n-1} a_i}, & \text{if } h_i(g) \begin{cases} < \\ = \end{cases} 0 \text{ for } i = 0, \dots, m \text{ and } g = (g_0, g_1, \dots, g_{n-1}) \\ -\infty, & \text{else} \end{cases}$$

wobei $h_i(g), 0 \leq i < m$, Gleichheits- und Ungleichheitsbedingungen darstellen und die Koeffizienten $a_i > 0, 0 \leq i < n$, die Gewichtungen der Kontext-Elemente darstellen. Ist der Prioritätswert $-\infty$, so wird der jeweilige Mitarbeiter im Benachrichtigungsprozess nicht berücksichtigt. Der Wertebereich von p_{Ae} liegt im Intervall $[-1,1]$.

Über Gleichungen und Ungleichungen $h_i(g), 0 \leq i < m$ können nun notwendige Kriterien formuliert werden, die erfüllt sein müssen, damit ein Mitarbeiter benachrichtigt werden kann. Zum Beispiel kann das Kriterium „Ein Mitarbeiter darf nicht in der Nacht benachrichtigt werden“ über die Ungleichung $h(g) = -1 - g < 0$ in die Berechnung einbezogen werden, wobei g dem Zeitkontext Gain entspricht. Der am besten geeignete Mitarbeiter e_A^* bezüglich A ist dann derjenige Mitarbeiter mit der höchsten

Benachrichtigungspriorität. Der Benachrichtigungsalgorithmus benachrichtigt diesen Mitarbeiter und wartet für eine vorgegebene Zeit (abhängig von der Alert-Priorität) bis der Mitarbeiter den Empfang des Alerts bestätigt. Bleibt dieser jedoch untätig, wird der am zweit besten geeignete Mitarbeiter benachrichtigt. Dieses Verfahren wird solange wiederholt, bis mindestens einer der Mitarbeiter den Empfang bestätigt.

3 Fazit

Dieser Artikel hat einen Benachrichtigungsalgorithmus vorgestellt, der den Incident Management Prozess durch gezielte Benachrichtigung beschleunigen soll. Der Algorithmus nutzt Kontext-Informationen um diese Aufgabe zu erfüllen. In der Master Thesis [8] wird ein weiterer Benachrichtigungsalgorithmus beschrieben, der auf den Bayes' Klassifikator aus dem Umfeld des Maschinellen Lernens basiert. Eine Evaluierung beider Algorithmen hat gezeigt, dass der hier vorgestellte Algorithmus flexibler ist und bessere Ergebnisse liefert. Neben der Anwendung des Bayes' Klassifikator sind auch weitere Klassifikationsalgorithmen denkbar (z.B. Entscheidungsbäume, Neuronale Netze, Support Vector Machines, ...). Das in dieser Arbeit entwickelte System ist bei einem weltweit agierenden Unternehmen, bei dem das beschriebene Problemszenario akut ist (verschiedene Zeitzone, unterschiedliche Skills der Mitarbeiter,...), erfolgreich im Einsatz.

Literaturverzeichnis

- [1] Abowd, G. D., Dey, A. K., Brown, P. J., Davies, N., Smith, M., and Steggles, P. 1999. Towards a Better Understanding of Context and Context-Awareness. In *Proceedings of the 1st international Symposium on Handheld and Ubiquitous Computing* (Karlsruhe, Germany, September 27 - 29, 1999). H. Gellersen, Ed. Lecture Notes In Computer Science, vol. 1707. Springer-Verlag, London, 304-307.
- [2] Lei, H., Sow, D. M., Davis, J. S., Banavar, G., and Ebling, M. R. 2002. The design and applications of a context service. *SIGMOBILE Mob. Comput. Commun.* Volume 6, Issue 4, 45-55, Oct. 2002.
- [3] MIR3. Enabling the next-generation of IT notification capabilities. MIR3. 2009. Whitepaper. URL: http://www.mir3.com/misc/ESB_whitepaper/pdf/WhitePaper_MIR3_ESB.pdf.
- [4] Steen, M., Eijk, R., Koning, N. and Reitsema, E. A We-Centric Telecom Service for Police Officers to Support Communication. Enterprise Information Systems: 9th International Conference, Iceis 2007, Funchal, Madeira, June 12-16, 2007.
- [5] Streefkerk, J., van Esch-Bussemakers, M. and Neerincx, M. Context-Aware Notification for Mobile Police Officers. *Engin. Psychol. and Cog. Ergonomics, HCII 2007*, LNAI 4562, pp. 436-445, 2007.
- [6] Williams, D. and Curtis, D. Using Alert Notification Tools to Enhance IT Operations Management. Gartner, Inc. 2008. Whitepaper.
- [7] Williams, D. The Evolving Role of Alert Notification Tools in IT Operations. Gartner, Inc. 2008. Whitepaper.
- [8] Yildiz, B. A Dynamic Mobile Notification System for Incident Management in Monitoring Systems. 2009. Master Thesis. URL: http://berrendorf.inf.h-brs.de/lehre/abschlussarbeiten/arbeiten/2009_Thesis_Birol_Yildiz.pdf

Generic Parallel Computation of Large Garbled Circuits using JaMP

Thorsten Blaß*

Thorsten.Blass@informatik.stud.uni-erlangen.de

Abstract: This thesis evaluates the parallelization capabilities and the use of efficient memory usage for Yao’s Secure Function Evaluation approach. Yao’s algorithm is a very time and memory consuming cryptographic protocol. Existing techniques aimed at decreasing communication bandwidth and computation complexity but not on parallelization. Our goal is to achieve parallelism without platforms dependence or parallel computing architectures. Our solution is to use JaMP for implementing the resulting parallelized algorithm. JaMP is an OpenMP Framework for Java that can exploit any available parallel execution unit (CPU, GPU) at runtime. Although JaMP will be extended to be multi-GPU and cluster enabled.

Introduction

Secure Function Evaluation (SFE) allows the secure evaluation of a function on private input without the need of a trusted third party. Every party learns only the result of the function evaluation but nothing on the inputs of the involved parties [Yao86, LP04, LP07]. This thesis will focus on a subclass of SFE where only two parties are involved, named Two-party SFE.

Yao’s approach uses acyclic boolean circuits for representing the function which should be securely evaluated. This acyclic boolean circuit becomes encoded into a garbled circuit by applying a fixed transformation rule. The protocol execution can be outlined as follows:

The circuit constructor creates a garbled circuit GC : Assume the circuit consists only of one gate with two input wires (w_1, w_2) and one output wire (w_3) . The constructor encodes each wire w_i by selecting two different pseudo-random keys k_i^0 and k_i^1 of length k . These two keys replace the bit value 0 and 1 of a wire with k_i^0 resp. k_i^1 . The evaluator won’t know which key correspond to which bit value. The gate itself is then replaced by a four entry table (garbled table - GT). A table entry is given by

$$E_{k_1^{b_1}, k_2^{b_2}} \left(k_3^{G(b_1, b_2)} \right)$$

*This Diplomarbeit is supervised by Ronald Veldema, Ph.D. and Prof. Dr. Michael Philippsen (Friedrich Alexander Universität Erlangen-Nürnberg) and Dipl. Inf Thomas Schneider and Prof. Dr. Ahmad-Reza Sadeghi (Ruhr-Universität Bochum)

where b_1 and b_2 are the user input, $E_{k_1, k_2}(m)$ is an encryption function with two input keys of length k and a message m . The gate output is defined by function $G(p_1, p_2) \in \{0, 1\}$, with bit parameters p_1, p_2 . The same also applies for more gates, see [PSSW09] for detailed explanation. The evaluator obtains the key(s) corresponding to his input(s) and uses these to evaluate the garbled circuit by applying the inverse transformation.

Algorithm design and implementation is cost and time expensive. Therefore it is not desirable to implement an algorithm for every platform and hardware configuration. Computing environments will become more and more heterogeneous, e.g. more and more computers have graphic adapters which are capable of doing generic computations (e.g. NVIDIA Tesla [LNOM08]). JaMP [MKRVMBMP08, Dot09, VDKP09] is a framework which transparently adapt to this heterogeneity. It combines Java's write-once-run-anywhere approach with the OpenMP [DM98] standard to allow portable parallel code. For example:

```
class VectorAddition{
    void add(){
        //#omp parallel for
        for(int i = 0; i < N; i++)
            a[i] = b[i] + c[i];
    }
}
```

is valid JaMP code which will execute N additions in parallel. For-loops which are not capable of execution on the GPU will be translated into a threaded version using the (multi-core) CPU. Because no compile time assumptions on whether and which type of accelerator or graphics adapter might be available at runtime can be made, a specialized class-loader generates the necessary code on-the-fly. This code remains translatable by native Java compiler, because the parallel statements are encapsulated in a comment which will then be simply ignored.

Current approaches

As the creation and evaluation of garbled circuits requires computationally expensive operations (e.g. SHA-256, AES-128) on a huge amount of data, a lot of research has been done to decrease communication bandwidth and computation complexity: point-and-permute [MNPS04], "free" evaluation of XOR gates [KS08, Sch08] and garbled row reduction [PSSW09]. For a long time, SFE based on Yao's protocol [Yao86] has been considered a purely theoretical concept. Demonstrating that SFE is moving towards practical use, it has been implemented in software for small circuits (< 100.000 gates). However, circuits for many practical problems consist of millions or billions of gates. The efficient evaluation of such huge circuits has not been considered yet.

Benchmarks in [Dot09] showing that JaMP can speed up Java applications dramatically. Yet, the framework is still a prototype. It currently supports only one graphic adapter and is not cluster enabled. Furthermore, not all OpenMP directives are implemented [Dot09] and there are restrictions on statements that might be used in parallel regions, for example, no meta-object access and object allocation is allowed on the GPU [VDKP09].

Contributions of this work

This thesis investigates how the creation and evaluation of large GCs (millions to billions of gates) can be (generically) parallelized to exploit the performance of today's architectures.

When circuits grow large, neither the circuit itself nor the intermediate values (e.g bit-strings) can be stored entirely in memory. Partitioning techniques, caching strategies and gate schedules have to be evaluated to partition the circuit into proper data-chunks. A further point of interest is the transmission of the garbled circuit to the evaluator. There are two possibilities: First, the garbled circuit is created completely before being transmitted. Or second, all calculated gates will be transmitted on-the-fly and immediately evaluated by evaluator. Both possibilities have influence on the computational complexity of the used gate scheduling algorithm.

On behalf of JaMP it has to be evaluated how the framework fits the needs for implementing the parallelized algorithm. Where necessary the missing functionality will be implemented (e.g object allocation). Due to graphic adapters having limited memory capacity (compared to main memory), allocation of big data structures could fail (1). Furthermore the bandwidth from main memory to graphic memory, and vice versa, is a bottleneck (2). To address these two problems, JaMP uses array packages. For problem (1) array packages semi-automatically fragment an array into suitable chunks. These chunks are big enough to fit onto the graphic adapter memory. To solve (2) many objects were packed into one big chunk for bulk transfer. Currently, these packages have no support for multi-GPU systems and clusters. Therefore the array packages have to be modified or replaced. The new array packages become more complex as, synchronization is needed to hold data consistently, values can be used on different devices which implies dependencies. In the case of cluster use the network latency has to be hidden, operations and data has to be stored on the same computing device to minimize message/memory transfers. Especially the synchronization on NVIDIA's graphic adapters is difficult because there exist only limited synchronization possibilities¹. To ensure the write-once-runs-anywhere approach all problems have to be solved without assumption on a specific architecture.

¹Barrier synchronization is only for threads within the same thread block possible [NVI09]

References

- [DM98] L. Dagum and R. Menon. OpenMP: an industry standard API for shared-memory programming. *IEEE Computational Science and Engineering*, 5(1):46–55, 1998.
- [Dot09] G. Dotzler. Run-Time Parallelization of OpenMP/Java-Programs for the Execution on GPUs. Master’s thesis, University of Erlangen-Nuremberg, September 1, 2009.
- [KS08] V. Kolesnikov and T. Schneider. Improved Garbled Circuit: Free XOR Gates and Applications. In *International Colloquium on Automata, Languages and Programming (ICALP’08)*, volume 5126 of *LNCS*, pages 486–498. Springer, 2008.
- [LNOM08] E. Lindholm, J. Nickolls, S. Oberman, and J. Montrym. NVIDIA Tesla: A Unified Graphics and Computing Architecture. *Micro, IEEE*, 28(2):39–55, May 2008.
- [LP04] Y. Lindell and B. Pinkas. A proof of Yao’s protocol for secure two-party computation. ECCC Report TR04-063, Electronic Colloquium on Computational Complexity (ECCC), 2004.
- [LP07] Y. Lindell and B. Pinkas. An Efficient Protocol for Secure Two-Party Computation in the Presence of Malicious Adversaries. In *Advances in Cryptology – EUROCRYPT’07*, volume 4515 of *LNCS*, pages 52–78. Springer, 2007.
- [MKRVMBMP08] Michael Klemm, Ronald Veldema, Matthias Bezold, and Michael Philippsen. A Proposal for OpenMP for Java. In Matthias S. Mueller, Barbara M. Chapman, Bronis R. de Supinski, Allen D. Malony, and Michael Voss, editors, *OpenMP Shared Memory Parallel Programming (International Workshops IWOMP 2005 and IWOMP 2006)*, pages 409–421, Berlin, Germany, 2008. Springer.
- [MNPS04] D. Malkhi, N. Nisan, B. Pinkas, and Y. Sella. Fairplay — a secure two-party computation system. In *USENIX*, 2004. <http://www.cs.huji.ac.il/project/Fairplay/fairplay.html>.
- [NVI09] NVIDIA. CUDA Programming Guide 2.3. http://developer.download.nvidia.com/compute/cuda/2_3/toolkit/docs/NVIDIA_CUDA_Programming_Guide_2.3.pdf, 2009.
- [PSSW09] Benny Pinkas, Thomas Schneider, Nigel P. Smart, and Stephen C. Williams. Secure Two-Party Computation is Practical. In *Advances in Cryptology – ASIACRYPT 2009*, LNCS. Springer, December 6-10, 2009. Full version available at <http://eprint.iacr.org/2009/314>.
- [Sch08] T. Schneider. Practical Secure Function Evaluation. Master’s thesis, University of Erlangen-Nuremberg, February 27, 2008. Available at <http://thomaschneider.de/theses/da/>.
- [VDKP09] R. Veldema, G. Dotzler, M. Klemm, and M. Philippsen. jCudaMP: OpenMP/Java on CUDA. Technical report, University of Erlangen-Nuremberg, Computer Science Department 2, 2009.
- [Yao86] A. C. Yao. How to Generate and Exchange Secrets. In *IEEE Symposium on Foundations of Computer Science (FOCS’86)*, pages 162–167. IEEE, 1986.

Dynamisierung von Probability-Impact-Matrizen

Michael Hanspach

michael.hanspach@bsi.bund.de

1 Einleitung

Während die Einführung von Risikomanagementsystemen für Kapitalgesellschaften bereits verpflichtend ist (vgl. [Sch97]), profitieren fast alle Arten von Organisationen von der Durchführung einer Bedrohungs- und Risikoanalyse.

Probability-Impact-Matrizen visualisieren die Ergebnisse einer Bedrohungs- und Risikoanalyse und ermöglichen den Verantwortlichen so, einen schnellen Überblick über die erwartete Bedrohungslage zu erhalten. Da Probability-Impact-Matrizen üblicherweise von statischer Natur sind, ist die Aussagekraft des Modells in schnell veränderlichen Umfeldern eingeschränkt.

Im Rahmen dieser Arbeit wird eine Lösung für die Anwendung von Probability-Impact-Matrizen in dynamischen Einsatzfeldern erörtert.

2 Grundlagen

In einer **Probability-Impact-Matrix** werden Erwartungswerte von Bedrohungen mit dem Schaden bei Eintritt der Bedrohung in Beziehung gesetzt. Ziel ist die Bewertung und Darstellung des kalkulierten Risikos (Kosten im jeweiligen Betrachtungszeitraum) (vgl. [Woh06] und [Bau02]).

Die **Szenarioanalyse** (vgl. auch [AF02] und [AF06]) ist eine Methode zur Entwicklung unterschiedlicher Zukunftsszenarien und bildet die Grundlage des Konzepts dieser Arbeit. In einer Szenarioanalyse werden Schlüsselfaktoren für die Entwicklung des betrachteten Assets definiert und darauf aufbauend Zukunftsprojektionen erzeugt. Schließlich werden aus diesen Zukunftsprojektionen mögliche Zukunftsszenarien gebildet (vgl. [Wil06]). Die Zukunftsszenarien eines sicherheitskritischen Assets können mittels Probability-Impact-Matrizen dargestellt und interpretiert werden, wie im Konzept beschrieben.

3 Konzept für die Dynamisierung von Probability-Impact-Matrizen

Kern dieser Arbeit ist die Entwicklung eines Anwendungskonzepts zur Lösung des Problems der Dynamisierung von Probability-Impact-Matrizen. Es werden Anforderungen an die Lösung des Problems sowie ein Phasenmodell zur Beschreibung des Dynamisierungs-Prozesses definiert.

Der hier entwickelte Dynamisierungs-Prozess besteht aus 5 Schritten, die in serieller Abfolge durchlaufen werden:

Prozessschritt	Beschreibung
1	Eine konkrete <i>Fragestellung für die Bedrohungs- und Risikoanalyse wird formuliert</i> und ein konkretes technisches oder organisatorisches Betrachtungsfeld sowie Betrachtungs- und Planungszeiträume werden definiert.
2	Es werden <i>Bedrohungsdaten und -statistiken gesammelt</i> . Mögliche Datenquellen werden nach Art und Umfang der Daten selektiert. Für die Weiterverarbeitung werden die Daten kategorisiert.
3	Aus den so gesammelten Daten werden die entscheidenden Schlüsselfaktoren extrahiert, sowie Zukunftsprojektionen und - in einem weiteren Schritt - <i>Zukunftsszenarien gebildet</i> .
4	Jedes Szenario wird mittels einer Probability-Impact-Matrix <i>visualisiert</i> .
5	Schließlich werden die <i>Ergebnisse bewertet</i> , um ein Gesamtbild der prognostizierten Bedrohungslage zu erhalten und den Verantwortlichen die Möglichkeit zu geben, entsprechende Maßnahmen gegen kritische Bedrohungen einzuleiten.

Für jeden Prozessschritt werden das notwendige Vorgehen sowie die verwendeten Formeln und Algorithmen definiert. Um das Konzept auf möglichst unterschiedliche Einsatzfelder verallgemeinern zu können, werden dabei möglichst allgemein und unabhängig vom technischen oder organisatorischen Umfeld gehaltene Methoden verwendet. Eine Fallstudie dient als Beleg der *Durchführbarkeit* des Konzepts.

4 Fallstudie: Linux 2.6.x

Ziel der Fallstudie ist es, die Durchführung des entwickelten Konzepts an einem praktischen Beispiel zu zeigen. Thema der Fallstudie ist die Bedrohungs- und Risikoanalyse mittels dynamisierter Probability-Impact-Matrizen am Beispiel der Software *Linux 2.6.x*, zu der bereits umfangreiche Statistiken für eine Prognose existieren (vgl. [KH08]).

Quelle der Bedrohungsdaten hier ist die Firma Secunia, die auf ihren Websites regelmäßig Advisories über neue Schwachstellen in Linux 2.6.x und anderen Software-Produkten veröffentlicht (vgl. [Sec09]). Aus diesen Daten werden mittels Anwendung der Szenarioanalyse die relevanten Kategorien von Bedrohungen extrahiert. Dies sind in diesem Fall die Kategorien: *Denial of Service (DoS)*, *Exposure of sensitive Information*, *Privilege Escalation* und *Security Bypass*.

Ergebnis der Szenarioentwicklung sind: Das *Trendszenario*, das *Best-Case-Szenario* und das *Worst-Case-Szenario*. Für jedes Szenario wird eine Probability-Impact-Matrix der relevanten Bedrohungskategorien erzeugt.

Die Interpretation der Probability-Impact-Matrizen zeigt schließlich, dass im Planungszeitraum *DoS*-Angriffe dominieren. Im *Worst-Case-Szenario* jedoch liegen die Erwartungswerte eines *Security Bypass* sogar noch höher als die einer *DoS*-Attacke. Sowohl das *Trendszenario* als auch das *Best-Case-Szenario* definieren sehr ähnliche Häufigkeiten für das Auftreten der Sicherheitsvorfälle *Exposure of sensitive information*, *Privilege Escalation* und *Security Bypass*. Einen deutlichen Unterschied zwischen den beiden Szenarien gibt es lediglich in der Häufigkeit neu entdeckter *DoS*-Attacken.

5 Fazit und Ausblick

In dieser Arbeit stellt der Autor ein Konzept für die Anwendung von Probability-Impact-Matrizen in dynamischen Umfeldern vor. Hierzu werden die Ziele der Betrachtung (*Assets*) definiert und Daten zur Beschreibung der historischen Bedrohungslage sowie aktueller Entwicklungen der Assets gesammelt. Mittels Durchführung einer Szenarioanalyse werden die Bedrohungslage eines zukünftigen Planungszeitraums prognostiziert und verschiedene Zukunftsszenarien entwickelt. Die Bedrohungsdaten der Zukunftsszenarios werden jeweils in eine Probability-Impact-Matrix übertragen, welche anschließend interpretiert und bewertet werden kann.

Die Durchführbarkeit des Konzepts wird mittels einer Fallstudie gezeigt. Im Ergebnis der Fallstudie wurde die Häufigkeit sicherheitskritischer Schwachstellen in Linux 2.6.x für den Zeitraum von 2010 bis 2014 prognostiziert und in mehreren Probability-Impact-Matrizen visualisiert. Je höher die Zahl der auftretenden Schwachstellen in diesem Planungszeitraum schließlich ist, desto höher sind auch die Kosten für den Einsatz der Betriebssystem-Software. Dieses Ergebnis könnte nun mit den Bedrohungsdaten ähnlicher Betriebssystem-Software verglichen werden, um eine Entscheidung für oder wider den Einsatz von Linux 2.6.x anstelle eines anderen Betriebssystems zu treffen.

Literatur

- [AF02] Andreas Siebe Alexander Fink, Oliver Schlake. *Erfolg durch Szenario-Management*. campus, 2002.
- [AF06] Andreas Siebe Alexander Fink. *Handbuch Zukunftsmanagement*. campus, 2006.
- [Bau02] Mick Bauer. *Practical Threat Analysis and Risk Management*. Linux Journal, 2002.
- [KH08] Greg Kroah-Hartman. *Linux Kernel Development, How Fast it is Going, Who is Doing It, What They are Doing, and Who is Sponsoring It*. Linux Foundation, 2008.
- [Sch97] Gerhard Scherrer. *KonTraG*. Universität Regensburg, 1997.
- [Sec09] Secunia. *Vulnerability Report: Linux Kernel 2.6.x*. Secunia, 2009.
- [Wil06] Falko E.P. Wilms. *Szenariotechnik*. Haupt, 2006.
- [Woh06] Stefan Wohlfeil. *Sicherheit im Internet II*. Fernuniversität in Hagen, 2006.

Streaming-based Processing of Secured XML Documents

Juraj Somorovsky
Ruhr-University Bochum
juraj.somorovsky@rub.de

Master Thesis Supervisors: Prof. Dr. Jörg Schwenk, Dipl.-Inf. Meiko Jensen

Abstract: WS-Security is a standard that allows flexible application of security mechanisms in Web Services. However, using sophisticated security algorithms can lead to high memory consumptions and long evaluation times. In combination with the standard XML processing approach (DOM), the processing of large XML documents becomes impossible.

This paper presents an approach for processing of large XML documents including XML Signatures. For implementation of our solution, we used the Event Pipeline Pattern. This pattern offers much flexibility by developing of streaming-based frameworks. The evaluation shows that such a framework enhances the performance and is much more efficient in comparison to standard DOM-based solutions.

1 Introduction

In Service Oriented Architectures, one of the most common technologies to exchange data and execute remote procedures is Web Services [HBN⁺04]. Peers using the Web Services communicate with each other by sending and receiving XML-based *SOAP messages* [NHK⁺07]. These messages are transported over standard protocols such as HTTP. As they often include confidential data, they have to be secured. Therefore, the Organization for the Advancement of Structured Information Standards (OASIS) defined the *WS-Security* standard [LK06]. This standard offers flexible application of security mechanisms on arbitrary document parts and ensures that SOAP messages provide integrity, confidentiality, and authenticity.

However, the use of WS-Security supports attacks on the Web Services' availability. The first reason contributing to the performance problem is the complexity of security algorithms. The second reason is the standard XML-processing model: *DOM* [BHH⁺04]. This model is applied in most of today's platforms. It parses the incoming SOAP message into an object tree. The object tree cannot be evaluated until the whole SOAP message is processed. The storage of all the elements' data causes high memory consumptions and eliminates high performance processing of large SOAP documents.

This paper provides a solution for the problems with processing of large XML documents, concretely SOAP messages. It presents an *Event Pipeline pattern* for evaluation of SOAP messages including XML Signatures [RSH⁺08]. This pattern uses a *streaming-based (event-based)* approach, which brings two major advantages. Firstly, it does not store the whole document in the memory. Additionally, the document is parsed step-by-step and the parsing can be interrupted as soon as a security violation is found.

2 Event Pipeline Pattern

Common usage of streaming-based XML parsers can be found in pipelined XML processors. The pipelined XML processor is a system of modules, which together create a pipe. The first module communicates with streaming-based XML parser and expects generated events. When the event comes, it is passed to the next modules in the pipe. This way, each module can get and process generated events. An example of such document processing gives the Event Pipeline pattern [GJI10].

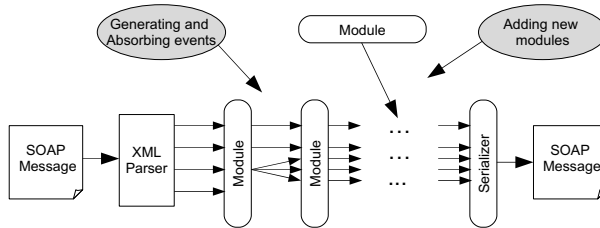


Figure 1: Streaming-based XML processing with Event Pipeline pattern

The application that uses Event Pipeline pattern consists usually of three parts: a document parser, some filtering or checking modules and a sink (e.g. serializer, which buffers a message). If the structure of the XML document is known, modules can be added statically to the pipe on the beginning of document parsing. Otherwise, there is also a possibility to insert modules at runtime during the parsing. When a module is added to the pipe, it automatically gets all the events passed or generated by its ancestors in the pipe. The implementation logic in the module tells, what to do with the event: It can be processed, absorbed, passed to the next module, or it can be generated to a completely new event (Figure 1).

3 Streaming-based XML Signature Processing

XML Signature is a W3C standard. It allows to sign arbitrary elements in XML documents. In SOAP messages, XML Signature is defined in the SOAP header (Figure 2). It consists mainly of these two parts: *SignedInfo* and *SignatureValue*. In *<SignedInfo>*, used cryptographic algorithms and references to the signed elements are defined. The whole *<SignedInfo>* block is signed with the peer's private key and the signature value is put to the *<SignatureValue>* element.

To verify XML Signatures in SOAP messages, first, the signature value generated over the *<SignedInfo>* block has to be validated. Then, all the references and their digest values have to be approved. For the validation framework, it has been decided to use these modules:

- *SAX Parser* [Meg04] parses the SOAP message into events and passes them to the next handlers.
- *SecurityExceptionHandler* handles security exceptions that are caused by security violation (incorrect reference or signature values) or by incorrect SOAP

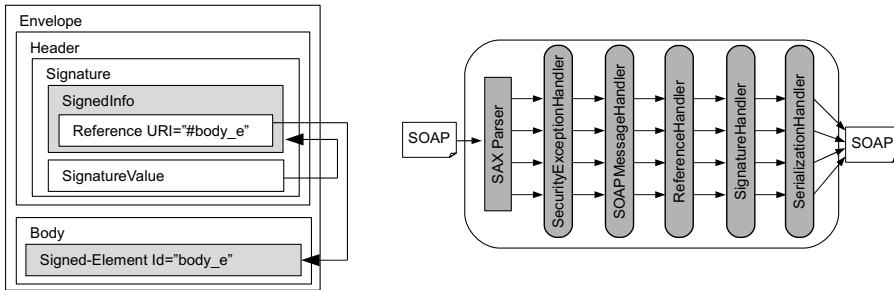


Figure 2: XML Signature in SOAP message and validation architecture

message format.

- `SOAPMessageHandler` controls the structure of SOAP messages.
- `ReferenceHandler` verifies referenced elements.
- `SignatureHandler` processes XML Signature and validates signature value.
- `SerializationHandler` serializes all the generated events to output a SOAP message.

4 Evaluation

The implementation was tested using a set of generated SOAP messages, in which the number of elements in the SOAP body was increased. Each SOAP body was signed using XML Signature. For evaluation purposes, the processing time was measured and compared to Java XML Digital Signature API [Mic06]. This API uses standard tree-based processing.

The test ran on a machine with 64 bit AMD Sempron 3400+ (1.8 GHz) processor and 1024 MB RAM. The machine used Linux Ubuntu system (version 9.04) with Java 6. The maximum Java Virtual Machine memory was set to 84 MB.

Figure 3 presents the results of the test comparing messages with up to 140,000 elements. As can be seen, the evaluation times measured with the streaming-based approach grow linearly with enlarging of the SOAP messages. The framework could validate a SOAP message with 140,000 elements in 625 milliseconds. On the other hand, the evaluation times of the DOM approach increase rapidly by the messages with more than 90,000 elements. By using the tree-based approach, processing documents documents with more than 140,000 elements caused out-of-memory exceptions. The streaming-based approach did not need more than 3 MB.

5 Conclusion and Outlook

In this paper, we discussed Event Pipeline Pattern and its application on the streaming-based XML processing. We developed a framework that is capable of XML Signature verification in SOAP messages. Our implementation was compared to the standard tree-

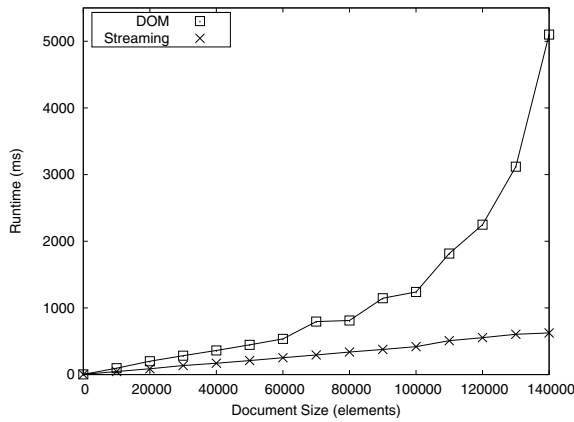


Figure 3: Evaluation time comparison with maximum 140,000 elements

based approach. The results show that the streaming-based approach is much faster and more memory efficient. Therefore, it could find application in Web Services firewalls to secure from Denial-of-Service attacks.

As XML Signature is one of the many standards used in WS-Security, our implementation is open for many other enhancements such as XML Encryption processing or WS-Security Policy validation.

References

- [BHH⁺04] Steve Byrne, Arnaud Le Hors, Philippe Le Hégarret, Mike Champion, Gavin Nicol, Jonathan Robie, and Lauren Wood. Document Object Model (DOM) Level 3 Core Specification. W3C recommendation, W3C, April 2004.
- [GJI10] Nils Gruschka, Meiko Jensen, and Luigi Lo Iacono. A Design Pattern for Event-Based Processing of Security-enriched SOAP Messages. In *SAGC*, 2010.
- [HBN⁺04] Hugo Haas, David Booth, Eric Newcomer, Mike Champion, David Orchard, Christopher Ferris, and Francis McCabe. Web Services Architecture. W3C note, W3C, February 2004.
- [LK06] Kelvin Lawrence and Chris Kaler. Web Services Security v1.1. Technical report, OASIS Open, February 2006.
- [Meg04] David Megginson. *SAX 2.0: The Simple API for XML*. SAX project, 2004.
- [Mic06] Sun Microsystems. *XML Digital Signature API*, 2006.
- [NHK⁺07] Henrik Frystyk Nielsen, Marc Hadley, Anish Karmarkar, Noah Mendelsohn, Yves Lafon, Martin Gudgin, and Jean-Jacques Moreau. SOAP Version 1.2 Part 1: Messaging Framework (Second Edition). W3C recommendation, W3C, April 2007.
- [RSH⁺08] Thomas Roessler, David Solo, Frederick Hirsch, Donald Eastlake, and Joseph Reagle. XML Signature Syntax and Processing (Second Edition). W3C recommendation, W3C, June 2008.

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: Biometrics and Electronic Signatures

- P-32 Peter Hubwieser (Hrsg.): Informatische Fachkonzepte im Unterricht – INFOS 2003
- P-33 Andreas Geyer-Schulz, Alfred Taudes (Hrsg.): Informationswirtschaft: Ein Sektor mit Zukunft
- P-34 Klaus Dittrich, Wolfgang König, Andreas Oberweis, Kai Rannenberg, Wolfgang Wahlster (Hrsg.): Informatik 2003 – Innovative Informatikanwendungen (Band 1)
- P-35 Klaus Dittrich, Wolfgang König, Andreas Oberweis, Kai Rannenberg, Wolfgang Wahlster (Hrsg.): Informatik 2003 – Innovative Informatikanwendungen (Band 2)
- P-36 Rüdiger Grimm, Hubert B. Keller, Kai Rannenberg (Hrsg.): Informatik 2003 – Mit Sicherheit Informatik
- P-37 Arndt Bode, Jörg Desel, Sabine Rathmayer, Martin Wessner (Hrsg.): DeLFI 2003: e-Learning Fachtagung Informatik
- P-38 E.J. Sinz, M. Plaha, P. Neckel (Hrsg.): Modellierung betrieblicher Informationssysteme – MobIS 2003
- P-39 Jens Nedon, Sandra Frings, Oliver Göbel (Hrsg.): IT-Incident Management & IT-Forensics – IMF 2003
- P-40 Michael Rebstock (Hrsg.): Modellierung betrieblicher Informationssysteme – MobIS 2004
- P-41 Uwe Brinkschulte, Jürgen Becker, Dietmar Fey, Karl-Erwin Großpietsch, Christian Hochberger, Erik Maehle, Thomas Runkler (Edts.): ARCS 2004 – Organic and Pervasive Computing
- P-42 Key Pousttchi, Klaus Turowski (Hrsg.): Mobile Economy – Transaktionen und Prozesse, Anwendungen und Dienste
- P-43 Birgitta König-Ries, Michael Klein, Philipp Obreiter (Hrsg.): 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 Fähnrich, Wolfgang S. Wittig (Hrsg.): Marktplatz Internet: Von e-Learning bis e-Payment
- P-73 Jan von Knop, Wilhelm Haverkamp, Eike Jessen (Hrsg.): "Heute schon das Morgen sehen"
- P-74 Christopher Wolf, Stefan Lucks, Po-Wah Yau (Hrsg.): WEWoRC 2005 – Western European Workshop on Research in Cryptology
- P-75 Jörg Desel, Ulrich Frank (Hrsg.): Enterprise Modelling and Information Systems Architecture
- P-76 Thomas Kirste, Birgitta König-Riess, Key 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 1st 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 3rd International Conference on IT-Incident Management & IT-Forensics
- P-115 Claudia Falter, Alexander Schliep, Joachim Selbig, Martin Vingron and Dirk Walther (Eds.) German conference on bioinformatics GCB 2007
- P-116 Witold Abramowicz, Leszek Maciszek (Eds.) Business Process and Services Computing 1st International Working Conference on Business Process and Services Computing BPSC 2007
- P-117 Ryszard Kowalczyk (Ed.) Grid service engineering and management The 4th International Conference on Grid Service Engineering and Management GSEM 2007
- P-118 Andreas Hein, Wilfried Thoben, Hans-Jürgen Appelrath, Peter Jensch (Eds.) European Conference on ehealth 2007
- P-119 Manfred Reichert, Stefan Strecker, Klaus Turowski (Eds.) Enterprise Modelling and Information Systems Architectures Concepts and Applications
- P-120 Adam Pawlak, Kurt Sandkuhl, Wojciech Cholewa, Leandro Soares Indrusiak (Eds.) Coordination of Collaborative Engineering - State of the Art and Future Challenges
- P-121 Korbinian Herrmann, Bernd Bruegge (Hrsg.) Software Engineering 2008 Fachtagung des GI-Fachbereichs Softwaretechnik
- P-122 Walid Maalej, Bernd Bruegge (Hrsg.) Software Engineering 2008 - Workshopband Fachtagung des GI-Fachbereichs Softwaretechnik

- P-123 Michael H. Breitner, Martin Breunig, Elgar Fleisch, Ley Pousttchi, Klaus Turowski (Hrsg.)
Mobile und Ubiquitäre Informationssysteme – Technologien, Prozesse, Marktfähigkeit
Proceedings zur 3. Konferenz Mobile und Ubiquitäre Informationssysteme (MMS 2008)
- P-124 Wolfgang E. Nagel, Rolf Hoffmann, Andreas Koch (Eds.)
9th Workshop on Parallel Systems and Algorithms (PASA)
Workshop of the GI/ITG Special Interest Groups PARS and PARVA
- P-125 Rolf A.E. Müller, Hans-H. Sundermeier, Ludwig Theuvsen, Stephanie Schütze, Marlies Morgenstern (Hrsg.)
Unternehmens-IT:
Führungsinstrument oder Verwaltungsbürde
Referate der 28. GIL Jahrestagung
- P-126 Rainer Gimnich, Uwe Kaiser, Jochen Quante, Andreas Winter (Hrsg.)
10th Workshop Software Reengineering (WSR 2008)
- P-127 Thomas Kühne, Wolfgang Reisig, Friedrich Steimann (Hrsg.)
Modellierung 2008
- P-128 Ammar Alkassar, Jörg Siekmann (Hrsg.)
Sicherheit 2008
Sicherheit, Schutz und Zuverlässigkeit
Beiträge der 4. Jahrestagung des Fachbereichs Sicherheit der Gesellschaft für Informatik e.V. (GI)
2.-4. April 2008
Saarbrücken, Germany
- P-129 Wolfgang Hesse, Andreas Oberweis (Eds.)
Sigsand-Europe 2008
Proceedings of the Third AIS SIGSAND European Symposium on Analysis, Design, Use and Societal Impact of Information Systems
- P-130 Paul Müller, Bernhard Neumair, Gabi Dreö Rodosek (Hrsg.)
1. DFN-Forum Kommunikationstechnologien Beiträge der Fachtagung
- P-131 Robert Krimmer, Rüdiger Grimm (Eds.)
3rd International Conference on Electronic Voting 2008
Co-organized by Council of Europe, Gesellschaft für Informatik and E-Voting.CC
- P-132 Silke Seehusen, Ulrike Lucke, Stefan Fischer (Hrsg.)
DeLFI 2008:
Die 6. e-Learning Fachtagung Informatik
- P-133 Heinz-Gerd Hegering, Axel Lehmann, Hans Jürgen Ohlbach, Christian Scheideler (Hrsg.)
INFORMATIK 2008
Beherrschbare Systeme – dank Informatik Band 1
- P-134 Heinz-Gerd Hegering, Axel Lehmann, Hans Jürgen Ohlbach, Christian Scheideler (Hrsg.)
INFORMATIK 2008
Beherrschbare Systeme – dank Informatik Band 2
- P-135 Torsten Brinda, Michael Fothe, Peter Hubwieser, Kirsten Schlüter (Hrsg.)
Didaktik der Informatik – Aktuelle Forschungsergebnisse
- P-136 Andreas Beyer, Michael Schroeder (Eds.)
German Conference on Bioinformatics GCB 2008
- P-137 Arslan Brömme, Christoph Busch, Detlef Hühnlein (Eds.)
BIOSIG 2008: Biometrics and Electronic Signatures
- P-138 Barbara Dinter, Robert Winter, Peter Chamoni, Norbert Gronau, Klaus Turowski (Hrsg.)
Synergien durch Integration und Informationslogistik
Proceedings zur DW2008
- P-139 Georg Herzwurm, Martin Mikusz (Hrsg.)
Industrialisierung des Software-Managements
Fachtagung des GI-Fachausschusses Management der Anwendungsentwicklung und -wartung im Fachbereich Wirtschaftsinformatik
- P-140 Oliver Göbel, Sandra Frings, Detlef Günther, Jens Nedon, Dirk Schadt (Eds.)
IMF 2008 - IT Incident Management & IT Forensics
- P-141 Peter Loos, Markus Nüttgens, Klaus Turowski, Dirk Werth (Hrsg.)
Modellierung betrieblicher Informationssysteme (MobIS 2008)
Modellierung zwischen SOA und Compliance Management
- P-142 R. Bill, P. Korduan, L. Theuvsen, M. Morgenstern (Hrsg.)
Anforderungen an die Agrarinformatik durch Globalisierung und Klimaveränderung
- P-143 Peter Liggesmeyer, Gregor Engels, Jürgen Münch, Jörg Dörr, Norman Riegel (Hrsg.)
Software Engineering 2009
Fachtagung des GI-Fachbereichs Softwaretechnik

- P-144 Johann-Christoph Freytag, Thomas Ruf, Wolfgang Lehner, Gottfried Vossen (Hrsg.)
Datenbanksysteme in Business, Technologie und Web (BTW)
- P-145 Knut Hinkelmann, Holger Wache (Eds.)
WM2009: 5th Conference on Professional Knowledge Management
- P-146 Markus Bick, Martin Breunig, Hagen Höpfner (Hrsg.)
Mobile und Ubiquitäre Informationssysteme – Entwicklung, Implementierung und Anwendung
4. Konferenz Mobile und Ubiquitäre Informationssysteme (MMS 2009)
- P-147 Witold Abramowicz, Leszek Maciaszek, Ryszard Kowalczyk, Andreas Speck (Eds.)
Business Process, Services Computing and Intelligent Service Management
BPSC 2009 · ISM 2009 · YRW-MBP 2009
- P-148 Christian Erfurth, Gerald Eichler, Volkmar Schau (Eds.)
9th International Conference on Innovative Internet Community Systems
I²CS 2009
- P-149 Paul Müller, Bernhard Neumair, Gabi Dreo Rodosek (Hrsg.)
2. DFN-Forum
Kommunikationstechnologien
Beiträge der Fachtagung
- P-150 Jürgen Münch, Peter Liggesmeyer (Hrsg.)
Software Engineering
2009 - Workshopband
- P-151 Armin Heinzl, Peter Dadam, Stefan Kirn, Peter Lockemann (Eds.)
PRIMIUM
Process Innovation for Enterprise Software
- P-152 Jan Mendling, Stefanie Rinderle-Ma, Werner Esswein (Eds.)
Enterprise Modelling and Information Systems Architectures
Proceedings of the 3rd Int'l Workshop
EMISA 2009
- P-153 Andreas Schwill, Nicolas Apostolopoulos (Hrsg.)
Lernen im Digitalen Zeitalter
DeLFI 2009 – Die 7. E-Learning Fachtagung Informatik
- P-154 Stefan Fischer, Erik Machle, Rüdiger Reischuk (Hrsg.)
INFORMATIK 2009
Im Focus das Leben
- P-155 Arslan Brömme, Christoph Busch, Detlef Hühnlein (Eds.)
BIOSIG 2009:
Biometrics and Electronic Signatures
Proceedings of the Special Interest Group on Biometrics and Electronic Signatures
- P-156 Bernhard Koerber (Hrsg.)
Zukunft braucht Herkunft
25 Jahre »INFOS – Informatik und Schule«
- P-157 Ivo Grosse, Steffen Neumann, Stefan Posch, Falk Schreiber, Peter Stadler (Eds.)
German Conference on Bioinformatics
2009
- P-158 W. Claupein, L. Theuvsen, A. Kämpf, M. Morgenstern (Hrsg.)
Precision Agriculture
Reloaded – Informationsgestützte Landwirtschaft
- P-159 Gregor Engels, Markus Luckey, Wilhelm Schäfer (Hrsg.)
Software Engineering 2010
- P-161 Gregor Engels, Dimitris Karagiannis, Heinrich C. Mayr (Hrsg.)
Modellierung 2010
- P-162 Maria A. Wimmer, Uwe Brinkhoff, Siegfried Kaiser, Dagmar Lück-Schneider, Erich Schweighofer, Andreas Wiebe (Hrsg.)
Vernetzte IT für einen effektiven Staat
Gemeinsame Fachtagung
Verwaltungsinformatik (FTVI) und
Fachtagung Rechtsinformatik (FTRI) 2010
- P-163 Markus Bick, Stefan Eulgem, Elgar Fleisch, J. Felix Hampe, Birgitta König-Ries, Franz Lehner, Key Pousttchi, Kai Rannenberg (Hrsg.)
Mobile und Ubiquitäre Informationssysteme
Technologien, Anwendungen und Dienste zur Unterstützung von mobiler Kollaboration

GI-Edition Lecture Notes in Informatics – Seminars

- S-1 Johannes Magenheimer, Sigrid Schubert
 (Eds.):
 Informatics and Student Assessment
 Concepts of Empirical Research and
 Standardisation of Measurement in the
 Area of Didactics of Informatics
- S-2 Gesellschaft für Informatik (Hrsg.)
 Informationstage 2005
 Fachwissenschaftlicher Informatik-
 Kongress
- S-3 Gesellschaft für Informatik (Hrsg.)
 Informationstage 2006
 Fachwissenschaftlicher Informatik-
 Kongress
- S-4 Hans Hagen, Andreas Kerren, Peter
 Dannenmann (Eds.)
 Visualization of Large and Unstructured
 Data Sets
 First workshop of the DFG's International
 Research Training Group "Visualization
 of Large and Unstructured Data Sets –
 Applications in Geospatial Planning,
 Modeling and Engineering"
- S-5 Gesellschaft für Informatik (Hrsg.)
 Informationstage 2007
 Fachwissenschaftlicher Informatik-
 Kongress
- S-6 Gesellschaft für Informatik (Hrsg.)
 Informationstage 2008
 Fachwissenschaftlicher Informatik-
 Kongress
- S-7 Hans Hagen, Martin Hering-Bertram,
 Christoph Garth (Eds.)
 Visualization of Large and Unstructured
 Data Sets
- S-8 Gesellschaft für Informatik (Hrsg.)
 Informatiktage 2009
 Fachwissenschaftlicher Informatik-
 Kongress
- S-9 Gesellschaft für Informatik (Hrsg.)
 Informatiktage 2010
 Fachwissenschaftlicher Informatik-
 Kongress

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

